Smbe 2019
21-25 JULY
MANCHESTER | ENGLAND

SMBE2019.org

ABSTRACT BOOK

SMBE 2019 21-25 JULY 2019

Manchester Central, Manchester, United Kingdom

@OfficialSMBE
#SMBE2019

SMBE 2019, 21-25 July 2019, Manchester, England

Within this document the following content:

Walter Fitch Award Abstracts

Understanding the genomics of climate change response

SMBE-WFA-001

The origin and spread of locally adaptive seasonal camouflage in snowshoe hares

M. R. Jones [1,*]

[1]University of Montana, Missoula, United States

Abstract: Adaptation is central to the persistence of biodiversity to changing environments, yet we rarely understand the mechanisms underlying the origin and spread of adaptive variation. At least 21 mammal and bird species undergo photoperiod-induced seasonal molts between brown (summer) and white (winter) coats. In snowshoe hares (*Lepus americanus*), direct field estimates of survival reveal that mismatch between coat color and snow cover increases predation risk, indicating seasonal color molts evolved to maintain camouflage in snowy environments. Interestingly, some snowshoe hare populations have adapted to mild coastal environments by molting into brown winter coats. We used population genomics, association mapping, and allele-specific expression to understand the molecular basis and evolutionary history of locally adaptive seasonal camouflage in snowshoe hares. We discovered that *cis*-regulatory variation at the pigmentation gene *Agouti* underlies alternative winter pelage morphs. Phylogenomic analyses and coalescent simulations demonstrated that brown winter pelage arose through relatively ancient hybridization with black-tailed jackrabbits (*Lepus californicus*). In contrast to these signatures of ancient introgression, we found evidence that strong positive selection for the introgressed *Agouti* allele may have enabled the recent colonization of coastal environments along the range edge during a climatic warming period. Additional whole genome sequencing has revealed that brown winter-camouflage has evolved independently along other parts of the range edge, suggesting limitations to the broad spread of locally adaptive variation may be counteracted by convergent adaptation. These discoveries provide important insights into how hybridization and convergence shape past and ongoing adaptation to changing seasonal environments.

Open Symposium

SMBE-WFA-002

Adaptive evolution at a meiosis gene mediates species differences in the rate and patterning of recombination

C. Brand [1,*]

[1]University of Pennsylvania, Pennsylvania, United States

Abstract: Crossing over between homologous chromosomes during meiosis repairs programmed DNA double-strand breaks, ensures proper segregation, enhances the efficacy of natural selection among genetically linked sites, and determines the genomic distribution of nucleotide variability in populations. Little however is known about the molecular genetic changes or population genetic forces involved in the evolution of recombination rates between species. We show that a dicistronic meiosis gene, *mei-217/mei-218*, with a history of rapid evolution acts as a global, trans-acting modifier of the rate and chromosomal distribution of crossing over between two closely related *Drosophila* species. Using transgenic flies, we find that species differences in crossing over are attributable to changes in the strengths of crossover assurance, crossover interference, and centromeric suppression of crossing over. We speculate that rates of crossing over evolved in part to mitigate fluctuating, species-specific risks of ectopic recombination between non-homologous transposon insertions. Regardless of its causes, the evolution of *mei-217/mei-218*-mediated changes in recombination landscapes may contribute to downstream species differences such as the chromosomal distribution of nucleotide variability and rates of nondisjunction. Finally, we investigate the deeper phylogenetic history, causes, and consequences of evolution of our meiosis gene and its interactors over the *Drosophila* phylogeny.

Genetic conflicts in molecular evolution

SMBE-WFA-003

Weird gene in a weird mammal: A highly divergent pancreatic duodenal homeobox 1 (Pdx1) gene in the fat sand rat

Y. S. Dai [1,*], P. W. Holland [1]

[1]Department of Zoology, University of Oxford, Oxford, United Kingdom

Abstract: Various forces leading to strong GC skew in local genomic regions cause conflict between increasing GC levels and alteration of conserved amino acids. In most cases, natural selection will purge any deleterious alleles that arise. However, in the gerbil subfamily of rodents, several conserved genes serving key functions have undergone radical alteration in association with strong GC skew. We present an extreme example concerning the highly conserved homeobox gene *Pdx1*, a key gene in initiation of pancreatic organogenesis in embryonic development. In the fat sand rat *Psammomys obesus* and close relatives, we observe a highly divergent *Pdx1* gene associated with high GC content. In this study, we investigate the antagonistic interplay between very rare amino acid changes driven by GC skew and the force of natural selection. Using ectopic protein expression in cell culture, pulse-chase labelling, *in vitro* mutagenesis and drug treatment, we compare properties of mouse and sand rat Pdx1 proteins. We find that amino acid changes driven by GC skew resulted in altered protein stability, with a significantly longer protein half-life for sand rat Pdx1. We show that both sand rat and mouse Pdx1 are degraded through the ubiquitin proteasome pathway. However, *in vitro* mutagenesis reveals that GC skew has caused loss of a key ubiquitination site conserved through vertebrate evolution, and we suggest that sand rat Pdx1 may have evolved a new ubiquitination site to compensate. Our results give molecular insight into the conflict between natural selection and genetic changes driven by strong GC skew.

Contemporary Evolution
SMBE-WFA-004
Aquatic adaptation and fur trade devastation:
a deep dive into the genomes of the sea otter and giant otter

A. C. Beichman [1,*], K.-P. Koepfli [2 3], G. Li [4], W. Murphy [5], P. Dobrynin [2 3], S. Kliver [3], M. T. Tinker [6], M. Murray [7], J. Johnson [8], K. Lindblad-Toh [8 9], E. Karlsson [8 10], K. Lohmueller [1], R. Wayne [1]
[1]University of California, Los Angeles, Los Angeles, [2]Smithsonian Conservation Biology Institute, Washington D.C., United States, [3]Theodosius Dobzhansky Center for Genome Bioinformatics, St. Petersburg, Russian Federation, [4]Shaanxi Normal University, Xi'an, China, [5]Texas A&M University, College Station, [6]US Geological Survey, Santa Cruz, [7]Monterey Bay Aquarium, Monterey, [8]Broad Institute of MIT and Harvard, Cambridge, United States, [9]Uppsala University, Uppsala, Sweden, [10]University of Massachusetts Medical School, Worcester, United States

Abstract: Despite its recent invasion of the marine realm, the sea otter (*Enhydra lutris*) evolved a suite of adaptations for life in cold coastal waters, most notably its dense insulating fur. This remarkable coat led to the near-extinction of sea otters during the 18th-20th century fur trade, causing an extreme bottleneck that sea otter populations are still recovering from. We compared the *de novo* genome of the southern sea otter (*E. l. nereis*) to that of the tropical freshwater giant otter (*Pteronura brasiliensis*) to reconstruct each species' evolutionary history, identify genes influencing ongoing aquatic adaptation, and assess the lingering impacts of the devastation caused by the fur trade. In both otter lineages, we found signals of positive selection in genes related to aquatic adaptations, including polygenic selection on hair follicle development genes and widespread pseudogenization of olfactory receptor genes. These findings illuminate genetic mechanisms that underlie the rapid evolution of otters to the aquatic realm. At the population level, the sea otter genome showed extremely low genomic diversity, signals of recent inbreeding, and elevated levels of deleterious variants. To explore these signals further, we generated exome and neutral sequence capture probes to sequence 130 sea otters across five bottlenecked populations. These data enable us to use the site frequency spectrum and simulations to carry out demographic inference and assess the impacts of the fur trade bottleneck on selection. The genomic legacy of the fur trade is a contemporary evolutionary challenge that could impact future recovery and resilience of sea otter populations.

Contemporary Evolution

SMBE-WFA-005

Multiple modes of convergent adaptation in the spread of glyphosate-resistant Amaranthus tuberculatus

J. M. Kreiner*, D. A. Giacomini, F. Bemm, B. Waithaka, J. Regalado, C. Lanz, J. Hildebrandt, P. Sikkema, P. Tranel, D. Weigel, J. Stinchcombe, S. Wright

Abstract: The selection pressure exerted by herbicides has led to the repeated evolution of resistance in weeds. The evolution of herbicide resistance on contemporary timescales provides an outstanding opportunity to investigate key open questions about the genetics of adaptation, in particular the relative importance of adaptation from new mutations, standing genetic variation, and geographic spread of adaptive alleles through gene flow. Glyphosate-resistant *Amaranthus tuberculatus* poses one of the most significant threats to crop yields in the midwestern United States, with both agricultural populations and resistance only recently emerging in Canada. To understand the evolutionary mechanisms driving the spread of resistance, we sequenced and assembled the *A. tuberculatus* genome and investigated the origins and population genomics of 163 resequenced glyphosate-resistant and susceptible individuals in Canada and the USA. In Canada, we discovered multiple modes of convergent evolution: in one locality, resistance appears to have evolved through introductions of preadapted US genotypes, while in another, there is evidence for the independent evolution of resistance on genomic backgrounds that are historically non-agricultural. Moreover, resistance on these local, non-agricultural backgrounds appears to have occurred predominantly through the partial sweep of a single amplification haplotype. In contrast, US genotypes and those in Canada introduced from the US show multiple amplification haplotypes segregating both between and within populations. Therefore, while the remarkable diversity of *A. tuberculatus* has facilitated geographic parallel adaptation of glyphosate resistance, different timescales of selection have favored either adaptation from standing variation or *de novo* mutation in certain parts of the range.

Novel insights into evolutionary genetics from emerging technologies
SMBE-WFA-006
Copy number variants are major drivers of evolutionary dynamics and diversity
S. Lauer [1,*], G. Avecilla, P. Spealman, D. Gresham
[1]New York University Langone Medical Center, New York, United States

Abstract: Copy number variants (CNVs) are an important, but understudied source of genetic variation. CNVs drive adaptive evolution in diverse scenarios ranging from niche specialization to speciation and tumor evolution. In microbial populations, CNVs containing genes that encode high-affinity nutrient transporters are beneficial and undergo positive selection during nutrient limitation. However, the dynamics with which these CNVs are generated and selected are poorly understood as existing methods are not sensitive enough to detect CNVs at low population frequency. To overcome this challenge, we developed a fluorescent reporter assay that allows detection of CNVs as they arise *de novo* in evolving populations. We performed experimental evolution in *Saccharomyces cerevisiae* limited for various nitrogen and carbon sources and tracked duplication and deletion of the general amino acid permease, *GAP1,* with single cell resolution. We found that early duplication events are highly reproducible in replicate evolving populations, but later dynamics are complex and CNVs exhibit variable longterm fates. Using a barcoding approach and fluorescence activated cell sorting, we quantified the extent of clonal interference and demonstrated that many independent CNV lineages arise, compete and replace one another during adaptive evolution. Using molecular methods and genome sequencing, we identified a diverse array of *GAP1* CNV alleles including chromosomal aneuploidies, nonreciprocal translocations, tandem duplications, and complex CNVs, some of which may be generated through neo-chromosome formation. Together, our results show that CNVs are generated repeatedly by diverse processes, resulting in predictable dynamics, but that long-term fates of CNV-containing lineages are shaped by clonal interference.

The evolution of senescence: from theory to molecular data

SMBE-WFA-007

Limiting the damage: longitudinal comparative transcriptomics reveals novel mechanisms underlying extended healthspan in mammals

Z. Huang [1,*], C. V. Whelan [1], N. M. Foley [1], D. Jebb [1], F. Touzalin [1], E. J. Petit [2], S. J. Puechmaille [3], E. C. Teeling [1]

[1]School of Biology and Environmental Science, University College Dublin, Dublin, Ireland, [2]Ecology and Ecosystem Health, National Institute for Agricultural Research, Rennes, [3]Univ Montpellier, Montpellier, France

Abstract: Bats are the longest-lived mammals given their body size and high metabolic rate. However, the underlying molecular mechanisms of their extended healthspans are poorly understood. To address this question we carried out an eight-year longitudinal study of ageing in long-lived bats (*Myotis myotis*). We deep sequenced ~1.7 trillion base pairs of RNA from 150 blood samples collected from known aged bats to ascertain the age-related transcriptomic shifts and the possible miRNA-directed regulation that occurred. We also compared ageing transcriptomic profiles between bats and other mammals by analyzing 298 longitudinal RNA-Seq datasets. Bats did not show the same transcriptomic changes with age as commonly observed in humans and other mammals, but rather exhibited a unique, age-related gene expression pattern associated with DNA repair, cell cycle regulation, immunity and tumor suppression that may drive their extended healthspans. We show that bats have naturally evolved transcriptomic signatures that are known to extend lifespan in model organisms and identify novel genes not yet implicated in healthy ageing. We further show that bats' longevity profiles are partially regulated by miRNA, thus providing novel regulatory targets and pathways for future ageing intervention studies. These results further disentangle the ageing process by highlighting which ageing pathways contribute most to healthy ageing in mammals.

Genetic conflicts in molecular evolution

SMBE-WFA-008

Balanced under the arms race: selfish Segregation Distorter chromosomes and their suppressors in Drosophila

C.-H. Chang [1,*], D. Pascua [1], T. Mouton [1], A. M. Larracuente [1]

[1]Biology, U of Rochester, Rochester, United States

Abstract: The autosomal *Segregation Distorter* (*SD*) chromosome is a selfish coadapted gene complex that can bias its transmission by killing other sperm in spermatogenesis. *SD* chromosomes segregate at low frequencies of 1-5% in natural populations worldwide. Our previous work showed that *SD* chromosomes appear to be in a dynamic equilibrium: independent *SD* haplotypes experienced recent selective sweeps in European and African populations, suggesting that *SD* chromosomes can replace each other over short time scales. In contrast to the sweep patterns on other continents, we discovered that North American populations maintain a high diversity of chromosomal inversions on *SD* chromosomes—in 14 *SD* chromosomes, we find four polymorphic inversions, including two that we describe for the first time. To understand why these inversions are balanced in these populations, we survey genetic elements that interact with *SD*. By crossing a marked *SD* chromosome to 87 inbred lines, we found that suppressors are widespread (74%) on X and autosomes in a population from North Carolina (DGRP). Using a GWAS approach, we determined that multiple genetic modifiers affect the driving ability of *SD* chromosomes in this population. Two GWAS peaks are located in regions with known enhancers of *SD*. Interestingly, we found that a strong X-linked suppressor has distinct effects on *SD* chromosomes that are from the same population, but bear different inversions. Our recombination mapping reveals that a single major locus contributes to the suppressing effect of this X chromosome. We conclude that the high frequency of suppressors in North America likely contributes to a low frequency of *SD* chromosomes. Our results suggest that multiple haplotypes can be maintained under the arms race between *SD* chromosomes and their suppressors in a population. *SD* chromosomes may acquire genetic modifiers from standing variation through chromosomal inversions to escape suppressors.

Oral Abstracts

Barriers and drivers of evolutionary innovation by horizontal gene transfer
SMBE-OR-052
Why so many different types of mobile genetic elements?
E. P. C. Rocha [1,*]
[1]Microbial Evolutionary Genomics, Pasteur Institute, Paris, France

Abstract: Horizontal gene transfer driven by phages or conjugative elements allows the acquisition of complex adaptive traits and their transmission to subsequent generations. This speeds up evolutionary processes as exemplified by the acquisition of virulence traits in emerging infectious agents and by antibiotic resistance in many human pathogens. I'll describe how differences between mobile genetic elements in terms of their mechanism of transmission between cells and mechanisms of stabilization within cells result in diverse co-evolutionary dynamics. For example, plasmids are more plastic than integrative elements, but their ability to replicate autonomously comes at the cost of narrower host ranges. Such trade-offs contribute to explain the diversity of mobile genetic elements. They also implicate that certain types of traits are more likely to be carried by elements that use certain types of molecular mechanisms.

Barriers and drivers of evolutionary innovation by horizontal gene transfer

SMBE-OR-053

The role of horizontal DNA transfer in the evolutionary dynamics of antibiotic resistance in Streptococcus pneumoniae

S. Lehtinen [1,*], C. Chewapreecha [2], J. Lees [3], W. Hanage [4], M. Lipsitch [4], N. Croucher [5], S. Bentley [2], P. Turner [1], C. Fraser [1], R. Mostowy [6]

[1]University of Oxford, Oxford, [2]Wellcome Sanger Institute, Hoxton, United Kingdom, [3]New York University, New York, [4]Harvard Universtiy, Boston, United States, [5]Imperial College London, London, United Kingdom, [6]Jagiellonian Univeristy, Krakow, Poland

Abstract: The horizontal transfer of DNA allows evolutionary innovation to spread onto new genetic backgrounds. The extent to which the frequency of such transfer events impacts evolutionary outcomes is an important and open question. One hypothesis, captured by Beijerinck's dictum "everything is everywhere, but the environment selects," is that evolution is essentially the deterministic outcome of competition between all possible allele combinations. An alternative hypothesis is that, in practice, competition is limited to the pool of available variants and this pool is constrained by the rate at which horizontal DNA transfer allows genes to move between bacterial lineages.

We address this question in the context of antibiotic resistance evolution in *Streptococcus pneumoniae*. The two views of evolution lead to different interpretations of observed patterns of resistance, in particular, of the heterogeneous distribution of resistance determinants among pneumococcal lineages. The 'everything is everywhere' perspective explains the distribution of resistance genes on lineages as the result of variation in the fitness benefit resistance confers on different lineages. This view is supported by theoretical work suggesting that the fitness benefit gained from resistance depends on a lineage's duration of carriage, and an observed correlation between pneumococcal serotypes' duration of carriage and resistance frequency. Conversely, the 'gene transfer as limiting factor' hypothesis posits that different lineages have different resistance levels because they are able to acquire resistance determinants at different rates. This view is supported by the observation that highly recombinogenic pneumococcal lineages also have higher resistance frequencies. In addition, an association has also been reported between pneumococcal duration of carriage and recombination rate, raising the possibility that either of the other two associations (resistance vs duration of carriage, resistance vs recombination rate) could be confounded.

Using genetic and epidemiological data from the largest pneumococcal carriage study to date (1086 carriage episodes from infants in Maela, a refugee camp on the Thailand-Myanmar border), we revisit these associations in a unified analytical framework allowing adjusting for potential confounding. We find strong support for an association between duration of carriage and antibiotic resistance (Kendall's rank correlation coefficient: 0.41, 95%CI [0.16,0.60]), which is robust to adjusting for the rate of horizontal DNA transfer. We do not find strong evidence for an association between resistance frequency and horizontal DNA transfer rate. These results suggest that in this setting, the rate of horizontal DNA transfer does not constrain evolution of antibiotic resistance in the pneumococcus, thus lending support to the 'everything is everywhere' eco-evolutionary perspective.

Barriers and drivers of evolutionary innovation by horizontal gene transfer

SMBE-OR-054

Highways of recombination and heterogeneous patterns of donor-recipient relationships in bacterial species

C. P. Andam[*], C. Park[1]

[1]Department of Molecular, Cellular and Biomedical Sciences, University of New Hampshire, Durham, United States

Abstract: Phenotypic and genetic variation in microbes can take tremendously complex forms even within a single species. An important process that generates this variation is recombination, defined as the re-assortment of DNA between strains of different genomic backgrounds. The rate of recombination of a species is important for estimates of mutation and genomic change, and therefore the capability of a species to respond and adapt to selective pressures. Current models of microbial recombination incorporate the null expectation that recombination is a homogeneous process across a species, whereby different lineages of the same species and different genes within a genome exhibit the same rates and patterns of DNA donation and receipt. We aim to elucidate the extent in which variation in recombination exists within a microbial species and the factors that drive this variation. We analyzed the frequency and characteristics of genome-wide recombination in Salmonella enterica, Staphylococcus aureus, Streptococcus pneumoniae, Klebsiella pneumoniae and Yersinia enterocolitica. We show that heterogeneity and biases in recombination exist among members of the same species or subspecies, with some pairs of strains linked by highways of recombination. Some strains exhibit significantly higher frequencies of DNA donation or receipt, and may vary depending on whether recombination occurs in core genes or accessory genes. We show that serotype, ecology (host or habitat) and geographical proximity can influence the formation of these highways. These hyper-recombinant strains are likely to act as hubs of gene flow, facilitating the rapid spread of certain genes (e.g., antibiotic resistance, metabolic genes, niche-specific genes). In conclusion, these results demonstrate that recombination in microbial populations and species is a heterogeneous process. Our findings provide valuable insight in developing a coherent model for genome evolution that integrates variation in recombination within and among microbial species as well as in our understanding of why microbes have pan-genomes.

Barriers and drivers of evolutionary innovation by horizontal gene transfer

SMBE-OR-055

Being a plasmid: consideration of copy number, DNA topology and segregation in plasmid genome evolution

T. Dagan*

Abstract: Plasmids are genetic elements that colonize and replicate in prokaryotic cells. Plasmids play a major role in prokaryote ecology and evolution, as they are ubiquitous in all phyla and across all habitats. Plasmids are distinct from chromosomes in several properties, including a relatively small genome size, their ploidy level that can be dynamic, their frequency in the population that can be spatially and temporally heterogeneous, and their mobility. These properties correspond to known determinants of molecular evolution such as mutational supply, segregational drift, population size, and lateral transfer. We studied the fundamental principles that govern plasmid genome evolution using experimental evolution of model plasmids as well as simulated evolution. Our results reveal that segregational drift of multicopy plasmids interferes with the retention and fixation of novel plasmid variants. Depending on the selection pressure on newly emerging variants, plasmid genomes may evolve slower than haploid chromosomes, regardless of their higher mutational supply. Our results further reveal dependency of plasmid stability evolution on the coordination of plasmid transcription and replication, which is challenging in small genomes. Our studies thus reveal plasmid properties that are important for the evolution of successful autonomously replicating genetic elements.

Barriers and drivers of evolutionary innovation by horizontal gene transfer

SMBE-OR-056

Genetic Dominance Shapes Plasmid-Mediated Evolution In Bacteria

J. Rodríguez-Beltran [1,*], V. Sørum [2], A. San Millan [1]

[1]Department of Microbiology, University Hospital Ramón y Cajal (IRYCIS), Madrid, Spain, [2]Department of Pharmacy, Artic University of Norway, Tromsø, Norway

Abstract: Plasmids are key drivers of bacterial evolution by acting as vehicles for horizontal gene transfer. Plasmids encode accessory genes such as antibiotic resistance genes that are often adaptive to their bacterial host. In contrast to the bacterial chromosome, plasmids are typically maintained at several copies per cell, thus providing an island of polyploidy in an otherwise haploid bacterial genome. We hypothesized that if new adaptive mutations are dominant, plasmids, with greater mutational target size compared to chromosomes, will be more likely to acquire beneficial mutations, and will have greater evolutionary potential. Alternatively, if beneficial mutations are mostly recessive, then chromosomes will have access to new beneficial mutations that have no selective benefit as heterozygotes in plasmids. Here, we experimentally test this possibility by comparing the evolution of antibiotic resistance genes in chromosomes and plasmids. Our results show that genetic dominance largely determines the chances of plasmid beneficial mutations to reach fixation, as recessive mutations have no effect on bacterial phenotype. Thus, plasmid-encoded traits evolve mostly through dominant mutations, whereas chromosomal genes have access to both dominant and recessive mutations. We experimentally showed that gene location (plasmid or chromosome) and genetic dominance strongly determine the evolutionary pathways followed by bacteria. More generally, we showed that genetic dominance dictates the distribution of genes on natural plasmids. Collectively, our results shed light on the forces that shape plasmid-mediated evolution in bacteria and contribute to explain why certain genes, but not others, readily spread on plasmids.

Barriers and drivers of evolutionary innovation by horizontal gene transfer

SMBE-OR-057

Exploring the parasitism-mutualism continuum within microbial comunities - Beneficial plasmids can increase community diversity

U. Klümper [1,*], B. Dawson [1], A. Newbury [1], E. Hesse [1], M. Castledine [1], A. Buckling [1]

[1]CLES, University of Exeter, Penryn, United Kingdom

Abstract: A main factor in plasmid ecology and evolution is host-plasmid symbiosis, where their relationship can change from parasitic to mutualistic depending on environmental conditions and plasmid encoded traits. Host-plasmid co-evolution and fitness effects have been extensively studied in single strains. However, host-plasmid and host-host interactions within more complex communities remain underexplored. Here, we evaluated how an invading plasmid affects community diversity, depending on the nature of host-plasmid interactions ranging from parasitic to mutualistic.

In evolution experiments we invaded artificial communities of 5 mainly antagonistic but stably co-existing environmental isolates with conjugative plasmid pKJK5, conferring resistance to tetracycline. To change the nature of the plasmid-host interactions from parasitic to mutualistic, sub-inhibitory concentrations of tetracycline were added to half the reactors, thus turning the initial metabolic cost of plasmid carriage into a fitness benefit. To account for potential donor effects, plasmids were added to the community using a multifactorial design with each of the 5 strains serving as plasmid donor in individual experiments. No plasmid controls were run for both the antibiotic and the non-antibiotic treatment. After 6 weeks diversity of the evolved communities was assessed.

In the absence of plasmids, antibiotics at sub-inhibitory concentrations decreased community diversity when compared to the non-antibiotic control. For all plasmid treatments results were donor independent. Under non-antibiotic conditions invasion of a parasitic plasmid had no effect on diversity compared to the non-plasmid control. However, adding a mutualistic plasmid in the presence of antibiotics significantly increased community diversity compared to any control treatment.

By adding a mutualistic, communally shared plasmid previous antagonistic relationships between strains within the community were reduced, thus increasing diversity.

Biochemistry, epistasis and the evolutionary process
SMBE-OR-155B
COLLATERAL FITNESS EFFECTS OF MUTATIONS
M. Ostermeier [1,*]
[1]Johns Hopkins University, Baltimore, United States

Are you a member of SMBE?: No
Poster Submission: Protein epistasis can result from the non-additive effects of mutations on the protein's biochemical properties (e.g. thermostability and specific activity) or the non-linear mapping of those properties on organismal fitness. Such effects derive from changes in the ability of the protein to perform its physiological function. Here, we reveal the importance of a mutation's collateral fitness effects, which we define as effects that do not derive from changes in the protein's ability to perform its physiological function. We comprehensively measured the collateral fitness effects of missense mutations in the E. coli beta-lactamase antibiotic resistance gene TEM-1 using growth competition experiments in the absence of antibiotic. At least 42% of missense mutations in TEM-1 were deleterious with a selection coefficient less than -0.01. Deleterious mutations caused improper post-translational processing, incorrect disulfide-bond formation, protein aggregation, changes in gene expression, and other phenotypic changes. The surprising prevalence of collateral fitness effects may explain why protein evolution rates are better predicted by protein expression level than by protein essentiality (i.e. the E-R anticorrelation).

Biochemistry, epistasis and the evolutionary process

SMBE-OR-154

Cryptic genetic variation accelerates adaptive evolution by opening alternative paths to diverse adaptive peaks

J. Zheng [1 2,*], J. Payne [2 3], A. Wagner [1 2 4]

[1]Department of Evolutionary Biology and Environmental Studies, University of Zurich, Zurich, [2]Swiss Institute of Bioinformatics, Lausanne, [3]Institute of Integrative Biology, ETH Zurich, Zurich, Switzerland, [4]The Santa Fe Institute, Santa Fe, United States

Abstract: Functional protein sequences are very sparse in protein sequence space. In addition, protein fitness landscapes are also highly rugged, a phenomenon that renders most mutational paths to functional proteins inaccessible. How populations of evolving proteins break such constraints to explore protein sequence space and evolve new functions is a central question in evolutionary biology. Here we used directed evolution of fluorescent proteins to demonstrate that cryptic genetic variation, the standing genetic variation that is not normally expressed, but that brings forth phenotypic variation after environmental change or genetic perturbation, can enable evolving populations to alleviate this problem. Such populations accumulate mostly neutral or deleterious mutations that later help by-pass or traverse selectively inaccessible paths to new peaks in a fitness landscape that harbor proteins with new phenotypes. By opening diverse adaptive peaks for exploration, cryptic variation also provides access to otherwise inaccessible regions of protein sequence space. Our observations demonstrate that cryptic variation can help populations quickly respond to environmental changes, and also suggest that cryptic variation can promote the evolution of diverse proteins.

Biochemistry, epistasis and the evolutionary process

SMBE-OR-155

The evolution of position-dependent codon usage and the biochemical drivers of emergent epistatic interactions

A. I. Teufel[1 2,*], N. T. Marrow[3], A. Diament[4], T. Tuller[4], C. O. Wilke[2]

[1]Santa Fe Institute, Santa Fe, [2]Department of Integrative Biology, [3]Department of Physics, The University of Texas at Austin, Austin, United States, [4]Department of Biomedical Engineering, Tel Aviv University, Tel Aviv, Israel

Abstract: The speed at which mRNA sequences can be translated is related to the sequence's codon composition. Codons that correspond to abundant tRNAs translate quickly and codons that correspond to less abundant tRNA translate slowly. While the positions of codons along an mRNA sequence impact the speed of the ribosome translating the sequence, this position dependence can also influence the appearance of ribosomal traffic jams, where ribosomes collide and form queues. To examine how position-dependent codon usage evolves to maximize translation efficiency and avoid traffic jams, we simulate the evolution of a network of mRNA sequences with the use of a ribosome flow model. We find that the evolution of position-dependent codon usage is reliant on a number of factors including, sequence length, ribosome availability, and relative differences in codon speed. We also investigate the role of epistasis in position-dependent codon evolution. Numerous studies have focused on epistasis in the evolution of proteins, and a core concept in these works is that substitutions that were neutral or nearly neutral at the time of fixation become entrenched, and the probability of reverting back to the substitution's predecessor diminishes as additional mutations accumulate. We apply this same concept to the evolution of mRNA sequences by examining the probability of fixation of reversion mutations in our simulated data. We observe epistatic dynamics similar to those of proteins and find that factors such as ribosome availability and relative differences in codon speed impact the emergence of epistasis in mRNA sequences.

Biochemistry, epistasis and the evolutionary process

SMBE-OR-155A

ORIGINS OF HEMOGLOBIN: MECHANISMS FOR THE HISTORICAL EVOLUTION OF A NEW MOLECULAR COMPLEX

J. Thornton [1,*], A. Pilai [1], G. Hochberg [1]

[1]University of Chicago, Chicago, United States

Are you a member of SMBE?: No

Poster Submission: Virtually all proteins assemble into multiprotein complexes, but we have no detailed knowledge of how multimers and their associated functions arose during evolution. Here we dissect the mechanisms that drove the historical evolution of hemoglobin (Hb), arguably the best-studied multimer in all of biochemistry. Hb, a heterotetramer of two α and two β subunits, mediates oxygen transport in vertebrates by cooperatively binding and releasing oxygen. Using ancestral reconstruction and biochemical experiments, we show that Hb evolved from a monomeric common ancestor with other globins. We identified the missing-link intermediate in Hb evolution, a noncooperative homodimer that existed before the gene duplication that generated separate Hb □ and □ subunits; selection for Hb's distinct physiological functions therefore could not have driven early steps required for the evolution of its ultimate structure. The genetic basis for the evolution of Hb from this dimeric intermediate was surprisingly simple: four historical substitutions, clustered on a small region of □'s surface, were sufficient to trigger acquisition of a new interface and assembly into a tetramer. By causing a new intersubunit interface to form, these substitutions also altered the complex's oxygen-binding function, conferring the foundations of cooperativity; this was possible because an ancient conformational linkage already existed, apparently fortuitously, between oxygen-binding at the globin active site and the surface where the new interface happened to evolve. Our observations reveal how easily evolution can produce new multimeric complexes and confer new functions, because just a few mutations can recruit existing surfaces and structural features into higher-level architectures with new conformational properties.

Biochemistry, epistasis and the evolutionary process
SMBE-OR-156
Super-Darwinian exploration of experimental fitness landscapes
S. Sinai [1 2 3,*], G. M. Church [2 4], M. A. Nowak [1 5], E. Kelsic [2 3]
[1]OEB, Harvard University, Cambridge, [2]Wyss Institute, Harvard Medical School, Boston, [3]Dyno Therapeutics, Cambridge, [4]Genetics, Harvard Medical School, Boston, [5]Mathematics, Harvard University, Cambridge, United States

Abstract: Modern biological techniques, such as array-based DNA synthesis and deep sequencing, now allow for high-throughput parallel search of fitness landscapes. Nonetheless, for sequences of typical biological size, it's infeasible to sample even a fraction of the sequence space. Given a limited number of experimental throughput and rounds, we would like an algorithm that finds top performing sequences with high probability. Choosing candidate sequences completely at random results in many sequences without functionality. Directed evolution is a popular tool for identifying improved sequences, but since randomly generated diversity has a low chance of generating optimal improvements, there is room for progress. Finally, epistasis futher complicates the process of optimization. Thus our goal is to identify empirical exploration strategies that can search the landscape more intelligently.

In this work we combine (i) scanning mutagenesis, where we can pick specific mutants with (ii) machine learning models that learn local (non-linear) sequence-to-function maps. We then utilize this tool to (iii) construct an exploration strategy that searches the landscape in batches for improved variants. We first demonstrate one iteration of this cycle by performing targeted mutagenesis on a 28-aa window of an Adeno-associated Virus (AAV) capsid protein, generating ~8e4 mutants, thereby improving the packaging efficiency over wild-type for variants as far as 13 edits. We then evaluate the exploration strategy for many iterations in simulated landscapes. Encouragingly, we show that the Boltzmann-Gumbel Explorer we propose outperforms standard evolutionary search and directed evolution, even with an imperfect local model of the landscape.

Biochemistry, epistasis and the evolutionary process

SMBE-OR-157

Experimental test of a universal biochemical mechanism of high-order epistasis in macromolecules

A. J. Morrison [1,*], M. Harms [1]

[1]Chemistry, University of Oregon, Eugene, United States

Abstract: High-order epistasis is ubiquitous in genotype-phenotype maps. It can make evolution deeply unpredictable as past mutations influence the effects of future mutations. The mechanisms that lead to high-order epistasis are unclear. Its pervasiveness suggests that it may arise from features shared across biological systems. In previous theoretical work (Sailer and Harms, (2017) PNAS 114 (45) 11938-11934), we hypothesized that high-order epistasis could arise for individual proteins that populate more than one conformation. Such conformational "ensembles" are fundamental thermodynamic properties of macromolecules that are intrinsically tied to function. A mutation affects all conformations of the ensemble, leading to nonlinear, non-additive effects on phenotype. We set out to experimentally probe the relationship between high-order epistasis and the molecular ensemble using a 4-site binary genotype-phenotype map between the wild-type lac repressor and a 4-way mutant. We can modulate how many conformations are in the ensemble by adding molecules that preferentially bind to one conformation of the lac repressor, shifting from a multi-conformation ensemble to an ensemble with a single dominant conformation. This allows us to toggle between conditions where we expect to see ensemble-induced high-order epistasis and conditions where we expect very little high-order epistasis. We found that the contribution of high-order epistasis to variation in phenotype drops from 19% to 5.4% when we shift to a single-conformation ensemble. Our data suggests that we can indeed modulate high-order epistasis by manipulating the ensemble. Further, the mathematical framework of statistical thermodynamics allows us to understand and decompose the effects of each mutation on the conformations of the ensemble. As a result, we may be able to treat mutations as additive in the thermodynamic context, and then calculate the expected high-order epistasis on observable features of the protein. By accounting for a fundamental biophysical property of macromolecules, we can both improve our predictions of evolutionary trajectories and provide insight into how chemistry shapes evolution.

Contemporary Evolution

SMBE-OR-142

Ecological genetics of recurrent evolution of (industrial) melanism in moths

I. Saccheri [1,*]

[1]Institute of Integrative Biology, University of Liverpool, Liverpool, United Kingdom

Abstract: The increased frequency of dark forms of moths in the heavily polluted industrial centres of 19th and 20th century Britain remains one of the more conspicuous examples of rapid adaptation to environmental change. The peppered moth has been the most intensively studied of these species and provides a rich testing ground for several ideas in ecological genetics. I will use our work on the peppered moth system and two other moths to consider the general problem of the genetic and environmental conditions under which contemporary evolution is likely to depend on de novo mutation vs ancestral alleles. The celebrated British carbonaria case is clearly due to the spread of a transposable element inserted into the gene cortex ca.1820. We also find that melanism consistently maps to cortex, in other regional populations of the peppered moth, and other moth species. This implies surprisingly strong canalisation of the genetic architecture of melanism in the Lepidoptera, perhaps coupled with relatively high mutation rate. However, patterns of linked diversity suggest older and more complex sources of exaptive alleles than represented by British carbonaria, potentially maintained through environmental heterogeneity.

Contemporary Evolution

SMBE-OR-146

Plastic and evolutionary processes underpinning contemporary evolution

M. Cuenca Cambronero [1,*], T. Nguyen [2], H. Marshall [2], N. Eastwood [2], H. Tomero, L. Orsini [2]

[1]Fish Ecology and Evolution, EAWAG, Kastanienbaum, Switzerland, [2]Environmental Genomics, University of Birmingham, Birmingham, United Kingdom

Abstract: Anthropogenic activities, such as land use and climate change, impose selection pressure on biodiversity at an unprecedented rate and intensity. This pressure may lead to extinction, alteration of species distributional ranges and disturbance in species-species interactions. However, species that persist despite this severe pressure, provide formidable models to study the mechanisms that underpin contemporary evolution.

Here, we use the ecological model species *Daphnia* to study phenotypic plasticity and genetically-based adaptation to anthropogenic activities. We are able to distinguish between these two mechanisms by using the practice of "resurrection ecology", by which historical and modern populations originating from the same genetic pool can be studied in the same experimental settings. *Daphnia*'s life cycle alternates sexual recombination with asexual (clonal) reproduction, providing the opportunity to reveal the relative contribution of plastic and genetic responses to multiple environmental factors through evolutionary time. We study the impact of eutrophication and pesticides, in isolation and combination, on life history fitness-linked life history traits. We show synergism between chemical pollution and other stressors, revealing that the use of single stressors essays may lead to biased estimates of species persistence. We also show that a complex interaction between plastic and genetic adaptation underpins contemporary evolution to anthropogenic stress in this species.

Contemporary Evolution

SMBE-OR-145

Rapid geographic expansion and evolution of the fungal maize pathogen Setosphaeria turcica in Europe

K. Schmid [1,*], M. Vidal Villarejo, F. Freund

[1]Institute of Plant Breeding, Seed Science and Population Genetics, University of Hohenheim, Stuttgart, Germany

Abstract: Modern agricultural systems are characterized by a small number of crops that are cultivated in large monocultures and simplified crop rotations. The employment of monogenic resistances and chemical plant protection in such environments exerts strong selection for resistance-breaking genotypes in agricultural pests. In Central Europe, the cultivation area of maize has expanded rapidly in recent decades and this expansion was accompanied by Northern Corn Leaf Blight. It is a major disease of maize and caused by the fungal pathogen *Setospherica turcica*. We used whole genome resequencing of 140 *S. turcica* isolates from Africa and Europe to analyse genomic footprints of expansion and selection. Coalescent modeling revealed a rapid spread of asexual lineages that originated several hundred years ago by sexual recombination and a very rapid population size increase over the last two decades. To test whether the spread of predominant asexual lineages resulted from strong selection of resistant races, we compared a neutral (bifurcating) with a sweepstake reproduction (multiple merger) coalescent model using Approximate Bayesian Computation (ABC). The ABC analysis provided strong support for a neutral demography with exponential growth dynamics. Race typing data of sequenced isolates indicated a rapid evolution of new resistance genes against the main monogenic resistance genes employed by maize breeders. K-mer analyses identified candidate regions of rapid evolution that also included effector genes. This work provides perspectives for pathogen management in modern agroecosystems and resistance breeding that is based on explicit population genetic modeling facilitated by DNA sequencing.

Contemporary Evolution
SMBE-OR-143
Chromosomal inversions and environmental adaptation in malaria mosquitoes
N. J. Besansky [1,*]
[1]University of Notre Dame, Notre Dame, United States

Abstract: The Afrotropical malaria vectorial system, responsible for >90% of disease transmission and deaths, is the most powerful available to Plasmodium parasites. Its most important component is the Anopheles gambiae complex, a group of at least eight cryptic mosquito species that radiated recently and rapidly. Although morphologically indistinguishable, profound physiological and behavioural differences affect their distribution with respect to humans and the environment, as well as their role in malaria transmission. Chromosomal inversions have been an important guide to understanding this group, both in terms of interspecific fixed inversion differences once used for cytotaxonomic identification, and contrasting intraspecific levels of inversion polymorphisms. The most chromosomally polymorphic species are those most geographically widespread, best able to adapt to anthropogenic modifications and environmental heterogeneities, and unfortunately also, those with the highest malaria vectorial capacity. Patterns of inversion polymorphism are non-random with respect to seasonal and spatial environmental heterogeneities, suggesting that inversions are the targets of spatially varying selection. Laboratory and field studies have implicated larval thermotolerance and adult resistance to aridity as adaptive phenotypes conferred by the most widespread inversions on chromosome 2. The application of modern genomic approaches to unraveling inversion history and underlying mechanisms are beginning to bear fruit, though much work remains.

Contemporary Evolution

SMBE-OR-144

The role of genome structural variation on plastic and constitutive phenotypic divergence in multifarious environments

A. Chaturvedi [1], J. Zhou [2], C. Jackson [3], B. Peter [4], O. S. Solari [5], V. Dhandapani [6], J. B. Brown [7], L. De Meester [8], L. Orsini[*]

[1]Department of Ecology and Evolution, Universite de Lausanne, Lausanne, Switzerland, [2]School of Computer Science, University of Birmingham, Birmingham, United Kingdom, [3]School of Public and Environmental Affairs, Indian University Bloomington, Bloomington, IN, United States, [4]Department of Evolutionary Genetics, Max Planck Institute for Evolutionary Anthropology, Leipzig, Germany, [5]Department of Molecular Ecosystems Biology, Lawrence Berkeley National Laboratory, Berkeley, United States, [6]School of Biosciences, University of Birmingham, Birmingham, United Kingdom, [7]Statistics, University of California, Berkeley, Berkeley, United States, [8]Laboratory of Aquatic Ecology, Evolution and Conservation, KULeuven, Leuven, Belgium

Abstract: Rapid environmental change, associated with anthropogenic activities has been shown to induce evolution on contemporary time scales. However, examples of contemporary evolution are mostly limited to model species which can be reared in the laboratory, whereas it is not well understood in nature. This is because Lack the evolutionary processes underpinning contemporary evolution requires the ability to distinguish phenotypic plasticity from genetically-based adaptation as well as to establish a clear link between evolving phenotypes and the underlying genetics.

We apply high throughput sequencing and phenotyping to dense geographic samples of the ecological model species and keystone aquatic crustacean *Daphnia magna* to investigate mechanisms of adaptation to multifarious environments. We measure fitness-linked traits in control and stressful conditions, as well as genome-wide polymorphism and structural variation on 19 populations distributed along three orthogonal gradients of selection, including land use, predation and parasitism. We identify genome variants underpinning plastic and constitutive phenotypic responses to each of the stressors in the landscape as well as to multiple stressors. Genes and pathways shared among populations experiencing the same stressor underpin local adaptation, whereas a very small proportion of alleles is shared among these populations. *Daphnia* shows a high number of structural variants linked to stress response. The high number of structural variants and repeatable patterns of standing genetic variation at gene and pathway, rather than allele level indicate that selection operates at high hierarchical levels in this species, enabling a fast adjustment to rapid environmental changes. Our findings have important implications for the impact of purifying selection on structural genome variants and their role in adaptive responses to anthropogenic stress. Further, they provide a conceptual framework of how genes might interact with the environment and evolve toward the development of plastic traits. Finally, our findings suggest that the effective population size of this species may be smaller than allele frequency estimates suggest as the target of natural selection are gene networks rather than single alleles.

Contemporary Evolution

SMBE-OR-147

Evidence that viruses, particularly SIV, drove genetic adaptation in natural populations of eastern chimpanzees

J. M. Schmidt [1,*], M. de Manuel [2], T. Marques-Bonet [2 3 4], S. Castellano [5 6], A. Andrés [1]

[1]UCL Genetics Institute/ Department of Genetics, Evolution and Environment , University College London, London, United Kingdom, [2]Institut de Biologia Evolutiva , Consejo Superior de Investigaciones Científicas–Universitat Pompeu Fabra, [3]National Centre for Genomic Analysis–Centre for Genomic Regulation, Barcelona Institute of Science and Technology, [4]Institució Catalana de Recerca i Estudis Avançats (ICREA), Barcelona, Spain, [5]Genetics and Genomic Medicine Programme, Great Ormond Street Institute of Child Health, [6]UCL Genomics, University College London, London, United Kingdom

Abstract: Viruses are a predominant factor behind recent and ongoing selection in mammalian genomes. Studying the effects such selection has had on chimpanzee genomes can provide valuable insights into how pathogens are affecting an endangered primate species. As there are many notable examples of cross species transmission between primates and humans - including the HIV/AIDS pandemic - these studies also have the potential to illuminate mechanisms of adaption to novel pathogens of medical and economic importance. By investigating patterns of genetic differentiation amongst the four chimpanzee sub-species, we show that highly differentiated SNPs in eastern chimpanzees are uniquely enriched in genic sites in a way that is expected under recent adaptation. These sites are enriched for genes that differentiate the immune response to infection by simian immunodeficiency virus (SIV) in natural vs. non-natural host species. Conversely, central chimpanzees exhibit selective sweeps at three cytokine receptors – paralogs of *CCR5* and *CXCR4,* the two major receptors utilized by HIV to enter human cells. Thus, we infer that SIV may be eliciting distinctive adaptive responses in different chimpanzee subspecies, and builds upon a developing literature suggesting that SIV elicits ongoing selection in African primates. Lastly, as central chimpanzee SIV is the source of the global HIV/AIDS pandemic, understanding the mechanisms that limit pathogenicity of SIV in chimpanzees can broaden our understanding of HIV infection in humans.

Evolution Ecology and Host-Virus dynamics in the Microbiome
SMBE-OR-190
Bayesian approaches to modeling complex microbiome dynamics
G. K. Gerber*, T. E. Gibson, R. Creswell, B. B. Hsu

Abstract: The human microbiome is highly dynamic on multiple timescales, changing dramatically during development of the gut in childhood, with diet, or due to medical interventions. However, inferring models from microbiome time-series data presents a number of challenges including high-dimensional but temporally sparse and non-uniformly sampled data; high measurement noise; and, nonlinear and physically non-negative dynamics. I will present Bayesian machine learning methods that we have developed to address these challenges and discuss some of the experimental systems we use for generating sufficiently rich longitudinal data for inference. In particular, I will present MDSINE (Microbial Dynamical Systems INference Engine), our method for efficiently inferring dynamical systems models from microbiome time-series data, which we have applied to developing bacteriotherapies for C. difficile infection and inflammatory bowel disease. I will also discuss our recent extensions to the method using Bayesian nonparametric techniques to enable scaling to large datasets while maintaining interpretability. Our contributions include a new type of dynamical systems model for microbial dynamics based on what we term interaction modules, or learned clusters of latent variables with redundant interaction structure; a fully Bayesian formulation of the stochastic dynamical systems model that propagates measurement and latent state uncertainty throughout the model; and introduction of a temporally varying auxiliary variable technique to enable efficient inference by relaxing the hard non-negativity constraint on states. Time permitting, I will also discuss some of our most recent work characterizing phage-microbiome dynamics in the mammalian gut.

Evolution Ecology and Host-Virus dynamics in the Microbiome

SMBE-OR-192

Ecological and Evolutionary Consequences of Viral Plasticity

M. Choua [1,*], J. A. Bonachela [2]

[1]Mathematics and Statistics - Population Modeling and Epidemiology Group, University of Strathclyde, Glasgow, United Kingdom, [2]Ecology, Evolution, and Natural Resources, Rutgers University, New Brunswick, United States

Abstract: Host-virus interactions can be found at any trophic level, from unicellular organisms, humans, whales, etc. The aim of a viral attack is to hijack the host machinery for virus reproduction. Lab experiments revealed that phage performance (i.e. infection time and offspring number) varies with physiological changes in their hosts, effectively the phage's environment. These variations are referred as viral phenotypic plasticity (i.e. the ability of an individual to vary its phenotype when environmental conditions change). In the past, models studying viral plasticity focused on intracellular dynamics. These latter are too detailed to be included in models that study host-phage interaction in the long term, which hinders our understanding of systems that range from pathogens infecting gut bacteria to marine phage shaping the ocean communities. Here, we compiled data for *Escherichia Coli* to deduce expressions that represent the dependence of lytic T-phage traits on host growth rate. We then included those expressions in a standard host-phage model to understand mechanistically how host physiology affects (i) the evolutionary response of the viral traits and (ii) the population dynamics associated with different scenarios including, e.g. nutrient pulses or host starvation. We show that plasticity on the offspring number drives the eco-evolutionary interactions and reinforce the feedback between host, virus, and environment. Our results reveal the limitation of standard models (i.e. models that neglect plasticity) in realistic scenarios and highlight the importance of viral plasticity to unravel host-phage interactions.

Evolution Ecology and Host-Virus dynamics in the Microbiome
SMBE-OR-241
Evolution and ecological interactions of gut bacterial taxa in challenged ecosystems
F. Hildebrand*, T. I. Gossmann, L. Moitinho-Silva, J. Huerta-Cepas, P. Bork

Abstract: The healthy human adult gut microbiome is taxonomically relatively stable over prolonged periods and the most abundant species are usually already described and cultured. However, abundances can change upon perturbation, such as antibiotics intake. I will present the dynamic changes in the gut microbiome in a long timeseries (>4 years, 24 timepoints), leading to the identification of novel class of Firmicutes that are normally low abundant. Due to their recurrent bloom in response to antibiotic interventions in multiple patients, we could reconstruct the genome of ^{U}*B. ceftriaxensis*. The patient specific reference genomes allowed us to identify the ecological interactions of the novel species with known probiotic bacteria, that led to ecological successions and ultimately the recovery of gut microbial diversity and a stable ecosystem.
To further understand the adaptive bacterial changes in these patients, we developed a novel high-resolution metagenomics approach, that can reliable separate strains with as few as 150 genome-wide nt differences. This allowed us to trace strain transfer between host family members and to show that the rare ^{U}*B. ceftriaxensis* was neutrally or adaptively evolving in the patient during the observation period, while other species appear to be under strong purifying selection. This is also reflected in differences in effective population size, that we can infer for the first time for a natural bacterial population, leading to new applications of population genetics.

Evolution Ecology and Host-Virus dynamics in the Microbiome
SMBE-OR-191
Bacteriophage-bacteria coevolution in the gut microbiota

L. De Sordi [1,*]
[1]Sorbonne Université, Paris, France

Abstract: Viruses that infect bacteria, or bacteriophages, are among the most abundant entities in the gut microbiome. However, the mechanisms by which they infect and persist as unique viral pools in the intestinal tract remain poorly understood.
We used comparative population genomics to study antagonistic co-evolutionary dynamics between networks of bacteria and bacteriophages, both *in vitro* and within the murine gut. We showed that intestinal bacteria are an evolutionary force driving the expansion of the bacteriophage host range by increasing the genetic variability of viruses which, in turn, exert a predatory pressure shaping the diversity of the microbiota.
Our data suggest that the modulation of bacteriophage-bacteria infection networks is relative to the opportunities for coevolution encountered in the intestinal tract, and that multiple predator-prey dynamics are perpetuated and differentiated in parallel, generating and maintaining intestinal microbial diversity and equilibrium.

Evolution Ecology and Host-Virus dynamics in the Microbiome

SMBE-OR-194

Community assembly in the microbiome: ecological insights into infant microbiome development

K. Coyte [1 2,*]

[1]Dept of Zoology, University of Oxford, Oxford , United Kingdom, [2]Boston Children's Hospital, Boston, United States

Abstract: In healthy adults, the gastrointestinal tract harbours a diverse community of microbes that play critical roles in health and wellbeing. However, we are not born with this microbiome. Over the first months and years of life the gut microbiome gradually develops, undergoing a process akin to classic primary succession. We here develop ecological theory to study the factors that drive these assembly processes. We find that interactions between species can enforce order on microbiome development, with interspecies dependencies driving the predictability of succession. We combine this theory with novel sample processing techniques to interrogate both bacterial and fungal microbiome development in premature infants. We utilize machine learning to infer how members of the microbiome are affected by both one another and clinical interventions. Preliminary results identify specific inter-microbial interactions that may in part drive community dynamics, and uncover the impact of antibiotic interventions in perturbing healthy microbiome development. Taken together, these results highlight the importance of disentangling microbial interactions if we are to understand and ultimately manipulate our microbiome communities.

Evolution Ecology and Host-Virus dynamics in the Microbiome

SMBE-OR-195

Dynamic interaction network inference from longitudinal microbiome data

J. Lugo-Martinez [1], D. Ruiz-Perez [2], Z. Bar-Joseph [1], G. Narasimhan [2,*]

[1]Carnegie Mellon University, Pittsburgh, [2]Florida International University, Miami, United States

Abstract: Several studies have focused on the microbiota living in environmental niches including human body sites. In many of these studies researchers collect longitudinal data with the goal of understanding not just the composition of the microbiome but also the interactions between different taxa. However, analysis of such data is challenging and very few methods have been developed to reconstruct dynamic models from time series microbiome data.

We propose a computational pipeline that enables the integration of data across individuals for the reconstruction of such models. Our pipeline starts by interpolating the samples using B-Splines. Then, the interpolated data for all individuals are aligned by warping the time scale of each sample into the scale of another representative sample. This is meant to compensate for the differences in rates with which biological events occur in different individuals or samples. The aligned profiles are then used to learn a Dynamic Bayesian Network, which represents causal temporal relationships between taxa and clinical variables, and are perfect for this task because of its interpretability. We tested our methods on three longitudinal microbiome data sets (infant gut, vagina and oral cavity). Different biological insights were obtained by the models which include several known and novel interactions. The extended CGBayesNets package is available with the documentation together in the journal paper.

Our results provide evidence that microbiome alignments coupled with dynamic Bayesian networks improve predictive performance over previous methods and enhance our ability to infer biological relationships within the microbiome and between taxa and clinical factors.

Evolution of phenotypes: understanding diversity and the role of plasticity in adaptation to new environments
SMBE-OR-071
Impact of the low frequency variants on the phenotypic landscape at population-scale

J. Schacherer*

Abstract: Understanding the rules laying behind the natural phenotypic variation has been a key point of modern genetics for decades. However, it is still difficult to precisely address and dissect the molecular bases underlying complex traits. Today, a better understanding of the genetic architecture of traits requires a precise estimation of genetic components governing phenotypes at a species-wide level. In this context, we took advantage of the large set of 1,011 natural *Saccharomyces cerevisiae* isolates that we completely sequenced. We selected a set of 55 isolates as genetically diverse as possible to generate a diallel cross panel of 3,025 hybrids. These hybrids were then phenotyped on 49 stress related traits resulting in 148,225 cross/trait combinations. The results clearly showed that although phenotypic variance is mostly governed by additivity, about a third of this variance can be explained by non-additive phenomena. This is confirmed by the fact that a majority of complete dominance is observed in 25% of the traits. The dataset we generated also allowed us to perform genome-wide association studies (GWAS) to uncover variants responsible for the tested phenotypes. Interestingly, 2,156 significantly associated variants were found and among them 12% are present in less than 5% of the 1,011 population. It clearly shows that those so-called low frequency variants represent an important source of phenotypic variance and can be mapped using GWAS on a diallel panel.

Evolution of phenotypes: understanding diversity and the role of plasticity in adaptation to new environments
SMBE-OR-072

Compensatory mutations drive morphological evolution

Z. Sarkadi*, Z. Farkas, K. Kovács, G. Fekete, C. Molnár, P. Horváth, D. Kalapis, Z. Bódi, C. Pál, B. Papp

Abstract: Microbes display great diversity in cellular morphology. Traditional explanations of this variation are adaptation to changing environments and neutral evolution. Here we propose an alternative scenario, where deleterious mutations play an important role: they induce selection pressure for compensatory evolution which may lead to novel cellular morphology. We tested this scenario with 180 budding yeast knockout strains which underwent compensatory evolution in the lab. We carried out high-throughput automated microscopy and image analysis on the ancestor and evolved lines, allowing us to measure 187 morphological traits. Restoration of the wild-type morphology was rare, while numerous lines evolved novel phenotypes. For example, several independently evolved lines that carried mutations in DNA repair genes obtained increased cell size and altered nuclear position. Strikingly, not only the cellular morphology, but also the capability to form invasive filaments changed during compensatory evolution. Specifically, some of the strains gained the ability to adopt filamentous growth, which is important for invasion of host tissue in pathogenic fungi. Taken together, deleterious mutations can promote the evolution of cellular morphology and may contribute to the emergence of fungal pathogenicity.

Evolution of phenotypes: understanding diversity and the role of plasticity in adaptation to new environments
SMBE-OR-073

Gene loss and duplication events have facilitated the evolution of multiple cone opsin genes in gobies

F.-Y. Wang [1,*], T.-W. Liu, T.-Y. Wang [2]
[1]National Applied Research Laboratories, Taiwan Ocean Research Institute, Kaohsiung, [2]Academia Sinica, Biodiversity Research Center, Taipei, Taiwan

Abstract: The photic environments of habitats shape the opsin gene evolution in fish. The divergent habitats and life-styles of gobies make them good candidates for studying the adaptive evolution of opsin genes. Eight goby species from different photic environments were collected to test whether their visual systems are tuned to the light spectra of their habitats. The opsin genes of the eight gobies were cloned and sequenced, and the light-spectra of their habitats were measured for comparison. First, we found all eight gobies possess single Rh1 (dim-light sensitive) rhodopsin gene, and lost SWS1 (short-wavelength sensitive 1) cone opsin gene. However, various gene numbers were observed in other three cone opsin genes, which are SWS2 (short-wavelength sensitive 2), Rh2 (middle-wavelength sensitive), and LWS (long-wavelength sensitive) opsin genes. These findings imply that the opsin gene duplication and loss in goby genome could be co-related with the complexity of habitat photic environment. Second, the amino acid sequence comparison shows that non-synonymous substitutions in the known spectral tuning sites were detected. Such substitutions can differentiate the λ_{max} between the duplicates of SWS2, Rh2, and LWS opsins. The diverse λ_{max} of goby opsins could result in the differential visual abilities in various spectral ranges, i.e. blue, green and red light. Finally, the evolutionary analysis of opsins shows that gene loss, duplication, and diversification work together to shape the visual capacities of gobies, especially in LWS opsin genes. We found that an ancient duplication and the follow-up parallel evolution within the two duplicate lineages, LWSA/B, could result in the intraspecific functional divergence or convergence of the LWS opsins a goby species possess. The aforementioned findings suggest that opsin gene evolution of gobies is highly species-specific, partly because of ecological differences in visual tasks.

Evolution of phenotypes: understanding diversity and the role of plasticity in adaptation to new environments
SMBE-OR-074

Altitude shapes local adaptation in Heliconius butterflies

G. Montejo-Kovacevich*

Abstract: *Heliconius* butterflies have long been studied for their Müllerian mimicry and are one of the most widely distributed genera of butterflies in the Neotropics. As such, they range across a vast number of habitats and climates, with some species that have specialised to high altitude environments in the Andes, and others, such as *H. erato*, that are found in continuous populations from 0 to 1600 m above sea-level. A key question is to disentangle heritable traits from plastic responses to the environment. Here I will present novel genomic signatures of local adaptation to altitude, the traits associated with it and dissect the plasticity and selection pressures shaping these traits. We have taken a three population approach, replicating sampling designs used to detect loci involved in extreme high altitude adaptation in Tibetan human populations. We use population branch statistics and excess to find signatures of selection to high altitudes in two species across four replicated altitudinal clines in the Andes. We found high levels of convergence and parallel adaptation between sides of the Andes. The dazzling diversity in colour patterns of *Heliconius*has perhaps obscured the less conspicuous variation in wing shapes and other locally adaptive traits. With a wild collection of over 3000 individuals, thermal tolerance tests in the wild, common-garden rearing experiments, and association studies with whole genomes of 300 wild *H. erato*, we show that wings are rounder at high elevations, both within and across species, that wing shape is heritable, and that several genomic regions are underlying variation in this trait. In contrast, our common-garden rearing experiments show that thermal tolerance is highly plastic, despite significant differences across altitudinally structured populations in the wild. Our study succeeds in linking signatures of selection to local environmental adaptation and reveals novel tractable adaptive traits in this system, making this an exciting new avenue for *Heliconius*evolution research.

Evolution of phenotypes: understanding diversity and the role of plasticity in adaptation to new environments
SMBE-OR-076

A virulence-factor associated component of variation in Helicobacter pylori

K. Thorell [1], D. Falush [2,*]
[1]University of Gothenberg, Gothenberg, Sweden, [2]University Of Bath, Bath, United Kingdom

Abstract: *Helicobacter pylori*, which colonizes the human stomach, is geographically differentiated, with patterns of variation that reflect the migration history of its host. However, differentiation can also arise because bacteria adapt to the specific challenges involved in colonizing hosts in their location, for example due to differences in diet or human genetics. We use a global collection of more than 1000 H. pylori genomes show that there is a distinct component of genetic variation that is extremely enriched for virulence factors, including cagA and vacA and lesser known and uncharacterized genes. A substantial fraction of the variants within the component are in frpB-4, which is involved in nickel uptake, suggesting that dietary factors may be important in determining the strategies that the bacteria adopt in colonizing their host and which also lead to different rates of progression to clinical diseases such as gastric cancer. Dissection of specific interactions between variants provide important clues about functional interactions underlying different strategies. Finally, we relate our findings to geographic variation in virulence, describing the most common strategies adopted by bacteria in South America, Europe and East Asia.

Evolution of phenotypes: understanding diversity and the role of plasticity in adaptation to new environments
SMBE-OR-097

Dramatic variations in extrachromosomal DNA caused by environmentally-responsive gene expression

R. Hull [1], G. Pizza [1], X. Vergara [1], M. King [1], J. Houseley [1,*]
[1]Epigenetics, Babraham Institute, Cambridge, United Kingdom

Abstract: We have recently demonstrated that budding yeast adapt to environmental copper through non-random transcriptionally-stimulated copy number variation (CNV) at the *CUP1* locus (Hull et al 2017), and have previously shown that ribosomal DNA CNV is similarly controlled in response to glucose. In both cases, CNV occurs through recombination events that are stimulated by transcriptional activity and chromatin state, raising the question of how widespread and locus specific gene amplification events are.

To answer this question, we have analysed another CNV outcome: the formation of extrachromosomal circular DNA (eccDNA), which accumulates to high levels during yeast replicative ageing. Remarkably, we find that ageing in the presence of copper induces the accumulation of *CUP1* eccDNA 10-fold. To determine locus specificity, we developed a quantitative eccDNA sequencing method that reveals *CUP1* to be the only locus in the genome from which eccDNA is differentially accumulated under copper treatment. Importantly, eccDNA accumulation is caused by transcriptional activation, and re-programming promoter specificity allows site-specific eccDNA accumulation to occur in the absence of selection. Mechanistic analysis reveals that eccDNA accumulation derives from increased rates of double strand break formation at some highly transcribed loci in aged cells, explaining the stimulation of eccDNA accumulation under specific environments.

We propose that differential formation of eccDNA provides an environment-specific reservoir of heterogeneous genetic material. Retention of eccDNA in aged cells is known to be compromised at extreme age and under stress, and therefore aged or stressed mothers would produce offspring with increased genetic diversity relevant to the current environment through eccDNA inheritance

Hull et al PLoS Biol 2017, 15, e2001333

Evolution of phenotypes: understanding diversity and the role of plasticity in adaptation to new environments
SMBE-OR-098

The impact of sub-continental ancestry on physical appearance in Latin Americans

J. C. Chacon-Duque [1 2,*], K. Adhikari [1], M. Fuentes-Guajardo [1], J. Mendoza-Revilla [1], V. Acuna-Alonzo [1], S. Canizales-Quinteros [3], G. Bedoya Berrio [4], M.-C. Bortolini [5], C. Gallo [6], G. Poletti [6], F. Rothhammer [7], R. Gonzalez-Jose [8], D. Balding [1], G. Hellenthal [1], A. Ruiz-Linares [9 10]

[1]Division of Biosciences, University College London, [2]Earth Sciences, Natural History Museum, London, United Kingdom, [3]UNAM, Mexico City, Mexico, [4]Universidad de Antioquia, Medellin, Colombia, [5]UFRGS, Porto Alegre, Brazil, [6]UPCH, Lima, Peru, [7]Universidad de Tarapaca, Arica, Chile, [8]CENPAT-CONICET, Puerto Madryn, Argentina, [9]Fudan University, Shanghai, China, [10]Aix Marseille Univ, Marseille, France

Abstract: The history of Latin America involved extensive genetic admixture, particularly between Native Americans, Europeans and sub-Saharan Africans. We have recently reported a comprehensive analysis of the genetic history of the region using a set of haplotype-based approaches, describing the sub-continental genetic ancestry of >6,000 Latin Americans across five countries (Brazil, Chile, Colombia, Mexico and Peru) with an unprecedented resolution. We were able to demonstrate how these results not only matched historical records regarding the precise origins of the ancestors of these populations, but also highlighted the extent of scarcely recorded migrations. Here we describe how, using measurements on the sampled individuals' physical appearances, we explored the impact of this fine-scale genetic structure on phenotypic variation across Latin America.

We evaluated these associations by performing linear regressions using the phenotypic measurements and sub-continental ancestry estimates, accounting for multiple covariates and collinearity, including broad continental ancestries. We found significant correlations between variation in southern and northern European ancestry sub-components and skin pigmentation, as well as between variation in Central Andean and Mapuche Native American sub-components and nose shape. For the latter, we also detected significant differences in allele frequencies between these groups at six loci previously associated with nose morphology in this sample. Furthermore, consistent with selection effects at these SNPs, the allele frequencies at this set of SNPs jointly were more differentiated between Central Andean and Mapuche than was the case in randomly selected sets of six genome-wide SNPs.

This research evidences the impact of fine-scale regional genetic variation on human phenotypic diversity and provides a template for future studies aiming to further understand the genetic architecture of complex traits. Such a template is vital given the ubiquity of recent admixture in nearly all world-wide human populations.

Evolution of phenotypes: understanding diversity and the role of plasticity in adaptation to new environments
SMBE-OR-099

Mutational dynamics remodeling transcriptional plasticity to drought stress in the Arabidopsis genus

J. de Meaux [1,*], K. Steige [1], F. He [1], U. Goebel [1], A. Beyer
[1]University Of Cologne, Cologne, Germany

Abstract: To describe how stress responses evolve and shed new light on the evolution of plastic reactions in ecologically divergent species, we undertook a comparative time-serie analysis of plant transcriptome responses to progressive dehydration in the *Arabidopsis* genus. Our study focused on the species *Arabidopsis lyrata*, which is robust to drought stress, and compared it to the response displayed by the outgroup species *A. thaliana* and the more sensitive sister species *A. halleri.* Our experimental design included interspecific hybrids along the parental genotypes, so that the cis- and trans-acting basis of regulatory differences could be assessed.

We observe that the evolution of the stress response has a widely polygenic basis, with a major contribution of numerous independent cis-regulatory modifications. This contribution reveals a broad mutational skew, which has independently decreased the magnitude of the plastic gene expression response in each of the two sister species *A. lyrata* and *A. halleri*. This mutational skew leads to widespread assimilation of the stress response. We however observe that a larger number of genes have evolved a magnified stress-response in *A. lyrata* compared to *A. halleri*. Although this pattern is largely sustained by an unknown and likely reduced number of *trans*-acting variants, evolutionary rates of *cis*-regulatory divergence point to the action of natural selection. This work sheds new light on the mutational dynamics of gene expression plasticity and presents new avenues for the study of its adaptive remodeling in novel ecological contexts.

Evolution of phenotypes: understanding diversity and the role of plasticity in adaptation to new environments
SMBE-OR-100

Mammalian habitat transitions and the power of the "LossOme"

F. Castro [1,*], M. Lopes-Marques [1], M. Fonseca [1], L. Alves [1], A. Machado [1], R. Ruivo
[1]CIIMAR, Matosinhos, Portugal

Abstract: Genomes are dynamic biological units, with processes of gene duplication and loss triggering evolutionary novelty. Mammalian radiation entailed the successful colonization of multiple and ecologically diverse habitats. The invasion of aquatic ecosystems by marine mammals such as Cetacea is a unique example of a radical habitat shift. This evolutionary path was accompanied by the appearance of novel phenotypic traits. To identify the critical genomic events associated with these changes is a central question in evolutionary biology. Here, we investigate how gene loss has impacted the genomes of marine mammals such as Cetacea in their adaptive road. By combining detailed comparative genomic analysis in multiple lineages, we show how key morphological (e.g. skin) and behavioral (e.g. circadian rhythmicity) adaptations emerged in these aquatic lineages as a result of gene inactivation processes. Our findings suggest that multiple gene loss events, ancestral and convergent, in mammalian evolution concurred with unique adaptive roads and associated phenotypes.

Evolution of phenotypes: understanding diversity and the role of plasticity in adaptation to new environments
SMBE-OR-101

A new mouse model sheds light on the evolutionary impact of an ancient deletion polymorphism in the human growth hormone receptor gene

M. Saitou [1,*], S. Resendez [1], K. Dean [1], F. Wu [2], X. Mu [2], O. Gokcumen [1]
[1]Biological Sciences, [2]Jacobs School of Medicine, State University of New York at Buffalo, Buffalo, United States

Abstract:
The growth hormone receptor (GHR) gene codes the receptor protein for the growth hormone. GHR is highly conserved among mammals. However, its third exon is polymorphically deleted with 30% allele frequency in the human population. Our lab has previously reported that Neanderthal and Denisovan genomes also carry the deletion allele (GHRd3), suggesting that the GHRd3 and ancestral alleles were segregating in the human-Neanderthal-Denisovan ancestral population.
Using population genetics analysis, we first resolved the haplotype architecture of the locus harboring GHRd3. We identified single nucleotide variants that tag the deletion allele (R2 with GHRd3 > 0.96). We showed that the haplotype carrying the GHRd3 allele is significantly associated ($p < 3e-09$) with "Standing height" in a cohort of 450,000 British individuals available through UK Biobank dataset.
To investigate the phenotypic impact of GHRd3 at the developmental and molecular level more thoroughly, we constructed a CRISPR-Cas9 based mouse where we deleted exon 3, modeling the human polymorphism. Using this model, we showed that there is a differential rate of growth between GHRd3 and wildtype mice. Moreover, comparative RNA-sequencing analyses from liver tissues showed that GHRd3 affects the expression of genes enriched for metabolic processes.
Taken together, our study suggests non-neutral evolution of GHRd3 in humans and verified previous associations with developmental phenotypes. Furthermore, we identified novel biological targets of GHRd3, affecting metabolic pathways in the liver. Our integrative approach sheds new light on the evolutionary impact of ancient exonic structural variants.

Evolution of phenotypes: understanding diversity and the role of plasticity in adaptation to new environments
SMBE-OR-103

Genomic architecture and evolution of a seasonal reaction norm

K. R. Van Der Burg*, R. D. Reed

Abstract: Phenotypic plasticity allows genomes to encode multiple distinct phenotypes that can be revealed in response to specific environmental cues. It has long been known that selection can cause an ancestrally plastic trait to become fixed, thereby decoupling environmental versus genetic induction—a process called genetic assimilation—yet the molecular basis of this process is still poorly understood. Here we used artificial selection to genetically assimilate a red seasonal wing phenotype from a naturally plastic population of butterflies. Through genetic mapping and gene editing, we show that three genes are responsible for fixing a phenotype. Combined with endocrine and chromatin accessibility assays, we infer that transition of wing coloration from an environmentally determined trait to a fixed, genetic trait can occur rapidly through selection on cis-regulatory alleles of trait-specific regulatory and effector genes, not environmental cue detection or endocrine genes. We propose that this mode of genetic evolution is favored by selection because it allows tissue- and trait-specific tuning of reaction norms, while avoiding pleiotropic effects that would be caused by upstream changes to core cue and response mechanisms.

Evolution of proteins and molecular pathways/networks under higher level selective effects
SMBE-OR-136
Mechanistic Modelling Without Prior Knowledge of Parameter Values: Global Robustness & Fitness Landscape of a Developmental Switch in Plants
M. A. Savageau [1,*]
[1]Univesity of California Davis, Davis, United States

Abstract: One of the defining characteristics of biological organisms is their robustness/evolvability. They typically tolerate wide environmental and genetic variations when robust, and yet change phenotype and evolve when they are not. It this talk I will (a) describe a new (phenotype-centric) modelling strategy to address these issues in a quantitative fashion, but without having to know values for mechanistic parameters in advance, and (b) illustrate the potential of this strategy with an analysis of gene circuitry implicated in xylem cell differentiation. The phenotype-centric strategy focuses on enumerating the phenotypic repertoire of a system and predicting parameter values by analytical methods, which is in stark contrast to traditional (simulation-centric) modelling strategies that focus on estimating parameter values and numerically exploring phenotypes by sampling high-dimensional parameter space and repetitively solving differential equations. Recent results from intact root and single-cell transcriptome experiments provide evidence for a bistable commitment to xylem cell differentiation at the cellular level. The gene expression data from single-cells implicates four genes in circuitry that underlies this cellular commitment. The analysis will show how alternative realizations of the circuitry correspond to hypotheses that can be evaluated and ranked by global tolerance to change that would lead to a loss of the phenotype. These results suggest a new way to view fitness landscapes; they also allow predictions of the most robust system design and guidance for experimental tests to establish its validity.

Evolution of proteins and molecular pathways/networks under higher level selective effects
SMBE-OR-137

Expansion of the corticosteroid synthesis pathway in primates through a multi-step enzyme

C. F. Olson-Manning [1,*], S. F. Kreutzmann [2], E. F. Bien [2], J. W. Thornton [3]
[1]Augustana University, Sioux Falls, United States, [2]Biology, Augustana University, Sioux Falls, [3]Ecology and Evolution, University of Chicago, Chicago, United States

Abstract: Metabolic networks are complex cellular systems and their function is dependent on the interactions among and regulation of the enzymes in the network. However, the mechanisms that lead to the expansion of networks are not well understood. Processes for expansion may include: co-opting components from other networks, a current network member gaining a novel function, or the specialization of paralogs from a generalized enzyme where the new network is now dependent on two enzymes instead of one. While gene duplication is a major driver of the expansion and functional evolution of metabolic networks, the effect and fate of retained duplicates in a network is poorly understood. To illuminate the mechanisms that constrain the specialization of a duplicated network component, we study the evolution of an enzyme family that performs multiple subsequent enzymatic reactions in the corticosteroid pathway in primates. The products of the pathway (aldosterone, corticosterone, and cortisol) are steroid hormones in tetrapods that regulate metabolism and stress. These steroids are synthesized by a multi-step enzyme Cytochrome P450 11B (CYP11B) that performs subsequent steps on different carbon atoms of the steroid derivatives. Through ancestral state reconstruction and *in vitro* characterization, we find that the ancestor of the CYP11B1 and 11B2 paralogs in primates had moderate ability to synthesize cortisol and aldosterone. Following a duplication event in the primate lineage one copy in humans (CYP11B1) specialized on production of cortisol while its paralog (human CYP11B2) maintained its ability to perform multiple subsequent steps as in the ancestral pathway. The maintenance of function in CYP11B2 was likely mediated by the change in pathwasy regulation following the duplication of the ancestral CYP11B. While the ancestral single-CYP11B pathway was regulated allosterically (by yet another pathway component), the two-CYP11B system maintains localization steroid expression in the adrenal cortex through different regulatory mechanisms. Our results suggest that pathway context and tissue-specific regulation both play a role in constraining potential outcomes of metabolic network elaboration.

Evolution of proteins and molecular pathways/networks under higher level selective effects
SMBE-OR-137A

The impact of gene network topology on the evolution of gene-specific expression noise

N. Puzović [1,*], G. V. Barroso [1], J. Y. Dutheil [1]
[1]Department of Evolutionary Genetics, Max Planck Institute for Evolutionary Biology, Plön, Germany

Abstract: Genetically identical cells grown under identical environmental conditions display significant variation in terms of the number of proteins in each cell. This variability of gene expression, or expression noise, originates from the inherent stochasticity of diffusion and binding of the molecular players involved in gene transcription and translation. We previously showed, using mouse single-cell transcriptomics data, that the main factor explaining gene-specific transcriptional noise is the position of the protein in the protein-protein interaction network, with more central proteins exhibiting less noise. We hypothesized that stabilizing selection at the network level may result in differential noise constraints at the gene level. Here, we develop models of gene expression in networks to test this hypothesis. We model noise propagation between genes and let the networks evolve while applying stabilizing or directional selection on the amount of product of one of the genes. We find that the gene regulatory function between two genes affects the output noise, and that the network topology upstream of the regulated gene can tune its noise. These findings suggest that (1) different selective pressures may shape the gene-specific expression noise by changing the topology of the network and (2) since expression noise is amenable to selection, it must be considered as an important parameter in evolutionary studies of gene expression.

Evolution of proteins and molecular pathways/networks under higher level selective effects
SMBE-OR-138

Can natural variation be used to reveal the genetic architecture of regulatory networks? A new comprehensive simulation approach

A. A. Teterina [1 2,*], P. L. Ralph [1], P. C. Phillips [1]
[1]Institute of Ecology and Evolution, University of Oregon, Eugene, United States, [2]Center of Parasitology, Severtsov Institute of Ecology and Evolution RAS, Moscow, Russian Federation

Abstract: Understanding the genetic architecture of complex traits remains one of the most challenging tasks in biology. From an evolutionary perspective, the structure of the gene regulatory networks that generate each phenotype is shaped by both adaptive and non-adaptive forces. Previous evolutionary simulations of gene networks have shown that pre-adapted networks can be robust against mutations (Wagner 1996) and that increase of selection, recombination, and mutation rates boost the speed of adaptation to the new optimum (Kioukis & Pavlidis 2018). A necessary next step is to more directly connect evolutionary dynamics of the networks themselves with signatures of adaptation and genetic variation that are emerging from population genomic studies. To this end, we have developed a new simulation framework in SLiM3 that includes selection on gene expression levels in evolvable, dynamic gene regulatory networks whose genes are distributed across realistically-sized genomes. Using this approach, we explore how genetic networks respond to adaptation to a novel environment to address the following questions: Can the signature of population genomic variation be used to determine the role of the gene in the network? What is the relationship between population genomic diversity metrics, traditional measures of quantitative genetic variation, and measures of network structure based on graph theory? With that, we are able to tie the evolutionary dynamics of network evolution directly to network-free population genetic measures of genetic diversity in a manner that points the way to unifying functional and population genomic approaches toward a more unified perspective on the evolution of regulatory networks.

Evolution of proteins and molecular pathways/networks under higher level selective effects
SMBE-OR-139

Protein melting temperature and protein folding free energy are not interchangeable in light of protein evolution

R. Razban [1,*], E. Shakhnovich [1]
[1]Department of Chemistry & Chemical Biology, Harvard University, Cambridge, United States

Abstract: The protein misfolding avoidance hypothesis (MAH) explains the universal negative correlation between protein abundance and sequence evolutionary rate across the proteome by identifying protein folding free energy (ΔG) as the confounding variable. Abundant proteins resist toxic misfolding events by being more stable, and more stable proteins evolve slower because their mutations are more destabilizing. Supporting evidence consists only of computer simulations. A study taking advantage of a recent experimental breakthrough in measuring protein stability proteome-wide through melting temperature (T_m) (Leuenberger et al. 2017), found weak MAH support for the *Escherichia coli* proteome, and no support for the *Saccharomyces cerevisiae*, *Homo sapiens* and *Thermus thermophilus* proteomes (Plata and Vitkup 2017). We find that the nontrivial relationship between T_m and ΔG and inaccuracy in T_m measurements by Leuenberger et al. 2017 can be responsible for not observing strong positive abundance–T_m and strong negative T_m–evolutionary rate correlations.

Evolution of proteins and molecular pathways/networks under higher level selective effects
SMBE-OR-140

Differences in expression across gene-networks lead to an extreme case of polyphenism

M. K. Priebe [1,*], C. Martinez-Ruiz [1], E. Vargo [2], R. Nichols [1], M. Robinson-Rechavi [3], Y. Wurm [4]
[1]Queen Mary University of London, London, United Kingdom, [2]Texas A&M University, College state, United States, [3]University of Lausanne, Lausanne, Switzerland, [4]Queen Maru University of London, London, United Kingdom

Abstract: Polyphenism, where a single genome produces two or more discrete phenotypes, provides insight into how variations in a gene network can generate diversity. The ants provide a unique opportunity to study this because their genomes encode three dramatically different but interdependent phenotypes.
Indeed, for 140,000,000 years, selection on ant genomes has been shaped by the need to produce sterile female workers that build and defend the nest, forage for food and nurse the young, reproductive females (queens) that lay up to thousands of eggs per day and can live for decades, and males that live long enough to fertilise a young queen. These different phenotypes are distinct in their morphology, behaviour and physiology. Phenotype is determined either by ploidy (males are haploid while females are diploid) or by the care an individual receives during the larval stage (differentiating workers from reproductive females).
To understand the evolution of genes, pathways and networks underpinning this extreme polymorphism, we used the invasive red fire ant *Solenopsis invicta*. We generated RNAseq from 19 distinct tissues from workers, males and reproductive females, including shared and caste-specific tissues. We extensively characterise differences in expression of genes and pathways between castes and tissues, as well as how network architecture has been shaped by the selection for different phenotypes. With our tissue-level resolution we can start to understand which networks are active in caste-specific tissues and how networks active in tissues shared across castes vary between them. Our results shed light on how a single genome responds to long-term high level diversifying selection encoded in distinct morphs.
We are interested in exploring how gene expression networks differ between castes and how these might lead to the discretely different phenotypes we observe. In order to understand this system, we generated a tissue-specific RNAseq from red fire ant workers, males and reproductive females. This dataset allows us to determine which genes are differentially expressed between castes that might underpin differences between them. Furthermore, we are able to identify groups of linked genes to form networks. With our tissue-level resolution we can start to understand which networks are active in caste-specific tissues and how networks active in tissues shared across castes vary between them. The results will shed light on how the observed phenotypes have evolved from a shared genome.

Evolution of proteins and molecular pathways/networks under higher level selective effects

SMBE-OR-141

Measuring the molecular effects of dosage imbalance inside essential complexes in budding yeast

D. I. Ascencio [1,*], G. Diss [2], A. DeLuna Fors [3], C. R. Landry [1]

[1]Institut de Biologie Intégrative et des Systèmes (IBIS), Université Laval, Quebec QC, Canada, [2]Friedrich Miescher Institute for Biomedical Research, Basel, Switzerland, [3]LANGEBIO-CINVESTAV, Irapuato, Gto, Mexico

Abstract: Certain genes are often retained as duplicates while others are more likely to be found as single-copy, which suggests that some duplications may cause an important and immediate effect that prevents their retention. We tested the extent of such phenotypic effects through large-scale experiments in the budding yeast. We found that around 10% of essential genes produce a fitness disadvantage when duplicated. Interestingly, genes that encode subunits of protein complexes are more likely to cause a disadvantage when duplicated than genes of other functional categories. The gene balance hypothesis, which predicts that the single-gene duplication of subunits is harmful because it perturbs the stoichiometric balance of the complex, may provide an explanation for this observation. Whereas the perturbation of protein-protein interaction networks caused by the deletion of complex subunits has been tested previously, the molecular effects of doubling the dosage of a single complex subunit are still poorly explored. We therefore systematically measured the perturbations in the protein-protein interaction network of protein complexes by increasing the dosage of every single subunit. We observe that most subunit duplications have small to non-detectable effects on the interaction network of their complex and, for most subunits, is not correlated with a significant fitness effect. Taken together, our results suggest the presence of post-transcriptional buffering mechanisms controlling the abundance of subunits of complexes that make fitness effects independent from the perturbation of the complexes themselves.

Evolutionary genetics and genomics of metabolic networks
SMBE-OR-234
Plant Metabolic Clusters – From Genetics to Genomics
A. Osbourn [1,*]
[1]John Innes Centre, Norwich, United Kingdom

Abstract: Plants produce a wealth of natural products. The vast majority of the natural product diversity encoded by plant genomes remains as yet untapped. The explosion in plant genome sequence data, coupled with affordable DNA synthesis and new DNA assembly technologies, now offer unprecedented opportunities to harness the full breadth of plant natural product diversity and generate novel molecules in foreign hosts using synthetic biology approaches. The recent discovery that genes for the synthesis of different kinds of natural products are organised in biosynthetic gene clusters in plant genomes opens up opportunities for mining for new pathways and chemistries. This advance, in combination with powerful new transient plant expression technology, is enabling the development of rational strategies to produce known and new-to-nature chemicals tailored for food, health and industrial applications. This presentation will focus on our work on developing a translational synthetic biology pipeline for rapid preparative access to plant natural products and novel analogs using synthetic biology approaches. It will also highlight recent advances in our understanding of the genomic rearrangements underpinning the formation of new plant biosynthetic gene clusters, and of the functions of plant natural products in nature.

Evolutionary genetics and genomics of metabolic networks

SMBE-OR-235

Enzyme evolution preceded genome evolution in the origin of catnip nepetalactone biosynthesis

B. R. Lichman [1,*], C. R. Buell [2], S. E. O'Connor [3] and Mint Evolutionary Genomics Consortium

[1]Department of Biology, University of York, York, United Kingdom, [2]Department of Plant Biology, Michigan State University, East Lansing, United States, [3]Biological Chemistry, John Innes Centre, Norwich, United Kingdom

Abstract: Catmints (*Nepeta* sp.), including catnip (*N. cataria*), are famed for their stimulatory effects on cats. The origins of this effect are nepetalactones, volatile iridoid stereoisomers which have been shown to play a role in plant-insect interactions. We have recently identified the enzymes responsible for the formation of nepetalactones[1,2]. Phylogenomic analysis[3] of the mint family (Lamiaceae) has revealed that whilst iridoid biosynthesis was an ancestral trait, and is still present in many mint species, it was lost in the Nepetoideae lineage and re-emerged in *Nepeta*. To uncover the mechanisms of metabolic loss and gain in the evolution of *Nepeta*, we generated draft genomes of two iridoid producing *Nepeta* species (*N. cataria* and *N. mussinii* syn *racemosa*) along with the closely related, non-iridoid producing hyssop (*Hyssopus officinalis*)[4]. This comparative genomic approach, together with biochemical and in-depth phylogenetic analysis, has revealed (a) conserved metabolic gene clusters in *Nepeta* which control iridoid formation and stereochemistry, (b) evolutionary mechanisms by which iridoid biosynthesis was lost and re-gained, and (c) a sequence of molecular and genomic events indicating that, in *Nepeta*, enzyme evolution preceded the formation of gene clusters. This study provides insight into plant metabolic evolution and, crucially, the interplay between enzyme evolution and genome evolution.

1. N. H. Sherden, B. R. Lichman *et al.*, Identification of iridoid synthases from Nepeta species: Iridoid cyclization does not determine nepetalactone stereochemistry. *Phytochemistry*. **145**, 48–56 (2018).

2. B. R. Lichman *et al.*, Uncoupled activation and cyclisation in catmint reductive terpenoid biosynthesis. *Nat. Chem. Biol.* **15**, 71–79 (2019).

3. Mint Evolutionary Genomics Consortium, Phylogenomic Mining of the Mints Reveals Multiple Mechanisms Contributing to the Evolution of Chemical Diversity in Lamiaceae. *Mol. Plant*. **11**, 1084–1096 (2018).

4. B. R. Lichman, Mint Evolutionary Genomics Consortium, The molecular, genomic and evolutionary origins of nepetalactone biosynthesis in catnip. *In preparation*.

Evolutionary genetics and genomics of metabolic networks
SMBE-OR-236
Molecular mechanisms and evolution of a novel floral volatile biosynthesis in wild tobacco
H. Guo, I. T. Baldwin [1], S. Xu[*]
[1]Molecular Ecology, Max Planck Institute for Chemical Ecology, Jena, Germany

Abstract: Flowering plants often produce diverse volatiles that are important for attracting their pollinators and attaining reproductive success. In the night-flowering wild tobacco, *Nicotiana attenuata*, benzyl acetone (BA) is a species-specific floral volatile, which can increase *Manduca sexta* moth mediated pollination success in nature by approximately 80%. However, the biosynthetic pathway of BA and its evolution remain a mystery. Here, using genetic mapping, gene co-expression network analysis and gene silencing, we demonstrate that phenylalanine ammonia-lyase 4 (*NaPAL4*), isoflavone reductase 1 (*NaIFR1*) and chalcone synthesis 1 (*NaCHAL1*) are necessary for the BA biosynthesis. Transient expression of these three genes together in *N. attenuata* leaves resulted in BA emission from leaves, suggesting that these three genes are sufficient for BA biosynthesis. Among natural accessions, expression changes of these three genes can explain more than 85% of the variation in floral BA emissions. Phylogenomic analysis showed that while *NaPAL4* evolved before the whole genome triplication that was shared among Solanaceae species, *NaIFR1* and *NaCHAL1* evolved from gene duplications that are specific to the *Nicotiana* genus. Comparative expression analysis among closely related *Nicotiana* species revealed that the increased expression of *NaIFR1* in *N. attenuata*, likely resulting from an insertion of a chloroplast DNA fragment in its promoter, was important for the evolution of this novel floral volatile biosynthetic pathway.

Evolutionary genetics and genomics of metabolic networks
SMBE-OR-237

THE E. COLI CLADE NEEDED ONLY A SINGLE HORIZONTAL DNA TRANSFER FOR EACH DETECTABLE METABOLIC INNOVATION, WHILE SIMPLER METABOLIC SYSTEMS WOULD REQUIRE DOZENS OF TRANSFERS

M. Lercher [1,*]
[1]Heinrich Heine University , Dusseldorf, Germany

Are you a member of SMBE?: No

Poster Submission: Anecdotal evidence suggests that some bacteria are more adaptable than others; for example, intestinal E. coli frequently spin off pathogenic strains adapted to other human tissues, while gastric Helicobacter pylori do not. Why? Fisher believed that complexity hinders adaptation due to increasing pleiotropic constraints; here, we argue the opposite, based on two complementary analyses of bacterial metabolic systems. Using pan-genome-scale modelling, we explore the ability of 71 unicellular organisms to adapt to 5000+ diverse nutritional environments. While the smallest metabolic systems – those of the endosymbionts Buchnera aphidicola and Helicobacter pylori – require on average 50 additional metabolic reactions to adapt to new environments, the metabolically much more complex generalist E. coli requires on average less than 5. During adaptation, organisms may accidentally acquire the ability to also grow in additional, non-selected environments; this “collateral adaptation” is much stronger for generalists than for specialists.We use a combination of evolutionary genomics and metabolic modelling to explore how different E. coli genomes introduced phenotypic innovations in their evolutionary history. We identify 3,000+ phenotypic innovations that arose through changes in accessory genome content; each arose through the horizontal acquisition of a single DNA segment less than 30kb long. 98% of metabolic phenotypes accessible to the combined E. coli pan-genome can be bestowed on any individual genome by transferring a single 30kb-DNA segment from one of the extant strains.

Evolutionary genetics and genomics of metabolic networks

SMBE-OR-238

How selection during the last 10,000 years shaped genetic variation associated with common metabolic and inflammatory diseases in present-day populations in Europe

A. Eriksson [1,*]

[1]Medical and Molecular Genetics, King's College London, London, United Kingdom

Abstract: Cardiovascular, metabolic and autoimmune diseases are among the leading causes of death globally and are increasing in importance in many parts of the world. These diseases are multifaceted, and involve a complex interplay between multiple genes and environmental factors that differ between populations and vary among individuals within populations. It has long been known that they have common risk factors and comorbidities, and recent genome wide association studies have revealed overlapping genetic pathways between these disorders: in particular, dysregulation of immune system signalling and resulting chronic inflammation has emerged as a common factor in the aetiology of a range of cardiovascular, metabolic and autoimmune diseases.

The increase of these diseases over the last few decades is often attributed to a mismatch between modern environments and life styles to those in which humans evolved. Support for this mismatch hypothesis have come from studies finding the signature of strong and recent positive selection around genetic variants associated with disease in contemporary populations and that many of these variants are relatively common in present-day populations. For example, adaptations that enhance the conversion of essential fatty acids from plant precursors have proinflammatory side effects that can contribute to diabetes and cardiovascular disease, and variants associated with coeliac disease have been shown to protect against bacterial infections. Thus, past adaptations to different environments have clearly been important, but their consequences for health in present-day populations, especially with regard to adaptations to changes in diet and pathogen exposure, has not been systematically investigated.

I used large datasets of genetic variation in ancient and modern populations to investigate the joint evolution of genes (and their expression) in pathways associated with common inflammatory and metabolic diseases over the last 10,000 years in Europe, Middle East and central Asia. The analyses revealed distinct evolutionary patterns associated with each archaeological epochs during this time period, characterised by different networks of co-evolving genes or regulatory variants affecting immune system functions, and with some genetic variants showing evidence of positive selection in some time periods and negative selection in others. Among the strongest signals are changes to genes involved in the regulation of the adaptive immune system that today are important genetic risk factors for several immune and metabolic diseases, such as multiple sclerosis and type 1 diabetes.

Evolutionary genetics and genomics of metabolic networks
SMBE-OR-239
Reconstructing the evolution of central carbon metabolism in budding yeasts
K. Correia [1,*], R. Mahadevan [1]
[1]Chemical Engineering and Applied Chemistry, University of Toronto, Toronto, Canada

Abstract: *Saccharomyces cerevisiae* has served as an important model organism for yeasts and eukaryotes. One consequence of its role in molecular biology is that current knowledge of budding yeast genomics and genetics is heavily skewed towards *Sac. cerevisiae*. To broaden our understanding of yeast metabolism beyond *Sac. cerevisiae* and the Saccharomycetaceae family, AYbRAH, an ortholog database spanning 600 million years of evolution, was used to refine the budding yeast species tree topology, identify important and reoccurring events in the radiation of budding yeasts, and partially reconstruct the genomes of proto-yeast and proto-fermenter, the first budding yeast and budding yeast that fermented glucose to ethanol, respectively. Analysis of budding yeast gene duplications place *Nadsonia fulvescens* in a clade sister to *Blastobotrys adeninivorans*, and *Bla. adeninivorans* in a clade sister to *Yarrowia lipolytica*. This topology has not been predicted by any concatenation or coalescent-based methods, indicating that genes chosen for phylogenetic reconstruction do not have a strong phylogenetic signal for *Nadsonia*. The shift from a citrogenic to ethanologenic lifestyle in budding yeasts is marked by an ancient duplication of *ACS1* to *ACS2*, which encode acetyl-CoA synthetase expressed during growth on acetate and glucose, respectively. This duplication led to the emergence of the pyruvate dehydrogenase (PDH) bypass in budding yeasts. Flux balance analysis (FBA) indicates the PDH bypass has a higher protein and phospholipid yield from glucose than ATP citrate lyase, the ancestral acetyl-CoA source in proto-yeast. Complex I and internal alternative NADH dehydrogenase (Ndi) orthologs have been independently lost in multiple yeast lineages. FBA simulations demonstrate Complex I has a significantly higher biomass yield than Ndi with glucose. The repeated loss of Complex I in yeasts, despite having a higher biomass yield than Ndi or aerobic fermentation, is proposed to be a result of its inactivation during cell proliferation to reduce harmful reactive oxygen species generation. Additional gene duplications have played a role in changing the expression of genes for different carbon sources, transitioning homomer enzyme complexes to heteromers with altered enzyme kinetics, possible ribosome heterogeneity in independent yeast lineages, changes in enzyme localization and cofactor preference. HGT appears to be an underappreciated factor in yeast evolution. The Dikarya pan-genome highlights the role chance duplications play in shaping the evolution of metabolism in yeasts and fungi.

Evolutionary Insights from Metagenonics

SMBE-OR-166

Metagenomics strain resolution on assembly graphs

C. Quince[*], S. Nurk[1], S. Raguideau[2]

[1]University of St Petersburg, St Petersburg, Russian Federation, [2]University of Warwick, Coventry, United Kingdom

Abstract: There are a number of algorithms that attempt to resolve strains *de novo* from short read metagenomics data. These either utilise co-occurrence of variant positions across samples or linkage of variants in reads. They are all, however, post-hoc solutions that are applied after metagenomics assembly usually through mapping onto contigs. I will describe an approach that resolves strains directly on metagenome assembly graphs by finding the set of paths and abundances that best explain the observed coverage on unitigs across multiple samples. I will show how this method can be used to resolve strain abundances and single copy core gene haplotypes within metagenome bins and discuss the feasibility of extending these methods to the entire genome. I will conclude by presenting the application of this algorithm to metagenome time series from the human gut and anaerobic digestion reactors.

Evolutionary Insights from Metagenonics
SMBE-OR-168
No more discrimination – including microbial eukaryotes in metagenomic surveys.
V. Rossetto Marcelino [1,*], T. C. Sorrell [1], E. Holmes [1]
[1]University of Sydney, Sydney, Australia

Abstract: Microbial eukaryotes are diverse, ubiquitous, and play important functional roles in environmental and host-associated microbial communities. However, micro-eukaryotes are far less studied than their bacterial counterparts, due in part to methodological challenges. Several methods are available to identify taxa directly from metagenome data, but their species-level identification tends to be highly inaccurate or relies heavily on reference databases of complete genomes, which are particularly scarce for microbial eukaryotes. We used a novel concept in read mapping to develop CCMetagen – a metagenome classifier that is highly accurate and fast enough to use the entire NCBI nucleotide collection as reference, facilitating the inclusion of microbial eukaryotes in metagenomic studies. High accuracy is achieved by assessing all read-mapping possibilities, rather than attempting to classify individual reads. Using simulated fungal and bacterial metagenomes, we found that species-level identifications obtained with CCMetagen are between 17x and 1580x more precise than other commonly used metagenome classifiers. We applied CCMetagen to characterize the gut microbiome of wild birds using RNA-based metagenomic data (metatranscriptomics). Besides prokaryotes, we found an abundant and diverse community of micro-eukaryotes, representing 45% of the family-level diversity in the bird microbiome. Interestingly, preliminary results suggest that the composition of the fungal community correlates with genetic similarity of the host, consistent with an evolutionary association – possibly phylosymbiosis – between bird species and their gut mycobiome. Our work opens possibilities to confidently include microbial eukaryotes in studies seeking ecological and evolutionary insights from metagenomes.

Evolutionary Insights from Metagenonics
SMBE-OR-172

Dental calculus as a tool to study the evolution of the oral microbiome in mammals

J. C. Brealey [1,*], H. Leitão [1], T. Hofstede [1], W. Xu [1], L. Dalén [2], K. Guschanski [1]
[1]Department of Ecology and Genetics/Animal Ecology, Uppsala University, Uppsala, [2]Department of Bioinformatics and Genetics, Swedish Museum of Natural History, Stockholm, Sweden

Abstract: The microbiome and its host have a long history of co-evolution, influenced by a complex interplay between host genetic and environmental factors. While the host-associated microbiome changes rapidly in response to stimuli, there is limited understanding of how it is influenced by extrinsic and intrinsic long-term processes, such as climate change and host demography. These processes can only be studied by analysing microbiome changes through time, over known historical events in a host species. Dental calculus, the calcified form of the plaque microbial biofilm, is one of few biological substrates that preserve information about the host and its associated microbiome in a virtually unchanged form through time. Although metagenomics sequencing of ancient dental calculus has provided valuable insights in humans, few studies to date have attempted to characterise the dental calculus microbiome of non-human animals. To establish dental calculus as a standard research tool, we identify important strategies and current limitations for working specifically with historical metagenomics data from non-human mammals. Despite biases in current reference databases, we determine that a recognisable oral microbiome signature can be retrieved from species as diverse as primates, bears, and reindeer. We also identify a subset of microorganisms that may be unique to each host species. Beyond taxonomic characterisation, we demonstrate that a wealth of information can be obtained from such metagenomics data, including detection of potential pathogens, antibiotic resistance genes, microbial metabolic pathways, and host dietary components. Our work establishes dental calculus as a valuable resource for the study of host-associated microbiome evolution.

Evolutionary Insights from Metagenonics

SMBE-OR-167

Meta-omics investigation of soil communities along a natural climatic gradient in the Finnish Arctic tundra

J. Hultman*

Abstract: Climate change is affecting the arctic dramatically as the warming is fastest in higher latitudes. Previously frozen ground is thawing and releasing substantial quantities of carbon which microbes can decompose. Microbes mineralize the carbon fraction and convert it to carbon dioxide and methane. Climate change models estimate that C released from thawing arctic permafrost can represent the largest future transfer of C from the biosphere to the atmosphere. In addition to temperature, also oxygen and moisture, among many others, affect the microbial activity. Understanding the drivers of arctic soil communities and especially how the microbial activity changes with warming and resulting changes in soil characteristics is needed. However, this kind of information in lacking for the arctic soil microbial communities.

We have analyzed over 100 soil plots from a large field site in Scandinavian low arctic with environmental gradient of microclimatic conditions. Metagenomic approach was used to create a database of microbial genes and metagenome assembled genomes (MAGs) found within this gradient. As DNA can originate from dead or dormant cells, we utilized metatranscriptomics to create comprehensive understanding on the metabolic activities of microbes within this fine-scale climatic variation.

Communities were dominated by Proteo-, Actino- and Acidobacteria and active functional pathways were mostly those involved in the breakdown of amino acids, carbohydrates and lipids. Recovered genomes were related to common soil taxa such as the nitrogen-fixing *Bradyrhizobium* (Alphaproteobacteria) and *Granulicella* (Acidobacteria), while others appear to represent distinct lineages distantly related to Candidatus *Koribacter* and Candidatus *Solibacter* (Acidobacteria). The metabolic potential of the soil communities showed, for example, the importance of carbon, nitrogen and sulphur cycles in the tundra ecosystem. By closing critical knowledge gap through integration of microbial activity from meta-omics data to proces model development will increase our general understanding about microbial community function in the changing Arctic.

Evolutionary Insights from Metagenonics
SMBE-OR-169
Large-scale reconstruction of ancient microbial genomes from the human gut

M. C. Wibowo [1,*], J. Luber [1], B. Tierney [1], Z. Yang [2], S. Zimmerman [1], S. Ballal [3], M. Snow [4], S. Leblanc [5], A. Kostic [1]
[1]Microbiology, Harvard Medical School, Boston, [2]Pathophysiology and Molecular Pharmacology, Joslin Diabetes Center, Brookline, [3]Gastroenterology, Hepatology and Nutrition, Boston Children's Hospital, Boston, [4]Anthropology, University of Montana, Missoula, [5]Peabody Museum of Archaeology and Ethnology, Harvard University, Cambridge, United States

Abstract: Studies have indicated industrialized humans have lost certain gut microbes, and this loss of microbial diversity is associated with chronic diseases. To discover extinct bacterial species, we performed shotgun metagenomic sequencing on 12 human paleofeces aged ~2,000 years from Utah and Arizona. We performed the first unbiased *de novo* assembly and the largest reconstruction of microbial genomes from paleofeces to date. We reconstructed 398 medium- and high-quality draft genomes, 207 (52.01%) of which are novel species. Phylogenetic analyses showed our novel genomes expanded the bacterial tree of life and formed new phylogenetic branches with large evolutionary distances to presently known bacteria. At equivalent sequencing depths of 40,000,000 reads, the average number of unique genes in the ancient gut microbiome was 331,648. In contrast, the number for 17 modern stool samples from the Human Microbiome Project was 2.34x lower at ~141,794. Therefore, our gene-level analyses support the hypothesis of extinctions in the modern gut microbiome. From metabolic pathway analyses, we found pathways targeted by antibiotics to be lower in abundance in the modern gut, whereas pathways involved in starch utilization were higher in abundance in the modern gut compared to the ancient gut. These are consistent with higher antibiotic usage and starch consumption in modern populations. We will further analyze the rate of evolution of various bacterial genes. Our work elucidates the evolutionary history of gut symbionts at the gene, pathway, and genome levels and may lead to discovery of extinct bacteria with the potential to restore human health.

Evolutionary Insights from Metagenonics
SMBE-OR-170
Small RNAs for the identification of viruses in the absence of detectable homology
D. J. Obbard [1,*]
[1]IEB, University of Edinburgh, Edinburgh, United Kingdom

Abstract: Metagenomic discovery has revolutionised our knowledge of virus diversity. It is in the process of re-shaping our understanding of host-virus (co-)evolution, and has prompted a renewed interest in the way we classify and describe virus lineages. However, virtually without exception, metagenomic discovery depends on detecting homology with known viruses: almost all virus-discovery pathways currently rely on blast-like sequence similarity searches. One alternative is to utilise immune-mechanisms to identify potential viruses, trusting the host to recognise viruses that we cannot.
The antiviral RNA interference pathway of insects is an ideal marker for viruses as the immune 'signature' of 20-23nt small RNAs is itself directly available from metagenomic sequencing. Using this approach, we have previously identified a number of 'candidate viruses' and virus sequences from the fruit fly, Drosophila. Around half of these candidates are now identifiable as viral on the basis of subsequent discoveries, but other siRNA 'candidates' are still of unknown origin. Among this latter group we have recently identified a new lineage of segmented single-stranded RNA viruses, whose polymerase shows barely-detectable sequence-level homology with that of Picornaviridae, Caliciviridae and Flaviviridae. Related (but unannotated) sequences are detectable in the transcriptomes of around 20 other arthropods, and can be present at extremely high levels.
Metagenomic discovery has revolutionised our knowledge of virus diversity. It is in the process of re-shaping our understanding of host-virus (co-)evolution, and has prompted a renewed interest in the way we classify and describe virus lineages. However, virtually without exception, metagenomic discovery depends on detecting homology with known viruses: almost all virus-discovery pathways currently rely on blast-like sequence similarity searches. One alternative is to utilise immune-mechanisms to identify potential viruses, trusting the host to recognise viruses that we cannot.
The antiviral RNA interference pathway of arthropods is an ideal marker for viruses, as the immune 'signature' of 20-23nt small RNAs is itself directly available from metagenomic sequencing. Using this approach, we have previously identified a number of 'candidate viruses' and virus sequences from the fruit fly, *Drosophila*. Around half of these candidates are now identifiable as viral on the basis of subsequent discoveries, but other siRNA 'candidates' are still of unknown origin. Among this latter group we have recently identified a new lineage of segmented single-stranded RNA viruses, whose polymerase shows barely-detectable sequence-level homology with that of Picornaviridae, Caliciviridae and Flaviviridae. Related (but unannotated) sequences are detectable in the transcriptomes of around 20 other arthropods, and can be present at extremely high levels.

Evolutionary processes and consequences of animal and plant domestication
SMBE-OR-064
Genomics of domestication in Africa
Y. Vigouroux*

Abstract: While there has been progress in our understanding of the origin and history of agriculture in sub-Saharan Africa, a unified perspective is still lacking on where and how major crops were domesticated in the region. Using whole-genome resequencing and statistical models, we investigated the domestication of African rice (Oryza galberrima), pearl millet (Cenchrus americanum) and African yam (Dioscorea rotundata), a set of key crops of early African agriculture. We identified convergence for the selection of key genes during Asian and African rice domestication. Our results supported also the hypothesis that the vicinity of the Niger River was a major cradle of African agriculture.

Evolutionary processes and consequences of animal and plant domestication
SMBE-OR-065

The genomic architecture of fitness decline during cacao domestication

O. E. Cornejo [1,*], J. C. Motamayor [2]
[1]School of Biological Sciences, Washington State University, Pullman, [2]Cocal Cola Global, Orlando, United States

Abstract: Domestication has had a strong impact on the development of modern societies. We have sequenced 200 genomes of the chocolate plant *Theobroma cacao* L. to show for the first time to our knowledge that a single population, the Criollo population (n = 4), underwent strong domestication ~3600 years ago (95% CI: 2481–13,806 years ago). Our analyses have shown that domesticated populations of *T. cacao* (Criollo) maintain a higher proportion of high-frequency deleterious mutations. We further show the negative consequences of the increased accumulation of deleterious mutations during domestication on the fitness of individuals (significant reduction in kilograms of beans per hectare per year as Criollo ancestry increases, as estimated from a GLM, $P = 0.000425$). We inferred the fine scale recombination map of cacao populations and show that consistent with other observations, domesticated cacao populations have higher recombination rates. We finally analyze the architecture of the accumulation of deleterious variants in regions of the genome with differing rates of recombination to show that the increase of recombination does not seem to have a significant impact on the distribution of deleterious mutations in the domesticated cacao.

Evolutionary processes and consequences of animal and plant domestication
SMBE-OR-066

The domestication history of B. oleracea: wild relatives, geography of domestication, and multiple origins of kale

M. E. Mabry [1,*], E. Y. Gallagher [1], S. D. Turner-Hissong [2], J. Labate [3], M. A. Gore [4], J. C. Pires [1]
[1]University of Missouri, Columbia, [2]University of California Davis, Davis, [3]USDA-ARS, Plant Genetic Resources Unit, [4]Cornell University, Geneva, United States

Abstract: *Brassica oleracea* is an important crop species that has intrigued researchers for centuries. When first introduced to the species, Darwin drew many parallels between his theory on natural selection and the domestication practices that led to the varied forms of this plant. *Brassica oleracea* is unique in that it has been domesticated into several different crop types, including broccoli, Brussels sprouts, cabbage, kale, and kohlrabi, as well as several lesser well-known types, such as walking stick kale and marrow cabbage. However, over a century and a half after Darwin, we still know very little about this crop species. Based on uncertainty of the wild relatives of this species, several conflicting hypotheses on the geographic origin have been proposed. Some evidence suggests a Sicilian origin, as there is a large diversity of wild relatives there, while others suggest England as the location of origin, however these are more likely feral populations of once domesticated cabbage. Using a diversity panel of 225 accessions from 19 different *B. oleracea* crop types and eight species of potential wild relatives, we integrate phylogenetic and populations genetic techniques to examine patterns of relationships among domesticated types and wild relatives. These analyses point to the closest living relatives of *B. oleracea* as *B. incana* and *B. cretica*, indicating support for origin of cultivation in the Mediterranean region, and recover evidence for multiple origins of kales. Together, these results put us one step closer to further understanding the evolutionary history of these fascinating and important crops.

Evolutionary processes and consequences of animal and plant domestication
SMBE-OR-067
Reconstructing the history of breed development in the domestic dog through genomic analyses.
H. G. Parker [1,*]
[1]National Human Genome Reaserch Institute, National Institutes of Health, Bethesda, United States

Abstract: The domestic dog offers a treasure trove of genetic variation covering a wide range of traits including morphology, behavior, and disease. The vast majority of this variation is contained with the breeds. Breeds are populations of dogs that are maintained to meet a specific and unique standard of appearance and behavior. Genome-wide haplotype analysis of breed dogs has revealed the patterns of hybridization and selection that created many of the breeds found in the world today. Correlating the abundance of haplotype sharing with available in-depth pedigree reports places the majority of breed creation events within the last 200 years, coinciding with the breed explosion that came after the development of registrations and breed-based competition. Conversely, the lack of haplotype sharing combined with highly consistent clade membership corroborates historical accounts describing specific shapes and behaviors in dogs dating back over two-thousand years. Combining these data gives us a window into the early years of the dog where proto-breeds appeared based on geography and utility, which were then refined through hybridization and selection into today's dogs. To examine breed relationships, we have expanded our canine studies to include over 200 recognized breeds and have added populations from a wide range of countries and continents. We are developing methods to create a stable breed-tree framework that will allow us to determine the timing of events older than 200 years and narrow down the source of some of the most common variants in dogs. Finally, we can use breed relationships in combination with whole genome data to identify variants that have been under selection to create many of the breed-defining traits that we have come to expect in our dogs.

Evolutionary processes and consequences of animal and plant domestication

SMBE-OR-068

Evolutionary changes in sequence, regulation and gene expression between Bos taurus and Bos indicus

M. Naval Sanchez [1,*], M. Menzies [1], L. Porto-Neto [1], D. F. Cardoso [1 2], C. Kern [3], M. Halstead [3], M. R. Fortes [4], A. Canovas [5], B. J. Hayes [6], P. J. Ross [3], H. Zhou [3], J. Kijas [1], A. Reverter [1]

[1]Agriculture and Food, CSIRO, St. Lucia, Australia, [2]Department of Animal Sciences, School of Agricultural and Veterinarian Sciences, Sao Paulo State University, Jaboticabal, Brazil, [3]Department of Animal Science, University of California Davis, Davis, United States, [4]School of Chemistry and Molecular Biosciences, The University of Queensland, St. Lucia, Australia, [5]Centre for Genetic Improvement of Livestock, Department of Animal Biosciences, University of Guelph, Guelph, Canada, [6]Queensland Alliance for Agriculture and Food Innovation, The University of Queensland, St. Lucia , Australia

Abstract: The majority of world cattle population can be divided as Taurine (*Bos taurus*) or Indicine (*Bos indicus*). Taurine cattle are mostly based in temperate environments, while Indicine breeds are highly adapted to environments where heat is a constant. Besides temperature, disease and parasite resistance, differences in human herd management and selection processes have shaped separate patterns of genetic variation in both cattle lineages. With the aim to explore how selection and genomic differences impact the function of regulatory elements and gene expression, we performed a comparative analysis at the level of sequence, open-chromatin and gene expression between these two cattle lineages. In particular, the genome sequences of 185 Taurine and 103 Indicine individuals were analysed. Genome-wide detection of candidate selective sweeps based on site frequency measures (Fst) and nucleotide diversity (π) resulted in 657 and 242 20Kb bins in the Taurine and Indicine genomes, respectively. A major finding is a large selective sweep in the Indicine genome located on chr5: 47,670,001-48,100,000 and spanning several genes including *HELB*, *IRAK3*, *GRIP1* and part of *HMGA2*. We performed ATAC-seq for 3 biological replicates of Indicine and 2 replicates of Taurine breeds, for four tissues: muscle, adipose, liver and hypothalamus. We assess the enrichment of previously identified selective sweeps for regulatory features, and confirm that differences between breeds involve mostly coding and proximal regulatory elements. Currently, we are exploring the impact of selection at the gene expression level through RNA-seq for the same individuals and tissues as the ATAC-seq data was generated.

Evolutionary processes and consequences of animal and plant domestication
SMBE-OR-069
Genetic legacies from unknown wolf-like canid facilitated high altitude adaptation of modern gray wolves and dogs
M.-S. Wang*, S. Wang, R. Nielsen, D.-D. Wu, B. Shapiro

Abstract: Hybridization among Canid lineages is considered to have played a crucial role in their evolutionary histories. Here, we explore the impact of admixture among canid lineages endemic to the unique high altitude location of present-day Tibet, We analyzed 20 wolf and 143 dog whole genomes, including individuals from Tibet and neighboring lowland regions to catalogue the genomic landscape of recent and ancient admixture. We found evidence of significant admixture among Tibetan wolves and Tibetan dogs. While most admixed regions of the genome are evolving under purifying selection, *EPAS1,* a hypoxia-inducible factor involved in the physiological response to oxygen concentration, has been positively selected in both lineages. Intriguingly, the *EPAS1* haplotype found in both Tibetan dog and Tibetan gray wolf appears to have originated from neither of these lineages, but instead from a third as yet undescribed canid lineage. In fact, we found that ~38% of the nuclear genomic ancestry leading to Tibetan gray wolves, as well as their mitochondrial haplotype, originates from this undefined canid, which we estimate diverged from living Eurasian gray wolves ~880kya. Our study provides new insights into the evolutionary history of gray wolf and dog populations, and highlights a compelling example how admixture from divergent lineages can provide adaptive advantages to new immigrants into extreme environments like the Tibetan Plateau.

From the rings of life to the new animal phylogeny: James A. Lakes's contributions to evolutionary biology
SMBE-OR-058
The Origins of Eukaryotes
J. Mcinerney*

Abstract: Eukaryotes were formed by the merger of a Bacterial and an Eocyte lineage. This event has had several implications for the eukaryotic cell - we know that for most eukaryotes, a majority of their genes with prokaryotic provenance are from the Bacterial lineage, though we find that genes with Eocyte provenance are more likely to be more highly expressed, more central in protein interaction networks and more likely to be lethal upon deletion in yeast. We are also coming to a greater understanding of the stem lineage from the First Eukaryotic Common Ancestor (FECA) to the Last Eukaryotic Common ncestor (LECA). In this talk, I will expand on our discoveries in this area.

From the rings of life to the new animal phylogeny: James A. Lakes's contributions to evolutionary biology

SMBE-OR-060

Phylosystemics: merging phylogenomics, systems biology and microbial ecology to study evolution

E. Bapteste [1,*]

[1]ISYEB, CNRS, Paris, France

Abstract: This theoretical talk proposes a general approach to further tackle fundamental, seemingly intractable, evolutionary questions: can we enhance our knowledge of the evolution not only of organisms, but also of processes, in particular during major evolutionary transitions? In the same spirit as James Lake's works, which connected phylogenetic inferences to molecular, cellular, and developmental events, I propose to introduce and generalize phylosystemics, a new multidisciplinary strategy that unites the short timescale of interactions studies from systems biologists and microbial ecologists with the longest timescale of studies familiar to evolutionary biologists, by taking advantage of methods from network sciences. To do so, phylosystemics will superimpose evolutionary information, e.g., origin, transferability/duplicability, chimerism and divergence, on entities/edges forming interaction networks produced by systems biology and by (microbial) ecology. I will suggest that this approach could have far-reaching consequences, since it could provide i) new evidence regarding the evolution within single branches of the web of life, for instance during a major transition such as eukaryogenesis, from the first eukaryotic common ancestor to the last eukaryotic common ancestor, ii) an additional way to distinguish some lateral (and endosymbiotic) acquired genes from events of loss of ancient gene families in protists, and iii) an original framework to study the evolution of metabolic pathways in microbial communities and possibly the evolution of geochemical cycles.

From the rings of life to the new animal phylogeny: James A. Lakes's contributions to evolutionary biology
SMBE-OR-061

Empirical mixture models improve model-fit and mitigate branching errors in partitioned phylogenomics

A. K. Redmond [1,*], A. McLysaght [1]
[1]Smurfit Institute of Genetics, Trinity College Dublin, Dublin, Ireland

Abstract: Concatenated phylogenomics is typically performed by either partitioning a dataset and modelling under site-homogenous amino acid substitution models, or by analyzing the entire dataset under CAT/GTR-CAT, a single site-heterogeneous infinte mixture model. Results can be incongruent between these approaches however, and their relative fit cannot be easily measured. This has led to debate on whether sponges or comb jellies are sister to all other animals. Here, we incorporate empirical mixture models, including derivations of CAT designed for single-gene phylogenetics, into partitioned phylogenomics. Using real datasets as examples, we show that this greatly improves model-fit and can relieve branching artefacts induced by systematic error in partitioned phylogenomic analyses. Our results suggest that using only site-homogeneous models drastically underappreciates the prevalence and extent of intra-gene site-heterogeneity, and can underestimate the number of hidden substitutions, even in alignments with >90% average identity. By applying better-fitting empirical mixture models to reanalyses of key datasets previously found to support comb jellies as sister to all other animals, we find that partitioned phylogenomics does not provide unequivocal support for this hypothesis of animal evolution. Furthermore, underlying support, in the form of partition-specific log-likelihoods, shifts towards sponges as sister to all other animals as model-fit improves. In summary, we suggest that the use of empirical mixture models is advisable in future studies using partitioned phylogenomics, and that support for comb jellies as sister to all other animals is likely artefactual.

From the rings of life to the new animal phylogeny: James A. Lakes's contributions to evolutionary biology
SMBE-OR-059
Three decades of progress using molecules to reconstruct the animal tree of life.
M. Telford*

Abstract: In the late 1980's the first credible molecular analyses of metazoan phylogeny appeareed. These were based on sequences from the small subunit ribosomal RNA gene (SSU). These studies began a transformation in our understanding of the history of animal evolution. I will consider the main areas of progress in data collection and in methods of analysis over the last three decades that have resulted in our current understanding of animal phylogeny and the challenges that we still face today.

From the rings of life to the new animal phylogeny: James A. Lakes's contributions to evolutionary biology
SMBE-OR-062

Estimating the divergence date of eukaryotes via the fossilised birth-death process

H. Betts [1,*], J. E. O'Reilly [1], T. A. Williams [1], P. C. J. Donoghue [1], D. Pisani [1]
[1]University of Bristol, Bristol, United Kingdom

Abstract: Eukaryotes comprise multiple multicellular clades as well as boasting a wide microbial diversity. The oldest crown eukaryote fossil is the red alga *Bangiomorpha pubescens* sitting at ~1.2 Billion years ago. However, the fossil record of acritarchs, probable eukaryotes with unknown affinities, stretches back into the late Palaeoproterozoic suggesting an earlier evolution of the last eukaryotic common ancestor (LECA). It is well known that fossils can never tell us when a clade evolved given the time lag needed to produce identifiable features. Hence, we can turn to molecular clocks which utilise both the fossil record and modern genetic data providing an integrative way to glean information about such divergences. Previously, node dating - where fossils are assigned to specific nodes in the tree - has been used to estimate the divergence of crown eukaryotes. Here, we used the fossilised birth death process which removed the need for setting priors on each fossil node calibration and allowed the incorporation of additional fossil material. We took a sample from the diversity of eukaryotic fossil data available on the Palaeobiology Database and added this to a eukaryote tree covering each of the major lineages. Previous estimates show that crown Eukaryota evolved in the Palaeo-Neo Proterozoic. Our results slightly exceed these estimates, pushing LECA back into the Palaeoproterozoic, but still with an outcome much younger than the great oxidation event, something which has been potentially linked with evolution of eukaryotes.

From the rings of life to the new animal phylogeny: James A. Lakes's contributions to evolutionary biology
SMBE-OR-063
From Ribosome Structure to Ancient Life
J. A. Lake*

Abstract: Reflections on my career and on rocking scientific boats with a lot of help from my friends.

The genetic architecture of polygenic adpatation: sweeps, small shifts and everything in between
SMBE-OR-198
Reconciling population and quantitative genetics perspectives on local adaptation: the importance of redundancy
S. Yeaman*

Abstract: Local adaptation involves a tension between divergent natural selection and the homogenizing effect of migration. In one of tne of the earliest models in population genetics, Haldane showed how a locally adapted allele would tend to be maintained as long as selection was strong relative to migration. Thus, large effect alleles with big differences in allele frequency between populations should be expected under local adaptation, especially when migration is high. However, even when migration rates are strong relative to selection, local adaptation can still persist at the phenotypic level, driven by small changes in allele frequency at many loci, which tend to be transient in their contributions to divergence. When this happens, the causal loci tend to be very difficult or impossible to detect using genome scans. Genetic redundancy is one of the main factors affecting which kind of architecture tends to evolve under migration-selection balance. If there are many ways to make a given phenotype, then highly polygenic responses with transient effects can be common, but if redundancy is limited, then such responses can be effectively constrained by migration. Here, I review theoretical models of local adaptation to explore the importance of redundancy, and present some new empirical approaches to quantifying redundancy by studying the repeatability of adaptation.

The genetic architecture of polygenic adpatation: sweeps, small shifts and everything in between

SMBE-OR-199

Adaptation dynamics of a polygenic trait

K. Jain [1,*]

[1]Jawaharlal Nehru Centre for Advanced Scientific Research, Bangalore, India

Abstract: Although many phenotypic traits are determined by a large number of genetic variants, how a polygenic trait adapts in response to the changes in the environment is still not well understood. While polygenic adaptation via small to moderate changes in allele frequencies at many loci is described by quantitative genetics, selective sweeps wherein large shifts in the allele frequencies occur at a few loci has been investigated within the framework of population genetics. Thus two qualitatively different modes of adaptation have been proposed, and there has been a general disconnect between these conflicting theories of adaptation. I will describe some recent work on a quantitative genetic model of polygenic adaptation that accommodates both subtle shifts and selective sweeps. I will address some key questions regarding the dynamics of the allele frequencies, and the mean and variance of a phenotypic trait when it is under stabilizing selection.

The genetic architecture of polygenic adpatation: sweeps, small shifts and everything in between

SMBE-OR-200

Cis and trans-regulatory variation and purifying selection are key drivers of the massively polygenic architecture of complex traits.

J. Pritchard [1,*], X. Liu [2], Y. Li [3]

[1]Biology and Genetics, Stanford University, Stanford, [2]Human Genetics, [3]Genetic Medicine, University of Chicago, Chicago, United States

Abstract: Early genome-wide association studies led to the surprising discovery that, for typical complex traits, most of the heritability is due to huge numbers of common variants with tiny effect sizes. Previously, we argued that new models are needed to understand these patterns. Here we provide a formal model in which genetic contributions to complex traits are partitioned into direct effects from core genes, and indirect effects from peripheral genes acting in trans. We propose that most heritability is driven by weak trans eQTL SNPs, whose effects are mediated through peripheral genes to impact expression of core genes. In particular, if the core genes for a trait tend to be co-regulated then the effects of peripheral variation can be amplified such that nearly all of the genetic variance is driven by weak trans effects.

A second key component of the model is to understand the role that polygenic stabiizing selection may play in shaping cis- and trans-genetic architecture. Here we show that under certain conditions, stabilizing selection can place an upper bound on the contribution that any single gene makes to trait heritability--thereby limiting the contributions that the biologically most important genes make to heritability. We present data analysis to measure the importance of this flattening effect in real traits. In summary, our model proposes a framework for understanding key features of the architecture of complex traits.

The genetic architecture of polygenic adpatation: sweeps, small shifts and everything in between
SMBE-OR-196

Tracking allele trajectories over 20 generations of selection for long limbs in mice

L. Hiramatsu [1,*], S. Belohlavy [2], M. Kučka [1], C. Rolian [3], N. H. Barton [2], Y. F. Chan [1]
[1]Friedrich Miescher Laboratory of the Max Planck Society, Tuebingen, Germany, [2]IST Austria, Klosterneuburg, Austria, [3]University of Calgary, Calgary, Canada

Abstract: Pedigreed selection experiments offer a powerful approach to directly track how genomes evolve under selection. An outstanding example is found in the Longshanks experiment, in which mice in two replicate lines were bred over 20 generations for longer tibiae relative to body mass, resulting in 13-15% increase. We build on our recently published dissection of this selection experiment to investigate the interplay between genetic drift, selection, and the underlying genetic architecture during response to selection. Combining low-pass and linked-read sequencing, we captured allele frequency changes and haplotype segregation across 20 generations. Our time-series reconstruction of allele trajectories allowed estimation of selection and dominance parameters under varying selection or drift models for every locus. Besides discovering additional candidate loci, we also leveraged the time-series information to determine the most likely mode of inheritance governing these loci. We will summarize the relative importance of different allelic interactions in the Longshanks selection response and discuss broader implications for adaptive evolution. Finally, we will present our fine-scale reconstruction of haplotype segregation across the pedigree to investigate how linkage and recombination and selection interact in an empirical, controlled and replicated setting. We show here that despite small population sizes, selection response is rapid and robust in the short term. With nearly complete phenotypic, genomic, and pedigree datasets, the Longshanks experiment provides a comprehensively detailed system to study adaptive genomic evolution.

The genetic architecture of polygenic adpatation: sweeps, small shifts and everything in between
SMBE-OR-197

Evaluating evidence for polygenic adaptation using GWAS cohorts from multiple ancestries

A. Refoyo-Martinez [1,*], F. Racimo [1], A. Martin [2], X. Jin [3], S. Liu [3], A. Albrechtsen [4]
[1]Geogenetics, Centre for Geogenetics at the University of Copenhagen, Copenhagen, Denmark, [2]Analytic and Translational Genetics Unit, MASSACHUSETTS GENERAL HOSPITAL & BROAD INSTITUTE, Boston, MA, United States, [3]Human genomics, Beijing Genomics Institute, Shenzhen, China, [4]Bioinformatics, University of Copenhagen , Copenhagen, Denmark

Abstract: Summary statistics from genome-wide association studies (GWAS) data have been utilized to detect and quantify polygenic adaptation for human complex traits. Previous studies have reported signatures of natural selection at sets of SNPs associated with complex traits such as height or BMI. However, some of these signals have lately been suggested to be caused by biases from population structure. Moreover, past studies have predominantly relied on SNP effect size estimates obtained from GWAS panels of European ancestry, which are known to be poor predictors of phenotypes in non-European populations. Here, we collated GWAS data from multiple anthropometric and metabolic traits that have been measured in more than one cohort around the world, including the UK Biobank, Finrisk, Chinese NIPT, Biobank Japan and PAGE. We then evaluated how robust signals of polygenic adaptation are to the choice of GWAS cohort used to identify associated variants and their effect size estimates, while using the same panel to obtain population allele frequencies (The 1000 Genomes Project). We observe many discrepancies across tests performed on the same phenotype. For example, global signals of polygenic adaptation on height-associated SNPs are strongly attenuated when using East Asian biobanks, relative to European biobanks. In general, we find that the more differentiated the members of two biobanks are, the less consistent tests of polygenic adaptation will be, which is consistent with publications reporting poor polygenic score prediction across continents. We are currently working on distinguishing among different explanations for these discrepancies, including differences in genetic architecture and population structure across cohorts, adaptive dynamics and gene-by-environment interactions.

Genetic conflicts in molecular evolution
SMBE-OR-158
GENETIC CONFLICT SHAPES DROSOPHILA TELOMERE EVOLUTION
M. Levine [1,*], B. Saint-Leandre [1], S. Nguyen [1]
[1]University of Pennsylvania, Philadelphia, United States

Are you a member of SMBE?: No
Poster Submission: Virtually all eukaryotes rely on telomerase to maintain chromosome length. Fruit flies are a widely studied exception. Instead of telomerase, Drosophila relies on domesticated retrotransposons that insert exclusively into telomeric DNA. Although hailed as an exemplary "genomic symbiosis," sporadic accounts of rapid evolution of both the telomere packaging proteins and the transposable element arrays to which they localize implicate instead intra-genomic conflict. Consistent with a molecular arms race model, we recently reported pervasive positive selection and recurrent gene turnover of Drosophila telomere protection genes. Here we report that telomere-specialized retrotransposons, too, evolve rapidly across only a few million years of evolution. Combining computational methods and cytogenetics, we discovered rapid diversification of telomere specialized retrotransposons, sporadic telomeric retrotransposon "escape" into non-telomeric locations, and one instance of specialized retrotransposon disappearance. The dynamic evolution of both telomeric sequence and telomere-localizing proteins suggest that sustaining retrotransposon-mediated telomere length requires a recurrently re-negotiated truce between host proteins and chromosome lengthening elements. To experimentally test this model, we leverage CRISPR/Cas9-mediated transgenesis to cleanly swap into D. melanogaster diverged versions of telomere integrity genes. For one fast evolving telomere end protection gene, a close relative's diverged allele fully complements the published mutant phenotype—we observe no catastrophic end-to-end chromosome fusions. However, experimental evolution of this genotype causes elevated telomere-specialized retrotransposon insertions. Moreover, this long-telomere genotype suffers a female fertility cost. These data are consistent with an ongoing molecular arms race shaping Drosophila telomere biology and reveals the genetic determinants of molecular domestication (and re-domestication) over evolutionary time.

Genetic conflicts in molecular evolution
SMBE-OR-159
THE P-ELEMENT INVASION OF DROSOPHILA SIMULANS
A. Betancourt [1,*]
[1]University of Liverpool, Liverpool, United Kingdom

Are you a member of SMBE?: No
Poster Submission: Transposable elements are widespread genomic parasites, and the archetypal example of a selfish gene, which can impose large fitness costs on their hosts. The key to their long-term persistence may therefore be the ability to spread between hosts, as exemplified by the invasion of Drosophila melanogaster by the P-element. This transposable element originated in a distant relative and spread through D. melanogaster in the latter half of the 20th century. Recently, we discovered a second invasion of a Drosophila species by the P-element. We find that the P-element spread through D. simulans rapidly and nearly simultaneously on three continents, with strains containing P-elements being rare in 2006 and common by 2014. Remarkably, the flies appear to have evolved to adapt to the presence of the P-element in this short time frame: fly strains collected from the early phase of this invasion are vulnerable to DNA damage from the P-element, while those from the latter phase are not. We investigate the genetic basis of this resistance, and find it appears to have little to do with the small RNA defence usually invoked in transposable element suppression.

Genetic conflicts in molecular evolution

SMBE-OR-160

Red queen and the lost boys: The role of RNAi in defending sex chromosome conflict

J. Vedanayagam [1,*], C.-J. Lin [1], C.-M. Lai [1], R. Dubruille [2], B. Loppin [2], E. C. Lai [1]

[1]Department of Developmental Biology, Sloan-Kettering Institute, New York, United States, [2]Laboratoire de Biométrie et Biologie Évolutive, Université Claude Bernard Lyon 1, Lyon, France

Abstract: Selfish meiotic drive systems (SMDs) that distort equal sex-ratio (SR) among progeny are known to be ubiquitous in nature. However, the molecular underpinnings of how SMDs distort SR, and how subsequent restoration of balanced SR is achieved, remain vastly mysterious. In our recent study, we unraveled a crucial endogenous role of hairpin-RNA(hpRNA)/RNAi pathway in taming SMDs in the non-model fly *D. simulans*, where two distinct SMDs are known to act in testis. The Winters suppressor not much yang (*Nmy*) encodes endo-siRNAs that can suppress the *de novo* genes Distorter on the X (*Dox*) and its progenitor Mother of Dox (*Mdox*) in trans. Curiously, we also uncovered overlapping *Nmy* endo-siRNAs at a region characterized as Durham sterility factor Too much yin (*Tmy*), revealing an evolutionary link between the Winters and Durham SMDs. In our current efforts, we are employing CRISPR/Cas9 and introgression approaches to obtain the now lost *Tmy* mutant to examine the complex distortion network inferred from our previous study. Furthermore, from exploring the transcriptome data from our CRISPR/Cas9 mutants in the core RNAi factors *Ago2* and *Dcr-2,* we unexpectedly uncovered an array of *de novo* hpRNAs that were born in the simulans clade and preferentially target X-chromosome genes, some of which may have likely distorter activity alongside *Dox*/*Mdox*. Broadly, we are seeking a functional understanding of these recently-emerged X-chromosome genes. Our findings support widespread and recent sex chromosome conflicts within *D. simulans* and/or simulans clade species and highlight the role of RNAi in resolving genetic conflicts.

Genetic conflicts in molecular evolution
SMBE-OR-161
The Molecular and Genetic Basis of Hybrid Incompatibility
J. C. Cooper*

Abstract: For two species to remain separate, their hybrid offspring must be less fit than either parent. This fitness decrease must come from an interaction of two wild-type parental genomes. Despite the identification of several hybrid incompatibility (HI) genes, there is almost no understanding of how dysfunction at the molecular level leads to hybrid sterility and inviability. To address this gap in understanding, my work has focused on hybrid male inviability in *Drosophila melanogaster* and *Drosophila simulans.* Though this is one of the oldest systems for studying the evolution of HI, the two known HI genes were not sufficient to explain the molecular mechanism of HI nor the evolutionary processes that drove isolation. I developed two new methods to map new hybrid incompatibility genes in this system, both of which highlight the power of combining whole genome sequencing with classic mapping schemes. Using these methods I uncovered a third and forth HI locus for this system. From this, I discovered that the massive variation observed in hybrid rescue between species is due to variation at only two loci. Going beyond identification, I have subsequently used genetics and molecular approaches to study the properties of the known HI genes. This work has implicated new processes at the center of hybrid incompatibilities, including transposable element regulation and chromatin boundary control. My work extends understanding of the evolutionary genetics of hybrid incompatibilities and builds toward a functional molecular model to explain hybrid inviability.

Genetic conflicts in molecular evolution

SMBE-OR-162

Evolutionary epigenetic effects of transposable elements in 3D nuclear space

G. Y. C. Lee [1 2,*], Y. Ogiyama [3], N. Martins [4], B. Beliveau [5], D. Acevedo [1], C.-T. Wu [4], G. Cavalli [3], G. Karpen [1]

[1]Lawrence Berkeley National Laboratory and University of California, Berkeley, Berkeley, [2]University of California, Irvine, Irvine, United States, [3]CNRS, Montpellier, France, [4]Harvard Medical School, Boston, [5]University of Washington, Seattle, United States

Abstract: Transposable elements (TEs) are ubiquitous genome parasites whose evolution is tightly intertwined with the evolution of host genomes. They are abundant in pericentromeric heterochromatin (PCH), which is enriched for the repressive epigenetic mark H3K9me2/3 and its reader protein, HP1a. TEs are also prevalent in the gene-rich euchromatic genome and, interestingly, we previously showed that they can lead to tens of Kb enrichment of H3K9me2/3 and HP1a at flanking euchromatic sequences in natural *Drosophila* populations. Due to the biophysical properties of HP1a, PCH of different chromosomes can coalesce into a single domain within the 3D nuclear space through liquid-liquid fusion (PCH domain). This domain is enriched with silencing proteins, which can significantly influence the function of genes that are brought into this 3D space. We hypothesized that euchromatic TEs enriched for H3K9me2/3 and HP1a could also spatially interact with the main PCH domain, influencing euchromatic genome function and ultimately genome evolution. To investigate the spatial contacts of euchromatic loci with PCH, we developed a novel analysis method for Hi-C, a high throughput approach for studying 3D genome organization. Despite being far from PCH on a linear chromosome, ~14% euchromatic TEs show 3D associations with PCH and are thus brought into the PCH domain enriched with silencing proteins. By leveraging polymorphic (presence/absence) TE insertions in natural *Drosophila* populations, we compared the 3D organization of homologous sequences with and without TE-induced H3K9me2/3 enrichment and demonstrated that the observed spatial contacts between euchromatic loci and PCH are indeed caused by TEs. Importantly, population genetic analysis revealed that TEs spatially interacting with PCH are more strongly selected against, suggesting functional consequences of 3D contacts with PCH. Our findings demonstrate that naturally occurring TEs could generate polymorphic 3D genome organization within and between species, having a far-reaching impact on the function and evolution of the gene-rich euchromatic genome.

Genetic conflicts in molecular evolution

SMBE-OR-163

An unprecedently powerful approach to measure biased gene conversion in mouse recombination hotspots provides evidence for the rapid evolution of the mismatch repair machinery leading to segregation distortion of weak (A or T) VS strong (G or C) alleles within the mammalian clade

M. Gautier [1,*], F. Baudat [2], N. Lartillot [1], B. de Massy [2], L. Duret [1]

[1]Laboratoire de Biométrie et Biologie Evolutive, CNRS, Villeurbanne, [2]Institut de Génétique Humaine, CNRS, Montpellier, France

Abstract: Meiotic recombination involves the formation of a DNA double-strand break (DSB) repaired using the homologous chromosome as a template. Thus, any variation between parental genomes appears as a mismatch when the two homologues pair. In a wide range of species, the repair of these mismatches is biased towards the favoured transmission of strong (G and C) alleles, a process named GC-biased gene conversion (gBGC). This fundamental distorter of allelic segregations during meiosis mimics the action of positive selection and plays a major role in the evolution of base composition in the vicinity of recombination hotspots.

We coupled sperm-typing, advanced sequencing and bioinformatic techniques to implement an innovative approach allowing to detect recombination events with an efficiency 100 times greater than that of current state-of-the-art methods.

We next measured the intensity of gBGC (*b*) in over 1,000 recombination hotspots. Notably, we observed that the contribution of very short non-crossover (NCO) events to *b* largely prevailed that of both longer NCOs and crossing-overs (COs), contrary to humans for which all three classes contribute equally. The repair machinery leading to segregation distortion thus evolved extremely rapidly within the mammalian clade. This work also evidences new relationships between the intensity of gBGC and other parameters of genome evolution, such as the recombination rate and the effective population size, hence suggesting that a second-order selection restrains the intensity of this deleterious process at the population-scale.

Overall, our work allows to better comprehend how the mismatch repair machinery evolved and shaped genomic GC-content through meiotic drive.

Genetic conflicts in molecular evolution

SMBE-OR-165

Sexual conflict drives the rapid coevolution of abalone sperm-egg recognition proteins creating species-specific binding and fertilization leading to speciation

J. Carlisle [1,*], J. Aagaard [1], D. Wilburn [1], W. Swanson [1]

[1]Genome Sciences, University of Washington, Seattle, United States

Abstract: Reproductive proteins mediating sperm-egg interactions are characterized by accelerated evolution. Strong selective pressure to maintain successful fertilization paired with differences in optimal male/female reproductive strategies can promote arms race dynamics that drive the rapid coevolution of proteins mediating sperm-egg recognition. This rapid coevolution could create molecular boundaries to hybridization and contribute to speciation. However, while many examples of protein divergence have been identified, an understanding of the functional consequences of reproductive protein diversification and its potential role in creating species boundaries remains to be established. As one of only a handful of known interacting sperm-egg protein pairs, abalone's lysin-VERL interaction is an important model for understanding the molecular mechanisms mediating fertilization. During abalone fertilization, sperm lysin dissolves the vitelline envelope of the egg by binding repetitive domains of the vitelline envelope receptor for lysin (VERL). Vitelline envelope dissolution by lysin is species specific, presumably due to the rapid coevolution of lysin and VERL, driven by sexual conflict, creating differences in conspecific and heterospecific binding affinities. Using surface plasmon resonance, a highly sensitive biochemical technique, I have quantitatively measured the binding affinities between heterospecific and conspecific lysin-VERL domain pairs and have detected species-specific binding between these rapidly evolving proteins. This is the first study to quantitatively investigate sperm-egg binding interactions between both conspecific and heterospecific proteins pairs and experimentally supports the hypothesis that weaker heterospecific binding of rapidly evolving sperm-egg recognition proteins can create boundaries to speciation by interfering with zygote formation.

Genome-wide methods for detecting selection

SMBE-OR-126

Detecting selection beyond sweeps: machine learning approaches for detecting balancing selection and comparing multiple modes of selection

K. Ahlquist [1 2,*], S. Ramachandran [1 3], L. A. Sugden [1 3]

[1]Center for Computational Molecular Biology, [2]Molecular Biology, Cell Biology and Biochemistry, [3]Ecology and Evolutionary Biology, Brown University, Providence, United States

Abstract: Most tools used for detecting selection in population genomic data focus on selective sweeps. Nearly all methods of detecting selection are developed to detect just a single mode of selection. However, the genome contains many sites, each experiencing different modes of selection, including positive selection, balancing selection, neutral evolution and more. Identifying the mode of selection acting at a genomic site means not only identifying patterns associated with a single type of selection, but also testing alternative explanations for the patterns observed. Previous work from our group (Sugden et al., Nature Communications, 2018) developed the SWIF(r) framework, which uses averaged one-dependence estimation to learn the pairwise joint distributions of summary statistics for selection and calculate the per-site probability of a selective sweep. Here, we extend this supervised classification framework to detect balancing selection in both simulated data and human genomes, and compare the probabilities of different modes of selection acting at a given site. In simulated data our approach outperforms state-of-the-art methods in detection of balancing selection signals, achieving an 81% true positive rate at a false positive rate of 1%. SWIF(r) calculates an interpretable site-based probability for each mode of selection tested, making comparison between multiple hypotheses of modes of selection possible. These advances will (1) improve our overall understanding of the footprint of selection in shaping human genetic variation, and (2) allow researchers to disentangle the histories of complex regions in the genome that may have experienced multiple selection types.

Genome-wide methods for detecting selection

SMBE-OR-127

Joint inference of divergent selection and demographic history in a two-deme model with isolation and migration

S. Aeschbacher [1,*], T. Grubinger [1], K. Lohse [2]

[1]Department of Evolutionary Biology and Environmental Studies, University of Zurich, Zurich, Switzerland, [2]Evolutionary Biology, University of Edinburgh, Edinburgh, United Kingdom

Abstract: Demography may cause patterns of genomic variation that mimic the signature of natural selection. Vice versa, natural selection introduces biases to demographic inference from genomic data. Breaking this vicious circle requires methods for jointly inferring demography and selection. Here, we present a population-genomic approach for quantifying the amount and timing of gene flow as well as the strength of divergent selection in a two-deme model with isolation and gene flow. We use the distribution of pairwise sequence differences within and between species (populations) to first infer effective population sizes, the split time, and effective migration rates based on genomic blocks previously stratified by recombination rate. We then use the relationship between effective migration rate and recombination rate to estimate the strength of divergent selection in the face of gene flow. We validate this approach using simulations, and illustrate it using whole-genome polymorphism and recombination data from *Mimulus guttatus* and its selfing sister species *M. nasutus*. Focussing on an area of sympatry in the northern range of the species' distributions in Western North America, we identify a strong signal of isolation after an initial split about 860,000 years ago, followed by recent secondary-contact gene flow from *M. nasutus* into *M. guttatus*. We find that a selection coefficient on the order of a few percent per megabase pair maintains the species barrier despite recent or on-going gene flow at a scaled rate of about 1.5 migrants per generation. We anticipate that our approach will benefit genome-scale studies of speciation and local adaptation.

Genome-wide methods for detecting selection
SMBE-OR-124
SELECTION ON GENOMIC TANDEM REPEATS
M. Anisimova [1,*]
[1]Zürich University of Applied Sciences (ZHAW), Zurich, Switzerland

Are you a member of SMBE?: No
Poster Submission: Genomic sequences are shaped by a complex interplay of various biological factors, among which natural selection plays an important role. Selection operates in many different ways, not only on the level of point mutations but also on genomic features. In particular, tandem repeats (TRs) are found abundantly in genomic sequences across all domains of life. Evidence suggests that some TRs are crucial for proteins with fundamental biological functions, and can be associated with virulence, resistance and infectious/ neurodegenerative diseases. Genome-scale systematic studies of TRs have the strong potential to unveil core mechanisms governing TR evolution and TR roles in shaping genomes. Indeed, a growing body of evidence suggests that selection may operate also on tandem repeats. Here I will review methods for analysing tandem repeats that explicitly rely on the evolutionary definition of a tandem repeat as a sequence of adjacent repeat units stemming from a common ancestor. This allows to define a powerful framework for statistical annotation and evolutionary analysis of TRs, potentially identifying TRs evolving under selective pressures. The discussed work has a focus on protein TRs, yet is generally applicable to nucleic acid TRs.

Genome-wide methods for detecting selection
SMBE-OR-128
Detecting selection with fluctuating population size phylogenetics codon models.

T. Latrille [1,*], N. Lartillot [1]
[1]Laboratoire Biométrie et Biologie Evolutive, Lyon, France

Abstract: Selection in protein-coding sequences can be detected based on multiple sequence alignments using phylogenetic codon models. Mechanistic approaches, grounded on population-genetics first principles, have been recently developed. These so-called mutation-selection models explicitly formalize the interplay between mutation, selection and drift, and return an estimate of the amino-acid fitness landscape, considered static along the phylogeny. They were recently proposed as a null (nearly-neutral) model against which to test for the presence of adaptation (Rodrigue, Lartillot MBE 2016, Bloom, 2016). However, these models rely on the assumption of multiplicative fitness landscapes (no epistasis) and constant population size; they also ignore polymorphism in extant species, with only one sequence representing the whole population. As a result, they return potentially biased estimates. We propose an extended mutation-selection model relaxing some of these assumptions, by accommodating for fluctuating population size and fluctuating mutation rate along the phylogeny, and by modeling polymorphism in extant species. The resulting mechanistic framework allows for a reconstruction of long-term trends in population size along the phylogeny. Simultaneously, it offers a better background for detecting adaptation across large clades, by correcting for local changes in the relative strength of selection and random drift. Finally, our work also points to important theoretical questions about how coding sequences respond to changes in effective population size and to selection.

Genome-wide methods for detecting selection

SMBE-OR-129

Inference of complex scenarios of adaptive evolution during rapid species radiation

A. Salman Minkov [1,*], H. A. Hejase [2], M. Hubisz [3], L. Campagna [4], I. J. Lovette [5], A. Siepel [2], I. Gronau [6]

[1]Efi Arazi School of Computer Science, The interdisciplinary center (IDC) Herzliya, Herzliya, Israel, [2]Simons Center for Quantitative Biology, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY, [3],Department of Biological Statistics and Computational Biology, Cornell University, Ithaca, NY, [4]Fuller Evolutionary Biology Program, Cornell Laboratory of Ornithology, Ithaca, NY, [5]Fuller Evolutionary Biology Program, Cornell Laboratory of Ornithology, Ithaca, NY, United States, [6]Efi Arazi School of Computer Science, Herzliya Interdisciplinary Center, Herzliya, Israel

Abstract: Species radiation provides us with a great opportunity to study the role of adaptation in shaping the landscape of genome sequence variation. However, the task of detecting genomic regions under positive selection is significantly complicated by population-specific selective forces as well as gene flow that takes place between emerging species. Conventional methods for detecting selective sweeps typically look for genomic regions with high levels of divergence from a closely related outgroup species and low levels of population diversity. This approach works well when studying a single speciation event, but it does not capture complex scenarios that occur in species radiation, such as selective sweeps shared between species, or parallel selective sweeps in different species. Thus, we developed a method for analyzing complex scenarios of selective sweeps that makes use of computationally inferred genetic genealogies together with a linear SVM classifier. Our approach takes advantage of recent progress in computational methods for genome-wide inference of ancestral recombination graphs (ARGs).

As a case-study, we analyzed a data set of 60 whole genomes sequenced from five recently diverged bird species of southern capuchino seedeaters from the genus *Sporophila* (Campagna *et* al. 2017). These finch-like bird species live mostly in sympatry, show low levels of overall genetic differentiation, but clear signs of differentiation in phenotypes such as male plumage coloration and song. We inferred genome-wide ARGs for this data set using ARGweaver (Rasmussen *et* al. 2014), and then used the inferred ARGs to examine detect and distinguish between selective sweeps in different species-combinations: (1) hard sweeps that include nearly all individuals from a given species; (2) partial sweeps that cover a sub-population within a species; (3) mixed sweeps that cover a population from two or more species; (4) parallel sweeps that occur in different directions in two or more different populations in the same locus. Using a linear SVM trained on simulated data we show that most divergence between these species is explained by soft sweeps. Moreover, using coalescent ages inferred by ARGweaver, we estimate that the great majority of sweeps are very recent. Interestingly, we do find some divergent loci that appear to have diverged much earlier in the evolution of these species. These deeply diverged loci indicate possible barriers to gene flow that emerge in these young species.

Genomic perspectives on plant and animal domestication
SMBE-OR-039
Pet dogs, citizen science, and the genetics of domestication
E. K. Karlsson [1,*]
[1]University of Massachusetts Medical School, Harvard, United States

Abstract: Pet dogs are an unparalleled genetic model for domestication. Over thousands of years, dogs have adapted to survive, and thrive, in human-modified environments, through changes in behavior, reproduction, and diet, among others. Furthermore, with the dog population now numbering in the millions, large-scale genomic studies with the statistical power to find genetic variants underlying polygenic traits are feasible. We have developed a citizen-science based approach to dog genetics that welcomes any dog, regardless of breed ancestry. In its first 3 years, Darwin's Ark (DarwinsArk.org) has enrolled more than 22,000 dogs, and their owners have cumulatively answered nearly 2.5 million survey questions. The first genomewide association studies, with 1400 DNA-sequenced dogs, finds genetic changes significantly associated with physical, behavioral, and food related traits. We are now pairing these results with our smaller-scale study of wolf-dog hybrids to identify changes also linked to domestication. By engaging directly with dog owners, Darwin's Ark is a powerful, open data resource for canine genomics, providing useful phenotypes and very large sample sizes. Paired with studies of non-domesticated and hybrid populations, it offers new insight on the biological pathways underlying one of the most important evolutionary transitions.

Genomic perspectives on plant and animal domestication
SMBE-OR-041

The Origins of Taurine Cattle: An Ancient Genomic Tale

V. E. Mullin [1 2,*], M. Pereira Verdugo [2], A. Scheu [3], D. G. Bradley [2]
[1]Earth Sciences, The Natural History Museum, London, United Kingdom, [2]Trinity College Dublin, Dublin, Ireland, [3]Johannes Gutenberg-University Mainz, Mainz, Germany

Abstract: The domestication of taurine cattle (Bos taurus) marks a significant period of time in human prehistory. Taurine cattle were domesticated in the Near East from their wild progenitor, Bos primigenius (aurochs), approximately 10,000 yrBP. Human mediated migration of taurine cattle into Europe started to occur c. 8,400 yrBP. Until now, the study of modern nuclear genomes and ancient mitogenomes have contributed to our understanding of the domestication of taurine cattle, subsequent migrations and admixture events. We have sequenced ~70 ancient cattle nuclear genomes from the Neolithic to the Middle Ages originating from the South West Asia and Europe. These ancient genomes provide a snapshot of the genetic diversity present in the past, allowing the timing of population events such as admixture and migration, that have shaped the genome of a globally important domesticate.

Genomic perspectives on plant and animal domestication

SMBE-OR-042

Ancient genomic structure of domestic dogs

A. Bergström [1,*], L. Frantz [2], R. Schmidt [3], E. Ersmark [4], L. Girdland-Flink [5], S. Amiri [6], G. Bar-Oz [7], J. Bulatović [8], S. Fedorov [9], I. Fiore [10], M. Germonpré [11], L. K. Horwitz [12], L. Janssens [13], Y. Kuzmin [14], R. Losey [15], M. Mashkour [16], V. Onar [17], D. Orton [18], M. Radivojević [19], M. Sablin [20], S. M. San José Santamaria [21], I. Stojanović [22], J. Storå [23], A. Tagliacozzo [10], K. Trantalidou [24], A. Villaluenga [25], A. Carmagnini [2], J. Haile [26], E. Irving-Pease [27], A. Linderholm [28], A. Götherström [23], L. Dalén [4], R. Pinhasi [29], G. Larson [27], P. Skoglund [1]

[1]The Francis Crick Institute, [2]Queen Mary University of London, London, United Kingdom, [3]University College Dublin, Dublin, Ireland, [4]Swedish Museum of Natural History, Stockholm, Sweden, [5]Liverpool John Moores University, Liverpool, United Kingdom, [6]University of Tehran, Tehran, Iran, Islamic Republic Of, [7]University of Haifa, Haifa, Israel, [8]University of Belgrade, Belgrade, Serbia, [9]North-Eastern Federal University, Yakutsk, Russian Federation, [10]Bioarchaeology Service, Museo delle Civiltà, Rome, Italy, [11]OD Earth and History of Life, Royal Belgian Institute of Natural Sciences, Brussels, Belgium, [12]The Hebrew University, Jerusalem, Israel, [13]University of Leiden, Leiden, Netherlands, [14]Institute of Geology and Mineralogy, Novosibirsk, Russian Federation, [15]University of Alberta, Edmonton, Canada, [16]CNRS, Paris, France, [17]Istanbul University-Cerrahpaşa, Istanbul, Turkey, [18]University of York, York, [19]University College London, London, United Kingdom, [20]Zoological Institute of the Russian Academy of Sciences, Saint Petersburg, Russian Federation, [21]Gordailua, the Gipuzkoa Heritage Collection Centre, Irun, Spain, [22]Institute of Archaeology, Belgrade, Serbia, [23]Stockholm University, Stockholm, Sweden, [24]Hellenic Ministry of Culture & Sports, Athens, Greece, [25]University of the Basque Country, Bilbao, Spain, [26]University of Copenhagen, Copenhagen, Denmark, [27]University of Oxford, Oxford, United Kingdom, [28]Texas A&M University, College Station, United States, [29]University of Vienna, Vienna, Austria

Abstract: The dog was the first animal to be domesticated, and has accompanied humans through major lifestyle changes and population movements. Yet very little is known about the population history of dogs and to what extent it reflects that of associated human populations. We sequenced 27 ancient dog genomes from the last 10,000 years to a median depth of 1.5-fold, including samples from the Near East, Siberia and across Europe. We find that the major lineages of dog ancestry cannot be related without invoking admixture events. Ancient West Eurasian dog genomes display a cline of dual affinities to early Near Eastern dogs and early Siberian dogs. Present-day African dogs share the majority of their ancestry with early Neolithic Near Eastern dogs. Modern European breeds display a remarkably uniform ancestry with structure among breeds uncorrelated to early European population structure, suggesting that they derive from a single, homogeneous source population. By co-analysing the dog genomes with ancient human genomes matched in time and space, we uncover several features of dog ancestry relationships that mirror known human population history, but also features where dogs display opposing trends. Overall, our ancient genomic data reveals that the divergent ancestries observed in European, Near Eastern, African, Siberian, East Asian, Oceanian, and American dog populations must have diversified already by 10,000 years ago.

Genomic perspectives on plant and animal domestication
SMBE-OR-040
Dynamic histories of adaptation and deterioration in plant domestication revealed through archaeogenomics
R. Allaby*

Abstract: The forces of selection involved in cereal domestication have varied over time in both strength and nature. It is becoming apparent that these forces of selection stretch back deep in time far beyond the onset of widespread cultivation. Using archaeogenomics it has become possible to gain insights into the way in which plants transitioned into domestication and the consequent impacts of that on their adaptive plasticity. A picture is emergent of domestication as a landscape scale process in which diversity was successfully brought wholesale in from wild populations. Adaptive episodes can be resolved which sequentially reveal the selection on fundamental domestication syndrome genes, environmental adaptations and later domestication improvement genes. The damage accrued in domesticated genomes through mutation load is not particularly associated with initial domestication, but with subsequent founder events and adaptive episodes reaching critically high levels in recent history. These insights both underline the importance of wild genetic diversity in rescuing core genomic functions of cultivars as well as improvement traits, and in some cases highlight how long term agricultural practices might be revised to support future sustainability.

Genomic perspectives on plant and animal domestication
SMBE-OR-043
Genomic Variation of Wine Grape after Phylloxera
H.-L. Liu [1,*], J.-M. Boursiquot [2], J.-P. Peros [3], M. Gaudeul [4], B. Shapiro [5], R. G. Allaby [1]
[1]School of Life Science, The University of Warwick, Coventry, United Kingdom, [2]Institut des Hautes Etudes de la Vigne et di Vin, [3]French National Institute for Agricultural Research, Montpellier, [4]Muséum national d'Histoire naturelle, Paris, France, [5]Ecology and Evolutionary Biology , University of California, Santa Cruz, Santa Cruz, United States

Abstract: As an endemic species, the wild and the domesticated form of the European grape（*Vitis vinifera ssp. sylvestris* and *ssp. vinifera*） are scattered across the Mediterranean region. In the 1850s, grape phylloxera （*Daktulosphaira vitifoliae*） was introduced into European vineyards and caused the Great French Wine Blight（GFWB）. Both domesticated grape *vinifera* as well as wild form *sylvestris* were decimated and leading to the establishment of grafting and nursery system in modern viniculture practice. North-American grape species *Vitis riparia* was also therefore introduced into Europe as a rootstock to assuage the agricultural damage.

Here, we carefully sequence and compare genomic content obtained from four European grape varieties, wild subspecies and rootstock species collected before and after the GFWB. This offers us the chance to track how genomic landscape evolve through time and allows us to explore whether artificial selection has re-shaped the genetic background of modern *V. riparia*. The reconstruction the somatic mutation rate and the potential regions under selection will allow us to infer the impact of the agricultural system on the genetic composition of a species. The finding will illuminate the critical role of introducing museum and herbarium collections for research into recent evolutionary change.

Genomic perspectives on plant and animal domestication

SMBE-OR-044

Ancient dispersal of Asian rice and abiotic forces underlying its distribution

R. M. Gutaker [1,*], E. Bellis [2], S. C. Groen [1], K. Bocinsky [3], J. d'Alpoim Guedes [4], J. R. Lasky [2], M. D. Purugganan [1]

[1]New York University, NEW YORK, [2]Pennsylvania State University, State College, [3]Montana Climate Office, Missoula, [4]University of California San Diego, La Jolla, United States

Abstract: Asian rice (*Oryza sativa*) was domesticated in the Yangtze Valley in East Asia. The onset of its cultivation corresponded to the warm Holocene climatic optimum between 9000-5000 years BP. Subsequently, rice dispersed to Northeast, Southeast and South Asia. Our knowledge of rice spread routes, the role of topological and climatic barriers in its dispersal is very limited. Here we reconstruct and time rice dispersal routes by analyzing over 1000 whole-genome sequences of Asian landraces, supplemented with analyses of migration barriers as well as abiotic factors that may impact geographic distribution of rice genomic variation. First, we genetically stratified rice populations and found they are geographically structured. We then elucidated relationships between rice populations using an admixture graph framework and established multiple waves of rice dispersal to Southeast and Northeast Asia. We reconstructed rice dispersal routes by mapping graph splits with archaeological data. Additionally, we associated genetic migration barriers with sea barriers and showed that these play an important role in shaping the geographic distribution of rice genomic variation. Our redundancy analysis showed that temperature, much more than any other abiotic variables, explains geographic distribution of japonica rice genomic variation. Finally, using a cross-coalescence method we found that the emergence of temperate japonica rice was a dynamic process that occurred between 4000-2500 years BP. The emergence of temperate japonica coincided with global cooling, and a probability drop for rice thermal niche in East Asia at around 4500 years BP. Our results show that genomic data are invaluable in studying crop histories and, in conjunction with geographic and climatic data, can elucidate factors responsible for the distribution of genomic diversity.

Inside Africa: Uncovering patterns of human genetic diversity

SMBE-OR-174

Selection, metabolism and resistance to infectious diseases in Africa

L. M. Pereira [1,*]

[1]Instituto de Investigação e Inovação em Saúde, Porto, Portugal

Abstract: The continuous characterization of genome-wide diversity in population and case-cohort samples, allied to the development of new algorithms, are shedding light on host ancestry impact and selection events on various complex diseases, namely infectious diseases. Pathogens and metabolites have been identified as the main selection motors in human evolution and, in many instances, their role is closely intertwined. We will illustrate such an example in dengue fever disease, which was identified when we conducted joint ancestry and association tests in an admixed Cuban cohort characterized for 2.5 million SNPs. We identified African-ancestry protection against the hemorrhagic phenotype through two genes intervening in lipid metabolism. Lipids are essential for the virus to enter in human cells and replicate therein. Functional tests have confirmed the involvement of these genes in dengue disease, and open up new avenues for the development of therapies. Dengue virus is not per se a significant selective motor, as its associated mortality rate is low. Other related viruses, such as yellow fever virus, could have been the drivers of the local African adaptation against several related infectious diseases. This African protection continued to favour African-descendants in the new world environment, when dengue virus was there introduced in the 19th/20th century.

Inside Africa: Uncovering patterns of human genetic diversity
SMBE-OR-176
Inferring deep population structure in Africa using linkage disequilibrium
A. P. Ragsdale [1,*], S. Gravel [1]
[1]Human Genetics, McGill University, Montreal, Canada

Abstract: Throughout history, populations have expanded and contracted, split and merged, and exchanged migrants. Because these events affect contemporary genetic diversity, we can learn about history by comparing predictions from evolutionary models to genetic data. We developed an approach to rapidly compute predictions for a wide range of diversity measures for many populations with complex demography, including for patterns of shared linkage disequilibrium between populations. These statistics are informative about deep population structure and archaic admixture, even when there is no available genetic data from diverge human lineages. Using this approach, we find evidence for substantial and long-lasting admixture from a deeply diverged lineage within Africa, and we infer detailed multi-population demographic models for a large set of diverse African populations. Our results underline the need for demographic models that better describe population structure within Africa, which can strongly affect predicted patterns of linkage disequilibrium and genetic diversity. More broadly, we highlight that by studying a wide variety of diversity statistics we can assess the robustness of commonly used evolutionary models and build more informed models of demographic history.

Inside Africa: Uncovering patterns of human genetic diversity

SMBE-OR-177

Sequencing hundreds of whole genomes from Bantu-speaking populations uncovers rare variation and informs on the recent demographic history of sub-Saharan Africa

S. Tallman [1,*], M. das Dores Sungo [2], F. C. Lubalo [2], S. Saranga [3], S. Beleza [1]

[1]Genetics & Genome Biology, University of Leicester, Leicester, United Kingdom, [2]Faculdade de Medicina, Universidade Onze de Novembro , Cabinda, Angola, [3]Faculdade de Educacao Fiscia e Desporto, Universidade Pedagógica, Maputo, Mozambique

Abstract: Recent research is beginning to catalogue the genetic diversity of Africa in increasing detail. However, further comprehensive sampling and generation of high-quality genomic data is required for a number of underrepresented populations. Today, a vast region of Africa spanning ~10,000,000km^2 and 27 countries is populatedlargely by speakers of Bantu languages; a distribution attributed to a massive diffusion of people that likely occurred after the development of agriculture in west-Africa approximately 4000-5000 BP. Indeed, the scale of this expansion was such that 1 in 3 Africans speak some variety of Bantu languages.

In order to support in the discovery of genomic variation across Bantu-speaking populations, we sequenced the genomes of 350 individuals (~12X mean coverage) from ethnolinguistic groups native to Angola and Mozambique. We show that, although genetic differentiation is low between Bantu-speakers, structure exists likely reflecting geographically-dependent admixture with autochthonous groups combined with the dynamics of population movements during and after the period of expansion; parameters of which are modelled using modern computational techniques. Furthermore, we describe millions of novel variants with the potential to improve our understanding of the genetic basis of medically and anthropologically relevant traits in these regions, as well as how variability in such traits may have been shaped by natural selection. Beyond this, we explore the viability of low-cost, ultra-low depth sequencing as a strategy for future genomics research in Africa by creating study-specific reference panels and sequencing a further 143 individuals to ~1X mean coverage.

Inside Africa: Uncovering patterns of human genetic diversity
SMBE-OR-175

The demographic et adaptive history of central African hunter-gatherers and farmers

E. Patin [1,*], P. Mouguiama-Daouda, M. Lopez, A. Kousathanas, J. Choin, M. Sikora, C. Harmant, H. Quach, J.-M. Hombert, A. Froment, G. H. Perry, L. B. Barreiro, C. D. Bustamante, P. Verdu, L. Quintana-Murci
[1]Institut Pasteur / CNRS, Paris, France

Abstract: Central Africa, a forested region that supports an exceptionally high biodiversity, hosts the world's largest group of hunter-gatherers, who live in close proximity with groups that have adopted agriculture over the past 5000 years. Our understanding of the prehistory of these populations has been dramatically hampered by the almost total absence of fossil remains in this region, a limitation that has recently been circumvented by population genomics approaches. Based on exome sequencing of 600 individuals, we estimate that ancestors of rainforest hunter-gatherers and Bantu-speaking farmers separated more than 100 000 years ago, supporting the occurrence of ancient population structure in Africa since the Late Pleistocene. Detailed demographic inference indicates that hunter-gatherers and farmers recently experienced strong population collapses and expansions, respectively, accompanied by increased gene flow. We show that the impact of the recent collapse of African hunter-gatherers on their mutation load has been modest. We provide evidence that the Holocene in central Africa was characterized by large-scale population migrations associated with the emergence of agriculture, and that early farmers first moved southward, through the equatorial rainforest, before spreading toward eastern and southern Africa. Genome scans detected several candidate loci for positive selection in these populations, as well as signatures of polygenic selection for short stature in rainforest hunter-gatherers. Finally, we provide evidence that adaptive variation has been acquired by central African populations through admixture, suggesting a previously unappreciated role of intraspecies gene flow in human local adaptation.

Inside Africa: Uncovering patterns of human genetic diversity
SMBE-OR-178

Isolation in the Fortunate Isles: genetic composition and migration patterns of the Canary Islands indigenous population

R. Fregel [1 2,*], A. C. Ordóñez [1 3], J. Santana [4], J. Gonzalez-Serrano [1], M. Arnay [3], B. Shapiro [5], C. D. Bustamante [2]
[1]Department of Genetics, Universidad de La Laguna, La Laguna, Spain, [2]Department of Genetics, Stanford University, Stanford, United States, [3]Department of Prehistory, Universidad de La Laguna, La Laguna, Spain, [4]Department of Anthropology, Durham University, Durham, United Kingdom, [5]Paleogenomics Lab, University of California Santa Cruz, Santa Cruz, United States

Abstract: The Canaries consist of seven main islands situated in the Atlantic Ocean, near the western Saharan coast of Africa. Although it is thought Plinio the Elder (23 - 79 AD) wrote about the Canary Islands in his Natural History, they were forgotten until re-discovered by European sailors in the 13th century and eventually conquered and colonized by the kingdom of Castile in the 15th century. The crushing of the resistance, and new diseases brought in by the conquerors, resulted in a high mortality rate among the native population, and the subsequent admixture with the European colonizers led to the loss of the indigenous culture and language.

Multidisciplinary research, including the analysis of ancient DNA, places the origin of the indigenous population in North Africa, mostly related to Berber communities, and radiocarbon dates on short-lived samples point to the 1st millennia for the peopling of the islands. Apart from the interest that the indigenous population of the Canary Islands raises by itself, they also represent an open window to the Prehistory of North Africa, a region that has been understudied from a paleogenomic perspective. The analysis of the Canarian indigenous population also offers a unique opportunity to understand how human populations are affected by isolation, migration and/or adaptation to a new environment. To provide detailed paleogenomic data on the Canary Islands indigenous population, we analyzed 48 ancient samples from 25 different archaeological sites from all seven islands.

Some mtDNA lineages present in the indigenous population of the Canary Islands have a clear North African adscription, pointing to a Berber origin. However, most indigenous lineages have a wider Mediterranean distribution, corroborating recent evidences of Neolithic dispersions in North Africa. Comparisons between islands indicate that the genetic composition of the indigenous population was variable, with some populations having high genetic diversity, while others were probably affected by genetic drift and/or bottlenecks. This result implies that, after the peopling of the archipelago, every island experienced its own evolutionary path, determined by the environmental conditions and limitations of insularity. We also observe an asymmetrical distribution of mtDNA haplogroups in the ancient population, with certain haplogroups appearing exclusively in the islands closer to the continent, implying the existence of more than one migration event. Although results are still preliminary, low-coverage whole-genome data starts to provide support to the conclusions reached using mtDNA, and suggests that the European Neolithic impact in North Africa was heterogeneous.

Inside Africa: Uncovering patterns of human genetic diversity
SMBE-OR-179
The genetic landscape of Ethiopia: diversity, intermixing and the association with culture
G. Hellenthal [1,*], S. Lopez [1], A. Tarekegn [2], N. Bradman [3], M. Thomas [1], E. Bekele [2]
[1]University College London, London, United Kingdom, [2]Addis Ababa University, Addis Ababa, Ethiopia, [3]Henry Stewart Talks, London, United Kingdom

Abstract: The rich linguistic, ethnic and cultural diversity of Ethiopia provides an unprecedented opportunity to understand the level to which cultural factors correlate with -- and shape -- genetic structure. Here we report results from analyses of new genome-wide autosomal DNA from >1100 Ethiopians representing 69 different ethnic groups. Using novel statistical approaches, we report the extent to which genetic similarity among Ethiopians correlates with birthplace, ethnic identity, shared language and the sharing of 31 cultural practices. We demonstrate significant associations between genetic similarity and each of geographic and elevation distance, both among present-day Ethiopians as expected and, more surprisingly, between present-day Ethiopians and a 4,500-year-old Ethiopian. We also show how ethnic groups that reported engaging in particular cultural practices are more genetically similar than expectations based on language and spatial distance. Furthermore, we give examples of how social behaviours have directly -- and strongly -- increased genetic differences among groups, though also providing evidence of recent intermixing among peoples from different language groups and religious affiliations.

In addition, we describe the ancestral history of different Ethiopian groups, revealing a southwest-northeast cline across Ethiopia defined by ancestry related to Central and West African groups versus Egyptian groups. We date distinct admixture events that show correlations with major language classifications and geography, including events dated to ~1600-2800 years ago involving sources carrying DNA similar to present-day Egyptians versus events starting ~1,450 years ago involving sources carrying DNA related to western Sub-Saharan Africans. Overall these results showcase the ability of genetics to help shed light on the ancestral histories of different ethnic groups, for example corroborating oral traditions. They also illustrate the necessity of accounting for ethnicity, geography, and sometimes even occupation, when designing sampling strategies (e.g. for genome-wide association studies, GWAS) to study Ethiopians.

Insights from fitness landscapes into evolutionary pathways

SMBE-OR-209

ONE (SMALL) STEP AT A TIME: THE DISTRIBUTION OF FITNESS EFFECTS OF YEAST MUTATIONS UNDER DIFFERENT STRESSES

C. Bank [1,*]

[1]Instituto Gulbenkian de Ciência, Oeiras, ,

Are you a member of SMBE?: No

Poster Submission: Recent technological advances make it possible to experimentally screen the fitness effects of thousands of engineered mutations simultaneously, which carries the promise to answer fundamental evolutionary questions regarding the potential for adaptation, the ubiquitousness of purifying selection, and the presence of epistasis. However, experimental estimates of the distribution of fitness effects have looked quite different from those obtained from statistical approaches using polymorphism data from natural populations. Here we systematically measured fitness effects of all 14598 single-step amino-acid changing mutations along the complete sequence of the heat-shock protein Hsp90 in yeast. Across all environments, we observed that the fitness of a large percentage (up to 80%) of mutations is indistinguishable from that of the reference genotype. Consistent with classical models of adaptation, we found the largest numbers of deleterious and beneficial mutations in the most stressful environments. By estimating the sensitivity of the distribution of fitness effects (DFE) to environmental changes at each amino-acid position, we found that the change of fitness effects is largely due to shifts in the mean DFE rather than shuffling of mutational effects. Thus, so-called "costs of adaptation" that appear when mutational effects change their sign across environments are rare. Our results support the hypothesis that structural properties of the mutated amino acid position are the dominant constraint on new mutations across all environments. We also compare our DFEs with expectations obtained from natural polymorphism data and discuss how these results support classical population genetic theories.

Insights from fitness landscapes into evolutionary pathways
SMBE-OR-210
Modeling higher order genetic interactions
J. Zhou [1], D. M. McCandlish [1,*]
[1]Simons Center for Quantiative Biology, Cold Spring Harbor Laboratory, Cold Spring Harbor, United States

Abstract: The outcome of the evolutionary processes depends, in large part, on the structure of the fitness landscape. Today, high-throughput experimental techniques are capable of illuminating the structure of very large fitness landscapes by simultaneously measuring phenotypes for tens of thousands to millions of genotypes. Nonetheless, interpreting the evolutionary implications of these experiments has proven difficult due to problems with (1) the substantial extent of missing data in these experiments, (2) experimental noise, and (3) the inherent difficulty in understanding the structure of high-dimensional landscapes. Here, I will present two principled methods for inferring complete fitness landscapes from incompletely sampled deep mutational scanning data. Importantly, both methods are capable of modeling not just simple pairwise interactions but also higher-order epistasis between any number of sites. The first method is called Minimum Epistasis Interpolation and attempts to infer the smoothest fitness landscape compatible with experimental observations by minimizing the average squared epistatic coefficient between all possible pairs of mutations in all possible genetic backgrounds. This approach results in a model that can behave in a very complicated manner where the data requires it but behaves nearly additively in regions of sequence space where data is sparse or absent. The second method is called Empirical Variance Component Interpolation. It can be viewed as an empirical Bayes procedure wherein one first estimates the fraction of variance due to each interaction order (additive, pair-wise, three-way, etc.) and then infers the maximum a posteriori fitness landscape using the inferred variance components as priors in a Gaussian process framework. I will demonstrate the performance and characteristics of the two methods by applying them to several datasets, including high-throughput assays of protein binding and pre-mRNA splicing. Moreover I will present low-dimensional visualizations of the resulting landscapes and provide a comparison of the differences in character between the dynamics of long-term molecular evolution induced by these landscapes.

Insights from fitness landscapes into evolutionary pathways
SMBE-OR-211
Is climbing up a fitness peak the same as falling down?
D. Lyons [1,*], Z. Zou [1], G. Zhang [1]
[1]University Of Michigan, Ann Arbor, United States

Abstract: Two of the most general findings in experimental evolution studies suggest that organisms do not climb up a fitness peak in the same way that they go down. In many adaptive evolution experiments, the rate of adaptation decelerates as mutations accumulate. It is commonly concluded that this reflects negative epistasis among beneficial mutations. On the other hand, organisms under extreme drift in mutation accumulation (MA) experiments generally exhibit decelerating rates of fitness decline, suggesting positive epistasis among random mutations. Thus, there is an apparent contradiction. As organisms "climb up" a fitness peak, their increase in fitness slows. We would expect that fitness decline accelerates in the reverse process as organisms "fall down" the peak. However, we observe that fitness decline also slows. Additionally, these observed "uphill" and "downhill" trajectories imply opposite types of epistasis. We explain this contradiction with simple evolutionary simulations on theoretical, computationally predicted, and experimentally measured fitness landscapes. Considering a house-of-cards model in which the fitness of a genotype is a uniform random variable between 0 and 1 reveals that decelerating adaptation can be a mathematical artifact rather than a manifestation of negative epistasis. We also find that clonal interference can produce slowing rates of both adaptation and fitness decline, assuming only variable mutational fitness effects and rare large effect mutations but no epistasis. Finally, because various adaptive evolution experiments do directly identify negative epistasis between beneficial mutations, we hypothesize a positive correlation between the effect size of a mutation and the likelihood or strength of negative epistasis. This would lead to a bias towards negatively-interacting mutations under adaptive clonal evolution, but not during MA experiments. In sum, we find that the relationship between fitness change and number of mutations does not necessarily reveal general patterns of epistasis as commonly assumed. Furthermore, we suggest that clonal interference with no epistasis and, possibly, biased adaptive processes contribute to the common finding of decelerating adaptation.

Insights from fitness landscapes into evolutionary pathways
SMBE-OR-212
Systematic analysis of high order epistasis using barcoded combinatorial gene drives
A. N. Nguyen Ba [1,*], C. W. Bakerlee [1], J. I. Rojas Echenique [1], K. Shulgina [1], M. M. Desai [1]
[1]Harvard University, Cambridge, United States

Abstract: We mapped the complete fitness landscape comprising ten beneficial single nucleotide polymorphisms spread over the whole genome using barcoded combinatorial assembly of CRISPR gene drives in yeast, generating 1024 genetic combinations in four mating steps in an exponentially growing plate array. This method makes use of recombining sgRNA to effectively link distant gene drives to a single benign locus. Barcode sequencing was then used to infer the fitnesses of all created genotypes in a pooled assay for homozygous diploids, heterozygous diploids and haploids in different environments. We find that, although the fitness landscape changes shape under these different contexts, fitness-mediated diminishing returns is the dominant predictor of mutational effects, with mutations changing sign over large fitness ranges. Consequently, pairwise and higher-order epistasis are also under fitness-mediated diminishing returns, giving rise to a form of phenotypic buffering. Single-effect mutations and fitness-mediated epistasis accounts for over 85% of the phenotypic variance between genotypes.

Insights from fitness landscapes into evolutionary pathways
SMBE-OR-213
Impacts of mutation and selection on the regulation of gene expression
P. Wittkopp[*], P. Vande Zande[1], M. S. Hill[2], F. Duveau[3]
[1]Molecular, Cellular, and Developmental Biology, [2]Ecology and Evolutionary Biology, University of Michigan, Ann Arbor, MI, United States, [3]Laboratoire Matière et Systèmes Complexes, Universitè Paris Diderot and CNRS, Paris, France

Abstract: Genetic variation affecting gene expression is wide-spread within and among species. This variation reflects the combined actions of mutation introducing new genetic variants and selection eliminating deleterious ones. Comparative studies of gene expression in fruit flies, yeast, plants, and mice have shown that the relative contributions of *cis*- and *trans*-acting variants to expression differences change over evolutionary time, indicating that selection has different effects on *cis*- and *trans*-regulatory variants. To better understand the reasons for this now widely observed pattern, we have been systematically studying the effects of mutation and selection on expression of the *TDH3* gene of the baker's yeast *Saccharomyces cerevisiae*. This work has revealed differences between *cis*- and *trans*-regulatory mutations in their frequency, effects, and dominance. Differences in pleiotropy are also generally assumed to exist between *cis*- and *trans*-regulatory that affect their evolutionary fate, but have been difficult to measure. In this talk, I will discuss how newly arising *cis*- and *trans*-regulatory mutations affecting expression of this focal gene are structured within the regulatory network, their pleiotropic effects on expression of all other genes in the genome, and how these pleiotropic effects influence fitness. A computational model of regulatory evolution integrating empirically observed differences in properties of *cis*- and *trans*-regulatory mutations will also be presented and discussed.

Insights from fitness landscapes into evolutionary pathways
SMBE-OR-214
Survival of the likeliest leads to sub-optimal phenotypes in adapting populations of Pseudomonas fluorescens
M. Barnett*, P. Rainey

Abstract: Why a particular phenotype arises during adaptation is not only a product of its fitness but also its likelihood of being realized by mutation. This difference in likelihood can be due to mutational biases and properties of the genotype-phenotype map, which translates genetic variation into its phenotypic consequence. Disentangling the relative contribution of fitness and likelihood to the outcome of adaptation requires the discovery of alternative paths evolution didn't follow and determining why. One way this can be achieved is through the exploration of 'latent' alternative phenotypes and their respective fitness. However, since such phenotypes are initially invisible to natural selection their discovery is in most cases not feasible. Here I employ experimental evolution of the model organism *Pseudomonas fluorescens* SBW25 as a means to explore possible latent phenotypes.
SBW25 is an obligate aerobe that when grown in static liquid media rapidly evolves the ability to colonize the oxygen replete air-liquid interface. Engineering strains that precluded pathways to the initially realized phenotypes and challenging the bacteria to again adapt to the same environment revealed numerous alternative phenotypes that could then be recreated in the ancestral genome. Some of these discovered mutants proved of higher fitness than initially realized adaptive types. Their prior absence is therefore explained by a lower likelihood of being manifest by mutation, or in the case of a phenotype requiring two or more mutations - a product of both the likelihood and fitness of any single-mutational step. It is demonstrated that the initially realized but less-fit phenotypes are more likely due to activation via mutational pathways requiring only single loss-of-function mutations e.g. to a gene encoding a negative-regulator. Such pathways provide a large mutational 'target' and so the associated phenotype is more readily manifested by mutation and amplified by selection, concomitantly reducing the population from which alternative phenotypes can arise.
These results demonstrate that evolution, even under strong selection and in large population sizes, may proceed along a mutational path of least resistance toward sub-optimal phenotypes.

Insights from fitness landscapes into evolutionary pathways

SMBE-OR-215

Secondary environmental variation creates a shifting evolutionary watershed for the methyl-parathion hydrolase enzyme

D. Anderson [1,*], G. Yang [2], F. Baier [3], N. Tokuriki [4]

[1]Biochemistry and Molecular Biology, University of Calgary, Calgary, [2]Michael Smith Laboratories, Univesity of British Columbia, Vancouver, Canada, [3]Microbiologie Fondamentale, Universite de Lausanne, Lausanne, France, [4]Michael Smith Laboratories, University of British Columbia, Vancouver, Canada

Abstract: Enzymes can evolve new catalytic activity when their environments change to present them with a novel substrate. Despite this seemingly straightforward relationship between environment and adaptation, it is not always obvious how other secondary environmental variation (i.e. factors besides the direct catalytic target) impacts enzyme evolution. Here, we characterize the adaptive landscape separating two metal-dependent enzymes (an ancestral dihydrocouramine hydrolase and a derived methyl parathion hydrolase) under eight different secondary environments defined by alternative divalent metals supplemented in growth media at physiological concentration. We find that different metal environments produce distinct adaptive landscapes across the same network of genotypes, however they vary in terms of both the number of "fitness peaks" (either local or global optimal genotypes) as well as the genotype of the "best" peak across the set of sequence space being examined. These differences are the result of significant genotype-by-environment (*G×E*) interactions, as well as environment-dependent epistasis (*G×G×E*). We find that variation in the metal environment shifts an evolutionary watershed, causing the outcome of adaptation to depend on the secondary environment in which it occurs. This suggests that adaptive landscapes may be fluid and that molecular adaptation is highly contingent not only on obvious factors (such as catalytic targets) but also on less obvious secondary environmental factors that can direct an evolutionary trajectory toward distinct outcomes.

SMBE Editors symposium
SMBE-OR-077
A domain-level analysis of low complexity regions
F. U. Battistuzzi [1,*], A. Mahama [1]
[1]Biological Sciences, Oakland University, Rochester, United States

Abstract: Repetitive regions belong to a wider category of sequences known for their lower information content (low-complexity regions) compared to their flanking regions. While past studies have focused on the role of these regions in large eukaryotic genomes, their presence, composition, and function within prokaryotes is less known. In eukaryotes, low complexity regions are considered drivers of genome evolution producing the genetic variability necessary for the evolution of new genes, gene functions, and gene families. It is likely that a similar role is played by these regions in prokaryotes but, for now, even the presence of these regions within these species is poorly understood. In this study, we performed a domain-level analysis of low complexity regions in prokaryotes to determine their presence, composition, and general evolutionary patterns compared to eukaryotes. We found that although the frequency of these regions in prokaryotic genomes is lower than in eukaryotes, they are widespread and present in virtually all species. Some general patterns are similar to those found in eukaryotes (e.g., species-specificity of their composition) but other patterns are unique and seem to be correlated with lifestyle. For example, some extremophilic Archaea have unique patterns of low complexity regions that may be important for the unique challenges faced by these species. Overall, analyses such as the one presented here will provide a comprehensive overview of low complexity regions within genomes and their potential evolutionary and functional roles.

SMBE Editors symposium
SMBE-OR-078
Detecting incomplete selective sweeps during modern human evolution
Y. Satta [1,*], W. Zheng [1], K. Nishiyama [1], N. T. Fujito [2], T. Hayakawa [3], N. Takahata [1]
[1]SOKENDAI (The Graduate University for Advanced Studies), Hayama, Japan, [2]University of California, San Francisco, United States, [3]Kyushu University, Fukuoka, Japan

Abstract: After the out-of-Africa dispersal of modern humans, physical and cultural environments have ever been changing drastically. Such changing environments must have been a cause as well as a driving force of adaptive evolution. As time elapsed is at most 50 K years (2000 generations), it is conceivable that many adaptive variants have not had enough time to be fixed. Although methods such as $F_{ST,}$ iHS and nSL are available to detect incomplete selective sweeps, robustness to recombination and distinguishing between hard and soft sweeps remain to be solved. To this end, we developed a method based on the two-dimensional site frequency spectrum. This method can also provide information necessary to estimate the TMRCA within an adaptive allelic group. Here we report three applications: (1) to demonstrate the principle and power of the method, we examined six loci that are thought to be under positive selection; (2) to render a genomic landscape, we tested the null hypothesis of neutrality for >150 intron SNPs randomly selected from the 1000 genome database; (3) to study the origin of mental disorders during human evolution, we tested 108 SNPs reported by theSchizophrenia Working Group of the Psychiatric Genomics Consortium

SMBE Editors symposium
SMBE-OR-079
Evolution of a young sex chromosome in a fish
D. Charlesworth [1,*]
[1]Institute of Evolutionary Biology, University of Edinburgh, Edinburgh, United Kingdom

Abstract: I will describe genetic and population genomic analyses of the evolution of the sex-linked region in the guppy, Poecilia reticulata.

SMBE Editors symposium

SMBE-OR-080

Human-specific functions of the transcription factor ZEB2 might be involved in brain and cognitive evolution

J.-E. Lee [1], V. Jovanovic [1], S. Berto [2], K. Nowick [1,*]

[1]Institute of Zoology, Freie Universität Berlin, Berlin, Germany, [2]University of Texas Southwestern Medical Center, Dallas, United States

Abstract: Humans differ from other primate species in many phenotypes, among them in their larger brain, cognition and behavior. However, the molecular basis underlying these differences is not well known yet. Analyzing transcription factor networks of the prefrontal cortex, we identified ZEB2 as one of the transcription factors with most differences in connectivity between humans and chimpanzees. Interestingly, ZEB2 is important during brain development and cell type specification and has been associated with microcephaly and reduced cognitive abilities, making it an interesting candidate for being involved in setting the stage for human cognition. In order to determine ZEB2 target genes, we performed Chromatin immune-precipitation followed by Next-Generation-Sequencing (ChIP-Seq) and knock-down experiments followed by RNA-Seq. To identify human-specific ZEB2 target genes, we conducted the same experiment in human, chimpanzee, and orang-utan cell lines with three biological replicates, each. Many ZEB2 target genes are gene regulatory factors, such as histone modifying enzymes and transcription factors, regulating various cell signaling pathway. Overall, the functions of human and chimpanzee ZEB2 are more similar to each other than compared to orang-utan ZEB2, which is in agreement with the phylogenetic distances. Human specific ZEB2 targets include genes involved in neuron projection, synaptic transmission, and autism. These functional changes on the human lineage might have contributed to the evolution of the human brain, its neural circuits, and its specific cognitive abilities.

SMBE Editors symposium
SMBE-OR-081
Gametic specialization of centromeric histone paralogs in Drosophila virilis
L. E. Kursel [1], H. McConnell [1], M. Zych [1], H. S. Malik [*]
[1]Basic Sciences, Fred Hutchinson Cancer Research Center, Seattle, United States

Abstract: In most eukaryotes, centromeric histone (CenH3) proteins mediate the highly conserved process of chromosome segregation as the foundational kinetochore assembly factor. However, in multicellular organisms, CenH3 proteins have to perform their essential functions in different chromatin environments. CenH3 proteins not only mediate mitosis and meiosis but also ensure epigenetic inheritance of centromere identity on sperm chromatin, which is highly compact and almost completely stripped of histones during spermiogenesis. We hypothesized that such disparate chromatin environments might impose different functional constraints on CenH3. If so, gene duplications could ameliorate the difficulty of encoding divergent and even potentially incompatible centromeric functions in the same gene. Here, we analyzed the cytological localization of two recently identified CenH3 paralogs, Cid1 and Cid5, in D. virilis using specific antibodies and epitope-tagged transgenic strains. We find that only ancestral Cid1 is present in somatic cells, whereas both Cid1 and Cid5 are expressed in testes and ovaries. However, Cid1 and Cid5 are alternately retained in male and female gametes; Cid1 is lost in male meiosis but retained throughout oogenesis, whereas Cid5 is lost during female meiosis but retained in mature sperm. Following fertilization, maternally deposited Cid1 rapidly replaces paternal Cid5 during the protamine-to-histone transition. Our studies reveal mutually exclusive gametic specialization of two divergent CenH3 paralogs. We suggest that centromeric histone duplication and divergence may allow essential genes involved in chromosome segregation to specialize and thereby resolve an intralocus conflict between maternal and paternal centromeric histone requirements in many animal species.

SMBE Editors symposium

SMBE-OR-082

Rise of simple substitution models for phylogenetics inference?

S. Kumar [1,*], Q. Tao [1], J. Barba-Montoya [1]

[1]Institute for Genomics and Evolutionary Medicine, Temple University, Philadelphia, United States

Abstract: In the early days of molecular evolutionary studies, many simple models were devised for phylogenetic inference. This simplicity has given way to an impressive diversity of complex models of nucleotide and amino acid substitution. These complex models are biologically more realistic and yield more accurate estimates of absolute sequence divergence. Such attributes have been the primary motivation for the use of increasingly sophisticated models for estimating evolutionary trees, divergence times, ancestral states, and relative rates of evolution among sites. Interestingly, some new findings are challenging the conventional wisdom that complex models are required for reliable inference. Emerging literature and new results will be presented that suggest that simple and complex models produce similar results in many types of phylogenetic analyses. Applying inductive reasoning, I find these trends to be encouraging because even the most complex models available are blatant approximations of the real evolutionary patterns.

SMBE Editors symposium

SMBE-OR-083

The Genomics of Selfing in Maize: Catching Purging in the Act

B. Gaut [1,*], A. Muyle [1], K. Roessler [1], C. Munoz-Diez [2], G. Gaut [3], A. Bousios [4], M. Stitzer [5], D. Seymour [1], J. F. Doebley [6], Q. Liu [7]

[1]Ecology and Evolutionary Biology, University of California, Irvine, Irvine, United States, [2]Dept. of Agronomy, University of Cordoba, Cordoba, Spain, [3]Dept. of Cognitive Sciences, University of California, Irvine, Irvine, United States, [4]School of Life Sciences, University of Sussex, Brighton, United Kingdom, [5]Dept. of Plant Sciences, University of California, Davis, Davis, [6]Dept. of Genetics, University of Wisconsin, Madison, United States, [7]College of Agriculture and Food Science, Zhejiang A&E University, Huangzhou, China

Abstract: In plants, self-fertilization is both an important reproductive strategy and a valuable genetic tool. In theory, selfing increases homozygosity at a rate of 0.50 per generation. Increased homozygosity can uncover recessive lethal variants and lead to inbreeding depression, unless it is countered by genetic purging. Here we investigated the dynamics of purging on genomic scale by testing three predictions. The first was that heterozygous, putatively deleterious SNPs in the parent are preferentially lost from the genome over time. The second was that the loss of deleterious SNPs varied as a function of recombination rate, because recombination increases the efficacy of selection by uncoupling linked variants. Finally, we predicted that genome size (GS) decreases during selfing, due to the purging of deleterious Transposable Element (TE) insertions. We tested these three predictions by following GS and SNP variants in a series of selfed maize (*Zea mays* ssp. *mays*) lines over six generations. In these lines, heterozygosity declined more slowly than expected; instead of decreasing by 50% each generation, it declined by ~27%. Heterozygosity was lost more rapidly, however, at sites with putatively deleterious SNP variants, particularly in chromosomal regions of elevated recombination. Finally, three lines lost 398 Mb from their genomes, on average, over the short timeframe of our experiment,. TEs were the principal component of loss, and genome loss was more likely for lineages that began with more TE and more chromosomal knob repeats. Overall, our work documented remarkable genomic loss - roughly three Arabidopsis genomes, on average - in only a few generations of selfing.

SMBE Editors symposium

SMBE-OR-084

Yaponesia Genome Project – genome sequence analyses of modern humans, ancient humans, animals, and plants, combined with archeological and linguistic data analyses

N. Saitou [1,*]

[1]Population Genetics, National Institute of Genetics, Mishima, Japan

Abstract: People reached Yaponesia (Japanese Archipelago) around 40,000 years ago, and many waves of migration has occurred since. Within this framework, we seek to decipher the genomic history of Yaponesians (people living on Japanese Archipelago) through determination and comparative analyses of many modern and ancient human genomes. We also analyze genome data of animals and plants introduced to Yaponesia by human migration. Temporal changes of population size are also estimated from genome sequence data using existing and newly developed methods. To have a better understanding of the history of Yaponesians, this interdisciplinary study will involve evolutionary genomics, archeology with special reference to age estimation (dating) of artifacts and ancient organisms, and linguistics with special reference to dialect analyses of Japanese and Ryukyuan languages. We aim to establish a new discipline, "historical genomics" of Yaponesians, through integration of these various disciplines. This multidisciplinary study is funded with 5-year MEXT/Japan grant nicknamed "Yaponesian Genome" started from FY2018. I am head of this project with six research groups, as well as head of modern human genome research group, who will collect DNA samples of 50 modern Yaponesian populations and analyze their genome-wide SNP data and whole genome sequences with many non-Yaponesian and ancient human genome data. Outline of this project is first introduced, followed by my own proposal of three-wave migration model to Yaponesia. This model, dubbed the "Inner Dual-Structure" will explain genetic diversity observed within mainland Yaponesia, and complements the established "Dual-Structure" model for the whole of Yaponesia. Implications to modern human evolution in East Eurasia is discussed.

SMBE Editors symposium

SMBE-OR-104

Cell-Type Specific Regulatory Evolution of Human Brains

I. Mendizabal [1], S. Berto [2], N. Usui [3], K. Toriumi [2], P. Chatterjee [4], C. Douglas [2], I. Huh [2], H. Jeong [4], T. Layman [5], C. Tamminga [6], T. M. Preuss [7], G. Konopka [3], S. Yi [2,*]

[1]School of Biology, Georgia Institute of Technology, Atlanta, [2]UT Southwestern Medical Center, Dalles, [3]UT Southwestern Medical Center, Dallas, [4]Georgia Institute of Technology, Atlanta, [5]Gerogia Institute of Technology, Atalnta, [6]UT Southwestern Medical Cer, Dallas, [7]Emory University School of Medicine, Atlanta, United States

Abstract: Recent advances in genetics and epigenetics highlight that molecular mechanisms of gene regulation are often cell-type specific. Here, we ask how regulatory mechanisms evolve in a cell-type specific way. Specifically, we investigate evolutionary changes of DNA methylation and gene expression in neurons and oligodendrocytes of human and non-human primate brains. Comparing DNA methylation maps of neurons and oligodendrocytes from multiple individuals of humans, chimpanzees and macaques, we can follow the evolutionary trajectory of DNA methylation and gene expression in a cell-type specific manner. Our results indicate highly distinct epigenetic regulatory mechanisms in the two cell types and evolutionary trajectories. We identified numerous divergently methylated regions between cell-types, in particular genomic regions previously identified as functionally relevant to early brain development of humans. Furthermore, we find that differentially methylated regions between cell types overlap with disease risk loci, demonstrating that altered gene regulation of distinct cell types might relate to the evolution of neuropsychiatric disorders.

SMBE Editors symposium

SMBE-OR-105

The role of natural transposable element insertions in stress response.

J. González [1 2,*]

[1]Institute of Evolutionary Biology (CSIC-UPF), Barcelona, Spain, [2]Comparative and Computational Genomics, Institute of Evolutionary Biology (CSIC-UPF), Barcelona, Spain

Abstract: Transposable elements are ubiquitous, abundant, and active components of genomes. Although most of the mutations caused by transposable elements are likely deleterious or neutral, adaptive mutations caused by transposable elements have also been identified. Our lab focuses on elucidating the role of transposable elements in adaptation at a genome-wide scale. We have used a computational pipeline to estimate population frequencies of reference transposable elements in 303 individual genome sequences and 83 pool-sequenced samples, collected from 60 worldwide natural *Drosophila melanogaster* populations. Taking into account the age and length of the insertions, and the evidence of selection in their flanking regions, we identified a subset of 300 polymorphic TEs likely to play a role in adaptation. Interestingly, a proportion of these insertions are located nearby stress-related genes. We thus investigated the role of transposable element insertions in adding transcription factor binding sites related to six stress responses. We are currently generating new *Drosophila melanogaster* reference genomes from several European populations to extend our analysis to non-reference transposable elements.

SMBE Editors symposium

SMBE-OR-106

The Major Histocompatibility Complex, a magic trait driving sympatric speciation?

M. Barluenga[*], M. J. Hofmann [1]

[1]Biodiversity and Evolutionary Biology, Museo Nacional Ciencias Naturales, CSIC, Madrid, Spain

Abstract: A thrilling topic in evolutionary biology is understanding the mechanisms generating biodiversity. Adaptive radiations are particularly good models for the study of speciation, since phenotypic divergence leading to speciation happens rapidly and repeatedly. The Neotropical Midas cichlid adaptive radiations in Nicaragua are driven by natural selection. Adaptation to alternative habitats and diets linked to morphological shifts has shaped species distribution, and has contributed to the development of reproductive barriers. However, whether this mechanism alone is sufficient to cause and maintain divergence in this system remains to be elucidated. Following on the strong evidence that ecological preference (habitat choice) is the main force shaping populations and driving reproductive isolation, we introduce a new environmental factor that causes differentiation in this system: parasite mediated selection. We evaluate the role of host-parasite interactions as a potential driving force for divergence and speciation mediated by the evolution the Major Histocompatibility Complex (MHC) genes. These genes are responsible of generating diversity (when responding and adapting to different parasites/mutualists), and at the same time of promoting assortative mating and barriers to reproduction in the host (mate preference based on olfaction cues) acting as magic traits.

SMBE Editors symposium

SMBE-OR-107

Insights on the mitochondrial genome of appendicularians

D. Huchon[*], M. Novosolov[1], G. Rade[2], T. Stach[3], C. Gissi[4]

[1]School of Zoology, Tel-Aviv University, Tel-Aviv, Israel, [2]Institute for Marine and Coastal Research, University of Dubrovnik, Dubrovnik, Croatia, [3]Institut für Biologie, Molekulare Parasitologie, Humboldt-Universität zu Berlin, Berlin, Germany, [4]Department of Biosciences, Biotechnology and Biopharmaceutics, University of Bari “Aldo Moro", Bari, Italy

Abstract: *Oikopleura dioica* (Appendicularia) is a model organism to understand the evolution of chordates. Although its nuclear genome has been sequenced, its mitochondrial (mt) genome has not been determined, probably because of their fast evolutionary rate and extensive RNA editing. We sequenced cDNA of the cox1 gene and nuclear genomes (using Illumina) from *O. dioica* and *Fritillaria pellucida*. Mapping the resulting DNA reads onto the cox1 cDNA sequences, we confirmed the previously-published observation that, at DNA level, long poly-T stretches interrupt the mt ORFs, hampering their identification. In addition, we noticed that mitogenomic poly-T correspond at RNA level to TTTTTT (6T) regions in *O. dioica* but to shorter TTTT (4T) regions in *F. pellucida*, leading to hypothesize species-specific RNA editings by deletion. Remarkably, different editing sites were observed between Mediterranean and Japanese populations of *O. dioica* suggesting a rapid evolution of these sites.

While tRNAs usually punctuate the protein-coding genes in chordate mt genome, only one tRNA, tRNAmet, was identified among *O. dioica* mt contigs assembled from published cDNA reads. To test if mt tRNAs could have been lost, we searched the genomic assembly for the presence of mt aminoacyl tRNA synthetase (aaRS) genes, which encode the proteins responsible for the aminoacylation of its cognate tRNA. No specific mt aaRS genes were found in the nuclear genome assembly of *O. dioica*, except for mt-MetRS, which catalyzes the mt-tRNAMet aminoacylation. This suggests that the mt genome of *O. dioica* encodes only the single tRNAMet.

SMBE Editors symposium

SMBE-OR-108

Evolution of the rate of copy-number and structural variant mutations under relaxed selection in Caenorhabditis elegans

A. S. Saxena [1], M. P. Salomon [1], C. Matsuba [1], M. Rajaei [1], C. F. Baer [1,*]

[1]Biology, University of Florida, Gainesville, United States

Abstract: Mutation Accumulation (MA) experiments have been a workhorse of evolutionary genetics for over fifty years, and much of what is known about the mutational process at both the phenotypic and molecular level has been learned from MA experiments. The advent of economical whole-genome short-read sequencing greatly increased our ability to characterize the rate and spectrum of base-substitution and short indel mutations. Copy-number variants (CNVs) and structural variants (SVs, e.g., inversions, translocations) have been more difficult to characterize. The paucity of information on the mutational properties of CNVs and SVs is significant because there is compelling evidence that those types of mutations underlie much variation in complex traits.

We previously demonstrated by means of a "second-order MA" experiment (i.e., sets of MA lines founded from individual MA lines) with the nematode *Caenorhabditis elegans* that the rate of base-substitution and (especially) small indel mutations increased over the course of ~250 generations of minimal selection. Preliminary analyses of CNVs based on short-read Illumina sequence data suggested that a similar trend holds for CNVs, but the quantitative estimates of CNV mutation varied by over an order of magnitude depending on the input parameters of the analytical algorithm. To attempt to get a better handle on the CNV and SV mutation rates, we sequenced a small number of MA lines (N=5) with Pacific Biosciences long-read sequencing and used the resulting estimates to attempt to inform the analysis of our more copious short-read data (N>100) MA lines. Preliminary results suggest a conservative long CNV rate of approximately 5% that of small mutations (base subs + short indels). The data are biased toward deletions, but comparisons of the sequence of the ancestor of our MA lines with the *C. elegans* reference genome indicates that our analytical method has the capacity to detect insertions and inversions as well as deletions.

We additionally report comparative MA data from an additional set of 25 C. elegans MA lines derived from a different starting genotype (PB306).

SMBE Editors symposium

SMBE-OR-109

Population genetic models of GERP scores suggest turnover of constrained sites along the human lineage

C. Huber [1], B. Kim [2], K. Lohmueller [3,*]

[1]Australian Centre for Ancient DNA, University of Adelaide, Adelaide, Australia, [2]Department of Biology, Stanford University, Stanford, [3]Department of Ecology & Evolutionary Biology, UCLA, Los Angeles, United States

Abstract: Comparative genomic approaches of searching for sequence conservation across multiple species have been used to identify sites where mutations could be under purifying selection and have functional consequence. However, the performance of these approaches has not been rigorously evaluated under population genetic models. Here, we use simulations to study how one measure of conservation, the GERP score, relates to the strength of selection (*Ns*). We show that the GERP score distribution is affected by the strength of purifying selection. However, turnover of functional sequence and missing data can strongly affect the GERP score distribution and lead to unexpected relationships between GERP scores and *Ns*. Our work points out several important limitations to using comparative genomic approaches for determining the fitness effects of individual mutations. Further, we show that for elements that have a high turnover rate, the optimal tree size to detect sequence conservation is not necessarily the largest possible tree, and more turnover reduces the optimal tree size. Finally, we use the distribution of GERP scores across the human genome to compare models with and without turnover of sites at which mutations would be under purifying selection. We estimate that 4.97% of non-coding human genome is under purifying selection and that most of this sequence has likely experienced functional turnover throughout vertebrate evolution.

SMBE Editors symposium

SMBE-OR-110

Ancient signature of B cell regulation in lampreys: evolutionary implications

S. Das [1,*], Y. Sutoh [1], M. P. Cancro [2], J. P. Rast [1], Q. Han [1], G. Bommakanti [1], M. D. Cooper [3], M. Hirano [1]

[1]Pathology, Emory University, Atlanta, [2]Pathology, University of Pennsylvania, Philadelphia, [3]Pathology, University of Pennsylvania, Atlanta, United States

Abstract: The TNFSF13 family members, BAFF (B-cell activating factor) and APRIL (a proliferation-inducing ligand) are important regulatory factors for lymphocyte activation and survival in mammals. A BAFF/APRIL-like relative called BALM has also been identified in cartilaginous and bony fishes. The TNF superfamily ligands BAFF and APRIL interact with three receptors, BAFFR, BCMA and TACI, to play discrete and crucial roles in regulating B cell selection and homeostasis in mammals. The interactions between these ligands and receptors are both specific and redundant: BAFFR (BAFF receptor) binds BAFF, while BCMA (B cell maturation antigen) and TACI (transmembrane activator and CAML interactor) bind to either BAFF or APRIL. Lampreys and hagfish are the only extant jawless vertebrates, both of which have B-like and T-like lymphocytes. To gain insight into lymphocyte regulation in jawless vertebrates, here we identified a *TNFSF13*-like and two *BCMA*-like genes in lampreys, *BCMAL1* and *BCMAL2*. Phylogenetic analysis and molecular cladistics markers suggest that the TNFSF13 protein in lampreys is more BAFF-like than APRIL-like. Similar to mammalian system, the lamprey *TNFSF13* homolog is expressed in T-like, B-like and innate immune cells, while both *BCMAL1* and *BCMAL2* are preferentially expressed by B-like lymphocytes. *In vitro* analyses indicated that the lamprey TNFSF13 protein can bind to a BCMA-like receptor Ig fusion protein and to both BCMAL1- and BCMAL2-transfected cells. Discriminating regulatory roles for the two BCMA-like molecules are suggested by their differential expression before and after activation of the B-like lymphocytes in lampreys. Our composite results imply that BAFF-based mechanisms for B cell regulation evolved before the divergence of jawed and jawless vertebrates.

SMBE Editors symposium
SMBE-OR-111
Chimeric origin of a proteinaceous pheromone gene family in murids and cricetids

Y. Niimura [1,*], M. Tsunoda [1], K. Touhara [1]
[1]Department of Applied Biological Chemistry, Graduate School of Agricultural and Life Sciences, The University of Tokyo, Tokyo, Japan

Abstract: Pheromones are chemical signals that have evolved for the communication between individuals of the same species. Not only small volatile molecules but also soluble peptides and proteins are used as pheromones. Exocrine-gland secreting peptide 1 (ESP1) was identified as a 7-kDa peptide pheromone secreted from the extraorbital lacrimal gland into tear fluid of male mice. ESP1 stimulates the vomeronasal sensory organ in female mice and enhances female sexual receptive behavior upon male mounting. ESP1 is a member of a multigene family consisting of >30 genes in mice and ~10 in rats. The coding sequences of ESP genes are separated into two exons, one encodes a signal peptide and the other encodes a mature sequence.

To elucidate the origin and the evolution of the ESP gene family, we extensively identified ESP genes from the genome sequences of >100 mammalian species and performed molecular evolutionary analyses. The results showed that ESP genes were found only from the genomes of the family Muridae including mice and rats and the family Cricetidae including hamsters, suggesting that the ESP gene family has been originated in the common ancestor of murids and cricetids. The lengths of the ESP mature sequences are highly variable, ranging from 42 to 131 amino acids. Murids tend to have a larger number of ESP genes than cricetids, while the lengths of murid ESP genes are significantly shorter than those of cricetid genes. Moreover, longer mature sequences tend to be more conservative in amino acid sequence than shorter mature sequences. Some ESPs in rats and hamsters are expressed in the lacrimal gland and the salivary gland; however, they did not induce a clear vomeronasal activity. It therefore appears that the pheromonal function of ESPs emerged in the mouse lineage after the divergence from rats.

We also found that the ESP mature sequence showed an overall weak similarity in amino acid sequence to the α-globin gene. Because ESP mature sequences are intronless, it is likely that the ESP mature sequence was inserted into the genome via retrotransposition. Moreover, the signal sequence of ESP genes encoded in the first exon was found to be similar to the first exon of the CRISP2 gene that is adjacent to the ESP gene cluster. These observations suggest that ESP genes were generated by recombining a retrotransposed α-globin gene with the first exon of the CRISP2 gene. Our study provides an intriguing example of molecular tinkering in rapidly evolving species-specific proteinaceous pheromone genes.

This work was supported by ERATO Touhara Chemosensory Signal Project, JST.

MEGE: Microbial Eukaryotic Genomic Evolution

SMBE-OR-118

THE RED ALGAE AS MODELS TO STUDY GENOME REDUCTION, HORIZONTAL GENE TRANSFER, AND BIOMINERALIZATION

D. Bhattacharya[*], J. Lee [1], H. S. Yoon [2]

[1]Biochemistry and Microbiology, Rutgers University, New Brunswick, United States, [2]Department of Biological Sciences, Sungkyunkwan University, Suwon, Korea, Republic Of

Abstract: Genome evolution is usually viewed through the lens of growth in size and complexity over time, exemplified by plants and animals. In contrast, genome reduction is associated with a narrowing of ecological potential, such as in parasites and endosymbionts. But, can nuclear genome reduction also occur in, and potentially underpin a major radiation of free-living eukaryotes? An intriguing example of this phenomenon is provided by the red algae (Rhodophyta) that have lost many conserved pathways such as for flagellar motility, macroautophagy regulation, and phytochrome based light sensing. This anciently diverged, species-rich, and ecologically important algal lineage has undergone at least two rounds of large-scale genome reduction during its >1 billion-year evolutionary history. I will discuss our recent work on the impacts of genome reduction on red algae, how these species have become models for studying horizontal gene transfer, and our initial insights into the genomic basis of biomineralization in coralline red algae, that are key components of healthy coral reefs. The presentation will stress work on the extremophilic red algae, the Cyanidiophytina that includes the well-studied *Galdieria* lineage. We show that the high conservation of genes comprising the spliceosomal machinery (SM) in *G. sulphuraria* is correlated with enrichment of small-sized spliceosomal introns and transcript variants derived from alternative splicing. Because many of the splice variants encode truncated proteins, we postulate they undergo nonsense-mediated mRNA decay. These results suggest the impact of genome reduction in free-living eukaryotes is ameliorated by retention of the SM that lays the foundation for intron-rich genes and access to alternative RNA-splicing to amplify genetic diversity.This, and other research I will discuss reveal the red algae to be an exciting, yet under-studied model that offers numerous novel insights as well as poses many unanswered questions that remain to be explored using genomic, genetic, and biochemical methods. The fact that a speciose lineage of free-living eukaryotes has spread throughout many aquatic habitats after having lost about 13% of its primordial gene inventory challenges us to elucidate the mechanisms underlying this remarkable feat.

MEGE: Microbial Eukaryotic Genomic Evolution

SMBE-OR-119

Comparative genomics of three vertically transmitted microsporidia, including Nosema granulosis, a feminizing parasite of amphipod crustaceans

A. Cormier [1,*], M. A. Chebbi [1], I. Giraud [1], R. Wattier [2], M. Teixeira [2], C. Gilbert [3], T. Rigaud [2], R. Cordaux [1]

[1]Laboratoire Ecologie et Biologie des Interactions, Equipe Ecologie Evolution Symbiose, Université de Poitiers, UMR CNRS 7267, CNRS, Poitiers, [2]Laboratoire Biogéosciences, Université Bourgogne Franche-Comté, UMR CNRS 6282, CNRS, Dijon, [3]Laboratoire Evolution, Génomes, Comportement, Ecologie, CNRS Université Paris-Sud, UMR 9191, IRD UMR 247, CNRS, Gif sur Yvette, France

Abstract: Multicellular organisms have been continuously involved in complex interactions with microorganisms during their evolution, the most intimate of which is endosymbiosis. Over the past years, evidence has been accumulating that endosymbionts affect animal biology in many ways, such as host nutrition, development, immunity and sex determination. In the latter example, endosymbionts are able to disrupt the sex determination of hosts in favor of females because they are predominantly transmitted vertically through female egg cytoplasm.

Current efforts in our laboratory are aimed at deciphering the genetic mechanisms underlying the ability of feminizing obligate intracellular endosymbionts (Microsporidia, Fungi) to reverse genetic males into functional phenotypic females in a freshwater amphipod (Arthropoda, Crustacea). Several Microsporidia species have been shown to be vertically transmitted, some of which are also able to induce feminization of male hosts. Here, we sequenced and analyzed the genomes of three vertically transmitted Microsporidia infecting the amphipod *Gammarus roeselii*, for which feminization induction has been demonstrated (*Nosema granulosis*) or is suspected (*Dictyocoela muelleri* and *Dictyocoela roeselum*). Unlike other Microsporidia, these three species do not have any extracellular stage and cannot be isolated from host cells, thereby substantially complicating genome assembling.

Using a combination of subtractive mapping, taxonomic assignment, metagenomic binning and oligonucleotide composition similarity, we obtained assemblies of the three Microsporidia. In particular, the feminizing *N. granulosis* assembly has a total length of 8.8 Mbp with N50 of ~13 kbp and it shows a very high level of completeness (BUSCO=98%). *D. muelleri* has a very large genome (41.9 Mbp, N50 of ~7 kbp) and also exhibits a high completeness level (BUSCO=87%). Finally, our *D. roeselum* assembly is partial (2.2 Mbp, N50 of ~3 kbp and BUSCO=50%), probably due to its very low titer in host oocytes causing insufficient sequencing depth.

We compared the *N. granulosis* genome to 32 available Microsporidia genomes, including five *Nosema* genomes from horizontally transmitted, non-feminizing species: *N. antheraeae*, *N. apis*, *N. bombycis*, *N. ceranae* and *N. sp. YNPr*. The phylogenetic tree of single copy orthologous genes indicated that *Nosema* is a monophyletic genus comprising two clades, one with *N. granulosis, N. antheraeae* and *N. bombycis* and the other with *N. apis*, *N. ceranae* and *N. sp. YNPr*. Analysis of distribution of clusters of orthologous genes among Microsporidia showed that most clusters with *N. granulosis* representatives were shared with a least one other Nosematidae (1,726/1,808), as expected. Interestingly, 39 clusters were uniquely shared by vertically transmitted Microsporidia (i.e. *N. granulosis* and the two *Dictyocoela*). The functional annotation of the genes within these clusters suggested in most cases functions related to transposition.

Our results provide a first insight into the evolutionary genomics of vertically transmitted and feminizing Microsporidia endosymbionts and they offer new genomic resources for further analysis of the genetic mechanisms underlying vertical transmission and host feminization.

MEGE: Microbial Eukaryotic Genomic Evolution

SMBE-OR-120

Oxford Nanopore sequencing and eukaryotic pangenomes – strain level genome variation in the harmful algal bloom causing species Aureococcus anophagefferens.

S. J. Sibbald [1 2,*], M. Adams [3], A. J. Roger [1 2], J. M. Archibald [1 2]

[1]Biochemistry and Molecular Biology, [2]Centre for Comparative Genomics and Evolutionary Bioinformatics, [3]Biology, Dalhousie University, Halifax, Canada

Abstract: The pan-genome refers to the sum of genes across all strains of a given species, only a subset of which reside in the genome of any given strain. The pan-genome concept readily applies to prokaryotes, where lateral gene transfer (LGT) can lead to enormous intra-species gene-content variability . However, the extent to which LGT-driven pan-genomes exist in eukaryotes is uncertain. Pan-genomic investigations rely on the availability of both a high-quality reference genome and multiple genomes from closely related strains – a breadth and depth of sequencing that is typically lacking in the realm of microbial eukaryotes. As more data on strain-level genomic diversity in microbial eukaryotes become available, it is becoming evident that their nuclear genomes, like those of prokaryotes, can be remarkably dynamic over short evolutionary timescales (e.g., in the haptophyte *Emiliania huxleyi*). Using Oxford Nanopore long-read sequencing technology, we produced a new high-quality reference genome (53 Mbp in size) for the harmful algal bloom causing species *Aureococcus anophagefferens*, as well as high-quality draft genomes for multiple closely related strains. Our comparative genomic investigation indicates strain level variation in both gene content and genome size, providing insight into how microbial eukaryotes adapt and diversify.

MEGE: Microbial Eukaryotic Genomic Evolution
SMBE-OR-123
Evolution of highly scrambled genomes
L. Landweber*, R. Neme, L. Bah, X. Chen

Abstract: Ciliates such as Oxytricha trifallax possess a dynamic pair of genomes, and massive DNA rearrangements produce a highly fragmented but functional somatic macronucleus from a complex germline micronucleus. This process eliminates nearly all noncoding DNA, including transposons, and rearranges (descrambles) hundreds of thousands of short DNA segments to build a second genome containing over 16,000 gene-sized "nanochromosomes." In the precursor, germline genome, the shattered segments of different genes often interweave with each other, frequently overlap and sometimes combinatorially assemble. A mature, somatic genome contains over 16,000 nanochromosomes. We have used PacBio sequencing of the germline genome of Tetmemena, a ciliate in the same hypotrich clade asOxytricha, to investigate the evolutionary history of genome fragmentation and scrambling. We find that the level of fragmentation is highly variable between species, as is the transposon content, while the overall level of scrambling is surprisingly conserved.

MEGE: Microbial Eukaryotic Genomic Evolution

SMBE-OR-121

Comparative Analysis of Oomycete Genome Evolution

J. McGowan [1,*], K. P. Byrne [2], D. A. Fitzpatrick [1]

[1]Department of Biology, Maynooth University, Maynooth, [2]School of Medicine, University College Dublin, Dublin, Ireland

Abstract: The oomycetes are a diverse class of microscopic, filamentous eukaryotes within the stramenopiles–alveolates–rhizaria (SAR) eukaryotic supergroup. They include some of the most destructive pathogens of animals and plants, such as *Phytophthora infestans*, the causative agent of late potato blight. Despite the threat they pose to worldwide food security and natural ecosystems, there is a lack of tools and databases available to study oomycete genetics and evolution. To this end, we have developed the Oomycete Gene Order Browser (OGOB), a curated database that facilitates comparative genomic and syntenic analyses of oomycete species. OGOB incorporates genomic data for 20 oomycete species including functional annotations and a number of bioinformatics tools. OGOB hosts a robust set of orthologous oomycete genes for evolutionary analyses. Here, we present the structure and function of OGOB as well as a number of comparative genomic analyses we have performed to better understand oomycete genome evolution.

We analyse the extent of oomycete gene duplication and identify tandem gene duplication as a driving force of the expansion of secreted proteins that facilitate infection. We identify core genes that are present and microsyntenically conserved (syntenologs) in oomycete lineages and identify the degree of microsynteny between each pair of the 20 species housed in OGOB. Our results reveal an extensive degree of microsyntenic conservation amongst genes with housekeeping functions. Furthermore, phylostratigraphy was used to estimate the age of all 319,881 genes in our dataset. This approach allows us to identify trends in de-novo gene formation and study the dynamics of gene family size evolution. OGOB is available at https://ogob.ie.

MEGE: Microbial Eukaryotic Genomic Evolution

SMBE-OR-122

Ditch trees, build networks: Paradigm shift in evolutionary studies

X. Xie [1,*]

[1]Department of Biology, Virginia State University, Petersburg, United States

Abstract: Phylogenetic trees have been the quintessential methods for evolutionary analysis, and the tree thinking has become ingrained in modern biology and served as the foundation of many research projects, like the “Tree of Life” project. However, the tree thinking and phylogenetic tree (re)construction might not represent the true evolutionary history of life as they have overlooked some essential processes in evolution, including the horizontal/lateral gene transfer that has happened frequently particularly in the early evolutionary history of life and among unicellular organisms (including eukaryotic lineages), as well as hybridization that happens frequently between reproductively compatible species/lineages, and recombination that also happens so frequently within cells. Furthermore, the phylogenetic tree construction in many cases have failed to address the incomplete lineage sorting problem among closely related groups. All these reasons make network analysis better alternative to tree construction. This talk will further discuss the progresses and challenges, as well as future directions, in evolutionary network analysis.

Microbial Evolution in Complex Environments
SMBE-OR-240
CHRONIC PSEUDOMONAS AERUGINOSA INFECTIONS IN THE LUNG ENVIRONMENT.
J. Fothergill*

Are you a member of SMBE?: No
Poster Submission: Chronic bacterial infections are a key feature of a variety of lung conditions in-cluding lung infections in cystic fibrosis patients. The opportunistic bacterium, Pseudomonas aeruginosa, is extremely skilled at both colonising and persisting in the airways of patients with lung damage leading to life-long infections. It has been suggested that the upper airways (including the paranasal sinuses and nasopharynx) play an important role as a silent reservoir of bacteria. Over time, P. aeruginosa can adapt during chronic infections leading to persistence in the face of the immune system and intense therapy regimes. Here we will discuss key changes that occur early in infection leading to a diverse population. By us-ing both in vivo and specialised in vitro models of infection, we can study the trade-offs of particular mutations within P. aeruginosa and the impact of the wider bacterial population. P. aeruginosa is a master of intrinsic resistance and persistence. Understanding these early adaptive changes and interactions may be central to future treatment and control measures and the use of models al-lows us an insight into the highly complex environment that is the respiratory niche.

Microbial Evolution in Complex Environments
SMBE-OR-240B
Bacterial adaptation is constrained in complex communities
T. Scheuerl [1,*], T. Bell [1], T. Barraclough [1], D. Rivett [1], R. Novell [1]
[1]Imperial College London, London, United Kingdom

Abstract: The evolution of single species in isolation has been well studied, however, in nature species are embedded within complex communities. In communities, biotic interactions may either facilitate or constrain evolution depending on whether the interactions expand or contract the range of ecological opportunities. A fundamental challenge is to understand how the surrounding biotic community alters evolutionary trajectories as species adapt to novel environmental conditions. Here we show how community context can dramatically alter the evolutionary dynamics of bacterial populations. We find that evolution of focal bacterial strains depends on properties both of the focal strain and of the surrounding community. In particular, there was a stronger evolutionary response in low-diversity communities, and when the focal species had a larger genome and were initially poorly adapted. The findings demonstrate that adaptation to new environmental conditions can only be understood in the context of interspecific interactions.

Microbial Evolution in Complex Environments

SMBE-OR-289

Population Genetics of Helicobacter pylori in the West African archipelago of Cape Verde

Y. L. Tam*, R. Zamudio, C. Davison, S. Tallman, N. Tavares, K. S. Santos, M. Rodrigues, I. Pina, I. I. Araújo, S. Beleza

Abstract: *Helicobacter pylor*i is a pathogenic and commensal organism that is known to be coevolving with humans since the earliest known human migrations out of Africa. This has generated worldwide genetic diversification of *H. pylori* strains as in their host, giving rise to complex interactions underlying pathogenicity that are still not fully understood. Cape Verde derives genetic ancestry to African and European continents. Due to an extensive history of admixture, it is expected that in Cape Verde the bacteria of the two ancestry origins have been brought into a common environment. The study of the genetic variation of *H. pylori* and of their host in Cape Verde may then be informative about how their coevolution modulates disease risk. We have collected whole genome sequences of 543 isolates from 186 Cape Verdeans. Population genetics analyses confirmed our hypothesis of the presence of both African and European bacterial strains on the islands. Due to their ability to recombine, admixture between strains was observed. In addition, we have found that founder effects during the colonization of the islands have given rise to divergent European *H. pylori* strains with specific phenotypes. Around 20% of the individuals demonstrate multiple colonisation, including individuals infected with strains of different ancestry. This evidence shows that cross infection is taking place in Cape Verde. Within-host diversity is informative about the history of recombination of *H. pylori*. This study demonstrates Cape Verde to be a natural model for the studying of the host-*H. pylori* interactions underlying the susceptibility of the host to the infection.

Microbial Evolution in Complex Environments
SMBE-OR-240A
EVOLUTIONARY ECOLOGY OF A COMMUNITY WIDE PUBLIC GOOD
A. Buckling [1,*]
[1]Univesrty of Exeter, Penryn, United Kingdom

Are you a member of SMBE?: No
Poster Submission: Cooperation between microbes has far-reaching consequences for ecosystem functioning, agriculture and health. While we have a good understanding of the ecological and genetic contexts that drive cooperation within-species, we have less understanding of the evolutionary ecology of cooperative behaviours that benefit entire microbial communities. Such behaviours are surprisingly common, and include detoxification of metal contaminated environments via the ubiquitous production of metal-chelating siderophores. We report how inter-specific and intra-specific social interactions might shape the evolution and ecology of siderophore production in soil communities, and the potential consequences this has for environmental remediation strategies.

Microbial Evolution in Complex Environments
SMBE-OR-290
Quantifying fitness of budding yeast in periodically fluctuating environments
F. Abdul-Rahman [1,*], D. Gresham [1], D. Tranchina [1]
[1]Biology, New York University, New York, United States

Abstract: In natural environments, organisms frequently face variable conditions that fluctuate periodically. While evolutionary dynamics in static environments have been well characterized using experimental evolution, the effect of fluctuating environments on selection remains less well understood. The length of the period of fluctuation has been shown to have a significant effect on evolutionary dynamics and fitness. Partly due to the challenges in experimental design, the focus of previous studies has been on periodic fluctuations that exceed generation time. However, multiple organisms have been shown to have rapid physiological responses to transient changes in the environment, indicating that organisms have been evolutionarily tuned to respond to intragenerational fluctuations in the environment.

Here, we characterize selection on *budding yeast* growing in continuous culture using various fluctuation regimes. We use BARseq, a pooled fitness assay of the single-gene deletion library, to assay fitness across ~4000 genotypes in three distinct selective regimes: 1) static nutrient levels, 2) transient nutrient fluctuations periodically occurring at intragenerational intervals, and 3) periodic nutrient fluctuations occurring at multigenerational intervals. Using an expanded model of the chemostat we defined, and experimentally verified, nutrient fluctuations. We quantified the distribution of fitness effects for each genotype in each selective condition. We find that intragenerational and multigenerational fluctuations result in decreased fitness variance among genotypes and as a result maintain greater genetic diversity than static environments. Using genome-wide fitness profiles we identified cellular processes that contribute to fitness effects unique to each condition. Our study highlights the importance of dynamic environments in the evolutionary history of budding yeast.

Microbial Evolution in Complex Environments

SMBE-OR-107A

Biodiversity in E. coli associated with passage through soil: a landscape genomics approach.

N. Dusek [1], A. J. Hewitt [1], K. N. Schmidt [1], P. Bergholz [1,*]

[1]Microbiological Sciences, North Dakota State University, Fargo, United States

Abstract: *Escherichia coli* is a commensal inhabitant of warm-blooded animals, a major cause of bacterial food- and water-borne illnesses, a platform for biotechnology, and a model of microbial evolution. In these diverse ecological roles, *E. coli* also exhibits enormous genomic biodiversity. *E. coli* are frequently deposited by hosts into secondary (i.e., extrahost) habitats where they must persist or adapt in order to reach a new host. Our overarching hypothesis is that passage through complex environments, like soil, may select for diversity in this species. Past studies have demonstrated that: *E. coli* from soil exhibit gene x environment interactions in a subset of genes, that the prevalence of *E. coli* varies dependent on soil properties, that subspecies (i.e., phylogroups) of *E. coli* may have differing ecology, and that *E. coli* isolated from complex extrahost habitats have different propensities to form biofilms than comparable sets of fecal/commensal isolates.

We have developed a well-curated, georeferenced set of > 3,000 *E. coli* isolates from surface soil with which to test our overarching hypothesis. A landscape genomics approach was used here to address the question, "how are *E. coli* genomes structured by the soil environment in nature?" For simplicity, we focused on one phylogroup, phylogroup D (*n* = 277) for this study. Here we show that, a) *E. coli* exhibit a gradient-of-patches spatial structure even at spatial scales as large as 10^2 to 10^4 m, b) the rivers may act as biogeographic barriers to the dispersal of *E. coli* residing in soils, and c) single-nucleotide and gene content variants in *E. coli* genomes are various chemical characteristics of the soil environment.

Although the main ecological process acting on *E. coli* in soil is rapid, frequent dispersal into soil followed by persistence and death, we interpret the above findings together to conclude that *E. coli* persist long enough in the complex soil environment that the overall spatial pattern of their dispersal can be detected. This suggests that persistence may be extended enough to achieve evolutionary rescue through gene acquisition and selection in some cases. Associations between soil environmental characteristics and specific genome variants suggests and basis for adaptation to soil. These adaptations can contribute to the global *E. coli* genome provided that strains residing in soil can disperse back into hosts, as has been observed in food- and water-borne *E. coli* disease outbreaks.

Microbial Evolution in Complex Environments

SMBE-OR-249A

CRISPR-Cas adaptive immune systems of bacteria accelerate the evolution of bacteria and phages

B. N. J. Watson [1 2,*], G. P. C. Salmond [3], E. R. Westra [2], R. H. J. Staals [1 4], P. C. Fineran [1 5]

[1]Department of Microbiology and Immunology, University of Otago, Dunedin, New Zealand, [2]ESI, University of Exeter, Penryn, United Kingdom, [3]Department of Biochemistry, University of Cambridge, Cambridge, United Kingdom, [4]Laboratory of Microbiology, Wageningen University and Research, Wageningen, Netherlands, [5]Bio-Protection Research Centre, University of Otago, Dunedin, New Zealand

Abstract: Phages outnumber their bacterial hosts ten-fold in most environments. CRISPR-Cas adaptive immune systems are widespread bacterial defence mechanisms that protect cells against phage predation. While the molecular mechanisms of CRISPR have been investigated extensively, we know relatively little about the evolutionary ecology of these systems. Here, we show that the type I-F CRISPR-Cas system in *Pectobacterium atrosepticum* has a profound impact on bacterial and phage population and evolutionary dynamics. First, we found that upon phage infection, this CRISPR system causes death of both the phage and the host, hence providing a population-level benefit at the cost of the infected individual (analogous to abortive infection systems). Second, we found that CRISPR systems can drive the evolution of extensive phage morphological diversity by selecting for mutants that carry point mutations or deletions in their genomes. Third, we demonstrated that CRISPR systems also profoundly impact bacterial evolution by enhancing the spread of bacterial genes between cells via phage-mediated transduction. These novel insights help us to understand how CRISPR-Cas immune systems influence the evolution of both bacteria and phages, and how they may be selected for due to the population-level benefits they provide.

Microbial Evolution in Complex Environments

SMBE-OR-291

Host-virus coexistence in phytoplankton through the genomic lens

S. Yau [1], M. Krasovec [2], F. Benites [3], E. Vancaester [4], M. Groussin [5], S. Rombauts [6], K. Vandepoele [7], S. Gourbiere [8], G. Piganeau[*]

[1]ISM, Barcelona, Spain, [2]Oxford university, Oxford, United Kingdom, [3]CNRS, Banyuls sur mer, France, [4]Gent university - VIB, Gent, Belgium, [5]MIT, Cambridge, United States, [6]VIB, [7]VIB - Gent university, Gent, Belgium, [8]LGDP, CNRS, Perpignan, France

Abstract: Microalga-virus interactions are major determinants of geochemical cycles in the oceans, as viruses are responsible for the redirection of carbon and nutrients away from larger organisms back towards microorganisms via the lysis of microalgae in a process coined the 'viral shunt'. While the acquisition of resistance to viruses has been reported to occur in many microalga-virus systems, the genetic underpinning of the coexistence of both the microalga and the virus in culture remains enigmatic.

Here I will present the genome analysis of a novel *Ostreococcus* microalga, *O. mediterraneus*, and its ~200 kb infective nucleo-cytoplasmic large DNA virus (NCLDV), OmV2. The microalgae and the virus have been coexisting in culture for over a decade. On a short evolutionary time scale, evolution experiments through single cell bottlenecks demonstrate that, in the absence of the virus, susceptible cells evolve from one ancestral resistant single cell, and *vice–versa*; that is that resistant cells evolve from one ancestral susceptible cell. This provides evidence that the observed sustained viral production is the consequence of a minority of virus-susceptible cells. Re-sequencing of one susceptible strain demonstrated that the phase switch involved a large 60 Kb deletion of one chromosome. Mathematical modelling of this mechanism predicts microalga–virus population dynamics consistent with the observation of continuous growth of both virus and microalga.

The long term microalga-virus coexistence may thus be explained by low-level phase switching between virus-resistant and virus-susceptible phenotypes, akin to a bet hedging strategy. Altogether, our results suggest a previously overlooked strategy in microalga–virus interactions involving structural chromosome rearrangement mutations.

Microbial Evolution in Complex Environments

SMBE-OR-292

Rapid adaptations to the human host in Legionella pneumophila

D. Leenheer[1 2,*], P. Pelaz[3], M. Morin[4], D. Klingenberg[4], S. Jarraud[5 6], C. Ginevra[5 6], L. Guy[1]

[1]Department of Medical Biochemistry and Microbiology, Uppsala University, Uppsala, Sweden, [2]Ph.D. Program in Human Biology, University of Tsukuba, Tsukuba, Japan, [3]Laboratorio de Legionella, Centro Nacional de Microbiología, Instituto de Salud Carlos III, Madrid, Spain, Instituto de Salud Carlos III, Madrid,, Spain, [4]Parasitology and Waterborne Pathogens, Public Health Agency of Sweden, Stockholm, Sweden, [5]French National Reference Center of Legionella, Hospices Civils de Lyon, [6]International Center of Infectiology Research, Université Lyon 1, Lyon, France

Abstract: *Legionella pneumophila* are host-adapted bacteria that infect, and reproduce primarily in aquatic amoebae. Using similar infection mechanisms, they infect human macrophages, and cause Legionnaire's disease and Pontiac fever in humans. We hypothesize that, despite these similarities, the hosts are different enough so that there exist a few mutations with high selective value, which dramatically increase *Legionella*'s fitness in the human host. As human-to-human transmission is rare, fixation of these mutations into the population is unlikely. Therefore, mutations in the same genes observed in independent human infections, could be examples of convergent evolution. Identifying these adaptive mutations would increase our understanding of the mechanism of *Legionella* infection in the human host. By comparing a large enough number of independent infections, we expect these highly adaptive mutations to appear several times, despite the short duration of a typical infection (7-10 days).

Clinical isolates and isolates identified as their respective environmental source were sequenced and assembled using SPAdes. Variants were called using RedDog. Variants between samples were compared to identify genes that are likely candidates for human-specific adaptations in multiple samples. Isolates from independent infections were obtained and sequenced from collaborators and complemented by isolates from published studies.

60 comparisons of clinical samples and their respective environmental source identified 240 SNPs, 126 of which were non-synonymous, 191 SNPs that occurred once, 5 SNPs occuring in 2 independent comparisons, 3 in 3, and 1 in 4 comparisons. The genes mutated independently multiple times, included genes involved in virulence and motility, and outer membrane proteins.

Microbial Evolution in Complex Environments

SMBE-OR-293

Mutualistic interactions substantially alter the tempo and mode of adaptation in a model microbial community

S. Venkataram [1,*], H. Kuo [1], E. Hom [2], S. Kryazhimskiy [1]

[1]Biological Sciences, UCSD, San Diego, CA, [2]Biology, University of Mississippi, University, MS, United States

Abstract: Mutualistic interactions are ubiquitous, yet we have little direct empirical knowledge of how novel mutualistic associations evolve, and how this evolution might differ from evolution in response to purely abiotic factors. Here we leverage a powerful model microbial mutualism comprised of two species, baker's yeast (*Saccharomyces cerevisiae*) and green algae (*Chlamydomonas reinhardtii*) to study the effect of mutualistic interactions on adaptation. We conducted laboratory evolution experiments of diploid yeast either in monoculture or in coculture with haploid algae under otherwise identical growth conditions to isolate the effects of mutualistic interactions from the abiotic environment. In addition, we use genetically engineered yeast carrying DNA "barcodes" – random chromosomally-integrated sequences that enabled us to track the dynamics of tens of thousands of yeast subpopulations within each replicate culture.
We find that yeast evolved in monoculture generate substantially fewer independent adaptive events and consequently adapt at a slower rate. We also conducted whole-genome sequencing of ~270 evolved isolates and identified copy number variants (CNVs) as the major mode of adaptation in both treatments. However, the two treatments differed in the specific loci targeted by the CNVs, a pattern which was also observed in the single nucleotide variants. Finally, we show that monoculture and coculture adapted yeast have significantly different effects on algal density, suggesting the existence of a rapid eco-evolutionary feedback between the species. Our results suggest that mutualistic interactions have a powerful impact on the rate, genetic basis and ecological outcomes of short-term evolution.

Mitochondrial-Nuclear Interactions

SMBE-OR-008

THE STOICHIOMETRY OF CYTONUCLEAR GENE EXPRESSION

D. B. Sloan [1,*]

[1]State Univesity of Colorado, Fort Collins, United States

Are you a member of SMBE?: No

Poster Submission: Key enzyme complexes within mitochondria and plastids are chimeric in the sense that they comprise gene products derived from two different genomes (nuclear vs. cytoplasmic). Like all multisubunit enzymes, these complexes require balanced expression of their subunits, but with the added challenge of coordinating this balance across genomes with radically different gene-expression mechanisms. We investigated the relationship between nuclear and cytoplasmic transcript abundances and found massive deviations from equimolarity. Across diverse eukaryotes, mitochondrial and plastid-encoded subunits have much higher transcript levels than their nuclear counterparts, despite the need for the resulting protein products to assemble into stoichiometrically balanced complexes. This asymmetry appears to be a general feature of nuclear vs. cytoplasmic genomes because genes that have undergone recent functional transfer from the mitochondrial genome to the nucleus have concurrently adopted typical nuclear-like transcript abundances. To investigate the functional components that may regulate subunit abundance and stoichiometry at the protein level, we have focused on the plastid caseinolytic protease (Clp), which exhibits enormous heterogeneity in evolutionary rates across angiosperms. Applying the principle of evolutionary rate covariation, we performed genome-wide screens for plant genes that co-accelerate with the Clp complex and detected strong associations with loci involved in both plastid and cytosolic protein translation. Overall, our results lead us to hypothesize that constitutively high levels of cytoplasmic transcript expression reflect relatively inefficient transcriptional-level gene regulation in mitochondria and plastids, whereas protein-level regulation of cytonuclear stoichiometry is likely both developmentally dynamic and subject to rapid evolutionary remodeling in groups like angiosperms.

Mitochondrial-Nuclear Interactions

SMBE-OR-010

Exploring the contribution of mito-nuclear ancestry discordance to disease risk in admixed individuals from the Penn Biobank

A. Zaidi [1,*], K. Makova [2], I. Mathieson [1]

[1]Genetics, University of Pennsylvania, Philadelphia, [2]Biology, Pennsylvania State University, University Park, United States

Abstract: Proper mitochondrial function relies on finely-tuned interactions between nuclear and mitochondrial genes as most mitochondrial proteins (>1,000) are encoded by the nuclear genome. Discordance in ancestry between mitochondrial and nuclear genes has been shown to lead to decreased fitness in non-human hybrid populations. In humans, mito-nuclear discordance has been shown to be associated with risk of pre-term birth in admixed populations. However, the extent to which mito-nuclear interactions contribute to variation in health-related outcomes in humans is broadly unknown.

We show that mito-nuclear discordance, measured as the degree of divergence between nuclear and mitochondrial ancestry, is negatively correlated with mtDNA copy number, estimated from whole-genome sequence alignments, in 341 recently admixed individuals from the 1000 Genomes Project. This suggests that mtDNA replication efficiency or mitochondrial biogenesis might be impacted by discordance between nuclear and mitochondrial ancestry.

To explore whether mito-nuclear discordance has an effect on disease risk, we carried out a phenome-wide association study (PheWAS) in 2,031 people of African-American ancestry from the Penn Biobank study. We used genotype data to call local ancestry and off-target reads from whole-exome data to call mtDNA haplogroups. These individuals have, on average, 79% African nuclear ancestry, 87% African mitochondrial ancestry, and mito-nuclear discordance of 0.26 (with standard deviation 0.21). Finally, we tested whether mito-nuclear discordance is associated with international classification of disease (ICD9) codes. Our study clarifies the relative contribution of mito-nuclear interactions to human disease risk.

Mitochondrial-Nuclear Interactions

SMBE-OR-011

Nuclear-encoded nutrition sensing pathways regulate transmission of 'selfish' mitochondrial mutations

B. L. Gitschlag [1], M. R. Patel [1,*]

[1]Department of Biological Sciences, Vanderbilt University, Nashville, United States

Abstract: Hundreds to thousands of wildtype mitochondrial genomes are normally present in each cell. However, a fraction of these genomes can acquire pathogenic mutations and coexist with wildtype copies in a state called heteroplasmy. Heteroplasmic variants that persist despite being deleterious to the host organism can be viewed as selfish genetic elements. We sought to determine the selection forces that allow such selfish mitochondrial genomes to persist in metazoan populations. In the model system *Caenorhabditis elegans*, we have discovered that nuclear-encoded nutrition sensing pathways are important determinants of heteroplasmic mutant levels. Specifically, we discovered that mutant genomes have a significant transmission advantage over wildtype mitochondrial genomes in the female germline. When exploring the physiological basis for this transmission advantage, we found that the advantage is substantially reduced under nutrient limited conditions. Subsequent imaging and genetic analyses revealed that transmission bias in favor of mutant mitochondrial genomes is similarly reduced in animals with defective nutrition sensing. Taken together, these data suggest that selfish mutant mitochondrial genomes exploit nuclear-encoded nutrition sensing pathways in the female germline to persist and proliferate. This is the first study to investigate the impact of nuclear-encoded nutrition sensing pathways on the regulation of mitochondrial genome transmission dynamics. By investigating impact of the nuclear genome on mitochondrial transmission dynamics, we provide insights into an important but understudied dimension of mitochondrial-nuclear interactions.

Mitochondrial-Nuclear Interactions

SMBE-OR-009

THE NUCLEAR SUPERVISION OF MITOCHONDRIAL DNA TRANSMISSION

H. Ma [1,*]

[1]Gurdon Institute, Cambridge, United Kingdom

Are you a member of SMBE?: No

Poster Submission: Fundamental biological processes rely on harmonic interactions between nuclear and mitochondrial genomes. These interactions, which are key to complex life, are under intense selection to maintain the integrity of the mitochondrial function. Several features of mtDNA make it likely to accumulate mutations. When this occurs, the nuclear genome sometimes will play catch-up to ensure the high compatibility of two genomes. Previously, a genetic system to isolate mtDNA mutants has been developed in Drosophila. We use these mutants to identify and study nuclear modifiers that can suppress detrimental effects of mitochondrial mutations. We also investigate how the nuclear genome influences the transmission of mitochondrial mutants in the female germline. These projects will contribute to our understanding of the mito-nuclear epistatic interactions. The knowledge of the mechanism by which a nuclear protein suppresses the detrimental effect of mitochondrial mutants can also contribute to future disease treatment.

Mitochondrial-Nuclear Interactions

SMBE-OR-012

Co-regulation of the mitochondrial and nuclear gene expression – adaptation to the host regulatory system

D. Mishmar*, G. Barshad [1], A. Blumberg [1], T. Cohen [1]

[1]Life Sciences, Ben-Gurion University of the Negev, Beer-Sheva, Israel

Abstract: Endosymbiosis was accompanied by transfer of most mitochondrial DNA (mtDNA) genetic information to multiple nuclear chromosomal locations, and establishment of an independent transcriptional regulation for each gene. This suggests sharp separation between the transcriptional regulation of nuclear DNA (nDNA)-encoded mitochondrial genes and their mtDNA-encoded partners, which are co-transcribed in polycistrones using dedicated transcription factors and a mitochondrial RNA polymerase. However, nDNA and mtDNA-encoded gene products physically interact to maintain mitochondrial function in different cells and tissues. Therefore, we hypothesized that coordination of gene expression regulation should occur between the two genomes, especially in the frame of the oxidative phosphorylation system (OXPHOS). If this is true, one expects adaptation of the ancient prokaryote (the mitochondrion) to the regulatory system of the eukaryote host. Our recent analysis of ~8500 RNA-seq experiments from 48 different human body sites, revealed evidence for co-regulation between bi-genomic OXPHOS subunits across most tissues. Our analysis highlighted several candidate regulators of the coordinated mito-nuclear gene expression, which are known transcriptional regulators of nDNA gene expression. Using ChIP-seq and DNase-seq experiments we found that these factors bind in vivo the human mtDNA outside of the control region, within the gene sequence, suggesting that the mtDNA gene-coding sequences also harbor overlooked regulatory information. Silencing of two such regulatory candidates, i.e. CCAAT/enhancer-binding protein beta (CEBPb) and c-Jun, led to elevated mtDNA transcript levels in human cells, suggesting a transcriptional repressive role. When CEBPb-silenced glioblastoma human cells were subcutaneously injected into mice, significant reduction in tumor size was observed (as compared to control), suggesting in vivo phenotypic impact for altered mito-nuclear co-regulation. The previously overlooked mitochondrial role of known nuclear transcriptional regulators will be discussed.

Mitochondrial-Nuclear Interactions

SMBE-OR-013

Cytonuclear interactions drive biased gene retention in the face of extensive nuclear introgression

E. S. Forsythe*, A. D. L. Nelson, D. B. Sloan, M. A. Beilstein

Abstract: Previous single-gene phylogenetic analyses of Arabidopsis and its close relatives revealed cytonuclear discordance, suggesting a potential history of introgressive hybridization. We applied phylogenomic techniques and several complementary statistical tests to show that introgression occurred, impacting our understanding of species relationships in the group. We used coalescent simulations to develop a novel statistical method, *Divergence-based Introgression Polarization (DIP),* which distinguishes introgression donors from recipients using whole genome sequences. Application of *DIP* indicated that cytonuclear discordance appears to have arisen via extensive asymmetric nuclear introgression, rather than cytoplasmic introgression, meaning that most of the nuclear genome was displaced by exogenous alleles during introgression, while only the cytoplasmic genomes and a small proportion of the nuclear genome were retained in their native state. The small fraction of the nuclear genome that was retained amidst the onslaught of foreign alleles is composed of ancient haplotype blocks that are enriched for genes that target to the mitochondria and chloroplasts, suggesting that selection favored compatible combinations of nuclear and cytoplasmic alleles during introgression, thus maintaining proper cytonuclear function.

Molecular basis of neural circuit and behavioral evolution
SMBE-OR-228
The evolution of predatory feeding and self-recognition in nematodes: from switch genes and a self-recognition peptide to neural circuits
R. Sommer [1,*], J. Lightfoot [1]
[1]Max Planck Insitute for Developmental Biology, Tuebingen, Germany

Abstract: Ever since Darwin, biologists are intrigued about evolution and its underlying mechanisms. In animals, neurobiological traits are key features that shape evolutionary diversity and novelty. We use an interdisciplinary approach that integrates development, neurobiology, ecology and population genetics to unravel the mechanistic changes that give rise to novelty and change. For this, we established *Pristionchus pacificus* as model system and combine laboratory studies building on genetic, genomic and transgenic tools with fieldwork. *P. pacificus* lives in association with scarab beetles. One key feature of its life style is a mouth-form dimorphism that enables predatory feeding. The development of teeth-like denticles of two different forms represents an example of developmental plasticity and we test the hypothesis that developmental plasticity is a facilitator of phenotypic diversification and the evolution of novelty. I will describe the molecular mechanisms underlying feeding plasticity and will show how the genetic machinery and the environment interact to specify this phenotypically plastic trait.
Studying different *Pristionchus* species and a diversity of *P. pacificus* wild isolates, we identified the first organismal self-recognition system in nematodes[1]. I will summarize our subsequent genetic studies that identified the peptide SELF-1 to represent the major ligand in *P. pacificus* that prevents cannibalism in this system. In the second part of my talk, I will describe our studies that investigate the neurobiology of predation and self-recognition[1]. Building on nervous system reconstruction in *P. pacificus*, we have shown that predatory feeding behavior can be induced by serotonin treatment. CRISPR/Cas9 induced mutants in serotonin biosynthesis genes and genetic ablation studies indicate the exact role of serotonin and serotonergic neurons and identify a novel role of serotonin in in nervous system function.

Reference (1): Lightfoot et al, *Science* 2019, in press

Molecular basis of neural circuit and behavioral evolution

SMBE-OR-230

The making of an olfactory specialist

T. O. Auer [1,*], M. A. Khallaf [2], A. F. Silbering [1], G. Zappia [1], K. Ellis [3], B. S. Hansson [2], G. S. Jefferis [4], S. Caron [3], M. Knaden [2], R. Benton [1]

[1]Center for Integrative Genomics, University of Lausanne, Lausanne, Switzerland, [2]Department of Evolutionary Neuroethology, Max Planck Institute for Chemical Ecology, Jena, Germany, [3]Department of Biology, University of Utah, Salt Lake City, United States, [4]Division of Neurobiology, MRC Laboratory of Molecular Biology, Cambridge, United Kingdom

Abstract: The evolution of animal behaviour is poorly understood. Despite numerous correlations of behavioural and nervous system divergence, demonstration of the genetic basis of interspecific behavioural differences remains rare. Here we develop a novel neurogenetic model, Drosophila sechellia, a close cousin of D. melanogaster that displays profound behavioural changes linked to its extreme host fruit specialisation. While D. melanogaster is a worldwide ecological generalist, D. sechellia, an island endemic, shows strong preference for a single food source, the “noni” fruit of the tropical shrub M. citrifolia.

Through calcium imaging, we identify olfactory pathways in this species detecting host volatiles. Mutational analysis indicates roles for individual receptors in long- and short-range attraction. Cross-species allele transfer demonstrates that differential tuning of one receptor is important for species-specific behaviour. We identify the molecular determinants of this functional change, and characterise their behavioural significance and evolutionary origin. Circuit tracing reveals that receptor adaptations are accompanied by increased sensory pooling onto interneurons and novel central projection patterns.

To pinpoint the underlying genetic basis of species differences in sensory neuron numbers we are employing a quantitative trait locus mapping approach to identify causal genetic loci. Our work links molecular and neuronal changes to behavioural divergence and defines a powerful model for investigating nervous system evolution and speciation.

Molecular basis of neural circuit and behavioral evolution

SMBE-OR-231

Revealing functional significance and driving force to maintain genetic variants of VMAT1 shed light on the evolution of psychological diversity in humans

D. Sato [1,*], Y. Ishii [1], T. Nagai [1], M. Kawata [1]

[1]Graduate School of Life Sciences, Tohoku University, Sendai, Japan

Abstract: Recent advances in genomic technologies allow us to elucidate genetic substrates underlying human psychology and its diversity. However, little is known about the evolutionary mechanism and history behind it. We have previously determined vesicular monoamine transporter 1 (*VMAT1*) as being positively selected in the human lineage with two human-specific amino acid substitutions (Glu130Gly, Asn136Thr/Ile). VMAT1 regulates the transport of monoamine neurotransmitters into synaptic vesicles, and several functional mutations have been reported. Among them, Thr136Ile (rs1390938) belongs to a luminal loop domain, which interacts with G proteins, and is known to up/downregulate the monoamine transport and to be associated with psychotic symptoms, such as anxiety and neuroticism. Interestingly, the genomic signature around VMAT1 suggests that 136Ile had later emerged along with Out-of-Africa migration of modern humans and that balancing selection has been acting on the Thr136Ile in non-African populations (Sato & Kawata, 2018 *Evol. Lett.*). In the present study, we aimed to elucidate the evolutionary history and mechanisms generating and maintaining the Thr136Ile variant and conducted two studies as below. (1) Functional assays measuring neurotransmitter uptake among reconstructed ancestral VMAT1 proteins showed that the transporter activity had initially declined, possibly suggesting that higher levels of anxiety had been favored through the early phase of human evolution. (2) Large Japanese cohort dataset revealed the signature of heterozygotic advantage (overdominance) of the Thr136Ile, which could be a reasonable mechanism to cause balancing selection on the site. In summary, our study presents an unusual evolutionary pattern of monoaminergic pathways, underlying psychological diversity in humans.

Molecular basis of neural circuit and behavioral evolution

SMBE-OR-229

Evolution of brain and behaviors in blind cavefish: the MAO (MonoAmine Oxidase) mutation.

S. Rétaux [1,*]

[1]Paris-Saclay Institute of Neuroscience, Gif sur Yvette, France

Abstract: mutation.Behavioral adaptation in response to environmental changes are crucial to the survival of species. Our group studies the micro-evolution of brain and behaviors in a fish model, Astyanax mexicanus. Within this species, there are several morphotypes: surface fish, living in the rivers of Central America and cavefish, which are blind and depigmented, living in Mexican caves. Cavefish and surface fish have evolved from eyed surface-like ancestors. During their rapid evolution, in less than 20,000 years in total and permanent darkness, cavefish have undergone numerous morphological and behavioral changes. In particular cavefish present a "behavioral syndrome", due to both sensory and central neuro-modulatory changes, and which includes increased locomotor activity, loss of aggressive behavior and schooling, and reduced sleep. We aim at understanding the genetic and neural bases of the behavioral adaptation of A. mexicanus to the cave environment.We have identified a point mutation in the coding sequence of the cavefish Monoamine Oxidase (MAO), the serotonin degrading enzyme. As accumulated evidences suggest that changes in neuro-modulatory systems generate significant variations in complex behaviors, we then addressed the following question: what are the consequences of the MAO mutation at the level of the serotonergic system and the behavior in cavefish? Combining protein structure 3D modelling and biochemical measurements, we found that the mutation changes the orientation of residues at the entrance of the active site and is exclusively responsible for increased levels of serotonin (5HT), dopamine (DA) and noradrenalin (NA) levels in the cavefish brain as compared to “wild-type” surface fish morphs. At behavioral level, the mutation does not seem to be responsible for the loss of aggressiveness and schooling in cavefish, but participates to its locomotor hyperactivity in specific conditions. Finally, the establishment of a phylogeographic map of MAO mutated alleles in various cave populations of partly independent origin allows discussing the origin of the mutation in terms of population biology. The contribution of the MAO mutation to the adaptation of cavefish to life in the dark will be discussed.

Molecular basis of neural circuit and behavioral evolution
SMBE-OR-232
Genetic and neural evolution of a mosquito odorant receptor associated with preference for human hosts
J. L. Zung [1,*], Z. Zhao [1 2], C. S. McBride [1 2]
[1]Department of Ecology and Evolutionary Biology, [2]Princeton Neuroscience Institute, Princeton University, Princeton, New Jersey, United States

Abstract: The mosquito *Aedes aegypti* offers a striking example of behavioural evolution in a genetically tractable organism. The "domestic" form of this mosquito vigorously seeks out human blood-meals, while the ancestral forest form often prefers biting non-human animals. Previous work implicated two odorant receptors in this host-preference shift (McBride et al. *Nature* 2014). The present study focusses on one of these candidates, Or103.

Or103 haplotypes segregating in domestic and forest populations are remarkably divergent. Putatively non-functional haplotypes are more abundant among human-preferring vs. animal-preferring F2 hybrids. This suggests that evolution of preference for humans is associated with selection for reduced Or103 function.

We also mapped Or103-sensory-neuron projections to the antennal lobe, the first relay point for olfactory information. Surprisingly, Or103 neurons project to multiple glomeruli in the antennal lobe. This unprecedented finding challenges the dogma that olfactory sensory neurons expressing a given receptor converge upon a single glomerulus. Furthermore, Or103 neurons' projection pattern is sexually dimorphic. At least one glomerulus is Or103-positive in females but not in males—and may therefore be particularly important for processing host cues. This same glomerulus is also the one most reduced in size in domestic vs. forest females.

Our results thus far suggest that a suppression of Or103 signal accompanied the evolution of human-preferring domestic mosquitoes from animal-preferring ancestors. Our next step is to find a ligand for Or103. In this ongoing work, we aim to understand how changes in a single odorant receptor help tweak an intricate olfactory code to reprogram a complex behaviour.

Molecular basis of neural circuit and behavioral evolution

SMBE-OR-233

Enzyme function shapes the evolution of bioluminescent courtship signals in sea fireflies

N. M. Hensley [1,*], E. A. Ellis [1], G. A. Gerrish [2], E. Torres [3], J. P. Frawley [2], T. J. Rivers [4], T. H. Oakley [1]

[1]Ecology, Evolution, & Marine Biology, University of California, Santa Barbara, Santa Barbara, [2]Biology, University of Wisconsin La Crosse, La Crosse, [3]Biological Sciences, California State University Los Angeles, Los Angeles, [4]Ecology and Evolutionary Biology, University of Kansas, Lawrence, United States

Abstract: Mating behaviours are diverse and noteworthy, especially within species radiations where they may contribute to speciation. Studying how differences in mating behaviours arise between species can help us understand how diversity is generated at multiple biological levels. The bioluminescent courtship displays of cypridinid ostracods (or sea fireflies) are an excellent system for this because amazing variety evolves while using a conserved biochemical mechanism. We find that the evolution of one aspect in this behavioural phenotype—the duration of bioluminescent courtship pulses—is shaped by biochemical function.We collected biochemical, phylogenetic, and behavioural data for 16 species, and used phylogenetic comparative models to show that differences in biochemical reaction are non-linearly correlated with the duration of courtship pulses. This relationship indicates that changes to both enzyme (c-luciferase) function and usage have shaped the evolution of courtship displays, but that they differentially contribute to, and may even constrain, phenotypic change. Next, we compared these in vivo functional results to c-luciferase sequences generated from de novo transcriptomes in a subset of species, predicting that certain sequence changes are correlated with shifts in biochemical function. From this, we reason that a small number of sequence changes are associated with divergent biochemistries, and which represent candidate sites to probe the relationship between c-luciferase sequence and function. Together, these findings demonstrate how unappreciated diversity of biochemical functions may influence the diversification of behaviors during evolution.

Mutation Rate Evolution

SMBE-OR-227

INFERENCE OF THE DISTRIBUTION OF FITNESS EFFECTS OF SPONTANEOUS MUTATIONS IN CHLAMYDOMONAS REINHARDTII USING MA LINE CROSSES

P. Keightley [1,*], K. Böndel, S. Kraemer, T. Samuels, D. McClean, J. Lachapelle, R. Ness, N. Colegrave

[1]Univesrity of Edinburgh, Edinburgh, United Kingdom

Are you a member of SMBE?: No

Poster Submission: Inferring the distribution of fitness effects of spontaneous mutations (the DFE) has been a long standing goal in evolutionary genetics. Spontaneous mutation accumulation (MA) experiments can provide information on the DFE, but typically each MA line contains many mutations, and only the combined effects of mutations on fitness have been estimated. To infer the effects of individual spontaneous mutations, we have crossed MA lines of the green alga Chlamydomonas reinhardtii to an unmutated ancestral strain to generate many recombinant lines each carrying an average of 50% of the accumulated mutations in a variety of combinations. We then determined the complement of mutations and estimated the fitness for each recombinant line. We inferred the DFE by a Bayesian mixture model assuming that the effects of mutations are sampled from a number of different distributions. A two sided gamma distribution provides a good fit to the data. We show that the DFE is highly leptokurtic, and that an significant proportion (about one-sixth) of mutations with absolute effects exceeding 1% increase fitness in the laboratory environment. The approach is limited by the precision of measuring phenotypes of mutations with very small effects.

Mutation Rate Evolution

SMBE-OR-222

Effects of nucleosome organization on human germline mutation rate variation

C. Li [1,*], N. Luscombe [1]

[1]Bioinformatics and Computational Biology Group, Francis Crick Institute, London, United Kingdom

Abstract: Understanding the patterns and genesis of germline de novo mutations is important for studying genome evolution and human diseases. Nucleosome organization is suggested to be a contributing factor to mutation rate variation across the genome. However, the small number of published de novo mutations and the low resolution of earlier nucleosome maps limited our understanding of how nucleosome organization affects germline mutation rates in the human genome. Here, we systematically investigated the relationship between nucleosome organization and fine-scale mutation rate variation by analyzing >300,000 de novo mutations from whole-genome trio sequencing and high-resolution nucleosome maps in human. We found that de novo mutation rates are elevated around strong, translationally stable nucleosomes, a previously under-appreciated aspect. We confirmed this observation having controlled for local sequence context and other potential confounding factors. Analysis of the underlying mutational processes suggests that the increased mutation rates around strong nucleosomes are shaped by a combination of low-fidelity replication, frequent DNA damage and insufficient/error-prone repair in these regions. Interestingly, strong nucleosomes are preferentially located in young SINE/LINE elements, implying frequent nucleosome re-positioning (i.e. shifting of dyad position) and their contribution to hypermutation at new retrotransposons during evolution. These findings provide novel insights into how chromatin organization affects germline mutation rates and have important implications in human genetics and genome evolution.

Mutation Rate Evolution

SMBE-OR-223

The adaptive architecture of intra-organism mutation rate variation in plants

L. Wang [1], Y. Ji [1], Y. Hu [1], H. Hu [1], X. Jia [1], M. Jiang [1], X. Zhang [1], L. Zhao [1], Y. Zhang [1], Y. Jiang [1], C. Qin [1], L. Yu [1], J. Huang [1], S. Yang [1], D. Tian [1], L. D. Hurst [2,*]

[1]State Key Laboratory of Pharmaceutical Biotechnology, Nanjing University, Nanjing, China, [2]Milner Centre for Evolution, University of Bath, Bath, United Kingdom

Abstract: Given the disposability of somatic tissue, selection can favour a higher mutation rate in the early segregating soma than in germline, as seen in some animals. While in plants intra-organismic mutation rate heterogeneity is poorly resolved, the same selectionist logic can predict a lower rate in shoot than in root, in longer lived terminal tissues (e.g. leaves) than in ontogenetically similar short-lived ones (e.g. petals) and that mutation rate heterogeneity should be deterministic with no significant differences between biological replicates. To address these expectations we sequenced over 750 genomes from various tissues of eight plant species. Consistent with a selectionist model, the rate of mutation accumulation per unit time in shoot apical meristem is lower than that in root apical tissues in perennials, where a high proportion of mutations in shoots are themselves transmissible, but not in annuals where somatic mutations tend very rarely to be transmissible. Similarly, the number of mutations accumulated in leaves is commonly lower than that within a petal of the same plant and there is no more heterogeneity in accumulation rates between replicate branches than expected by chance. High mutation accumulation in runners of strawberry is, we argue, the exception that proves the rule, as mutation transmission patterns indicate that runner has a restricted germline. We conclude that some mutation rate variation between tissues is consistent with selectionist theory.

Mutation Rate Evolution
SMBE-OR-226
MODELLING MUTATION FROM INDIVIDUALS TO SPECIES
A. Scally [1,*]
[1]University of Cambridge, Cambridge, United Kingdom

Are you a member of SMBE?: No
Poster Submission: Germline mutation processes connect developmental, ecological and evolutionary geneticsacross the broadest range of scales. Thanks to the widespread availability of genomesequencing, we now have estimates for present-day de-novo mutation rates in each of thehominid genera as well as one or two other primate taxa. For humans in particular we haveincreasingly detailed knowledge of the genomic distribution of de novo mutation, and arebeginning to explore the structure of the germline and its influence on mutation rates.However we are still limited in our ability to extrapolate these phenomena into theevolutionary past, and key aspects of the germline even in present-day organisms remaindifficult to access. I will discuss current and competing hypotheses about the origins ofgermline mutation, including multi-nucleotide mutation processes, and how computationalmodelling approaches combined with deep sequencing can help us infer mutationalparameters and processes on large and small scales. In particular I will focus on whetherevolutionary variation in mutation rates can be fully explained by life-history phenomena, andthe inference of germline parameters in the early and developing embryo.

Mutation Rate Evolution

SMBE-OR-224

Archaic fragment lengths and mutation spectrum differences suggest historical generation time changes among human populations

M. Coll Macià [1,*], L. Skov [1 2], M. H. Schierup [1]

[1]Bioinformatics Research Centre, Aarhus University, Aarhus C, Denmark, [2]Evolutionary Genetics, Max Planck Institute, Lepzig, Germany

Abstract: Modern humans and Neanderthals interbred around 50,000 years ago, leaving 1-3% Neanderthal DNA in all non-African populations. These Neanderthal fragments have since been broken down into small pieces by recombination. The fragment length distribution (FLD) is informative about the number of generations that has passed since the admixture event. We detect Neanderthal fragments in the individuals of Simons Genome Diversity Project and find that FLD differs among the different continental populations, e.g. with a mean length in West Eurasian population, present-day and ancient samples, smaller than the East Asian. We find that the difference in FLD cannot be explained by multiple pulses of archaic introgression into East Asian populations or gene flow from West Eurasian populations into Africa. An alternative explanation is that the recombination clock must have run at different speeds across the world, with the fastest in West Eurasia and slowest in East Asia.

The number of *de novo* mutations and their spectrum depend on parental age. Therefore, comparing the accumulation of derived alleles and their type among populations is also informative about changes in generation time. After masking regions with archaic introgression and variants seen in Africa, we find that the West Eurasians have more derived alleles than East Asians. We then focused on the mutational spectrum and first replicate the differences between populations previously reported, including the European enrichment of C > T transitions in a TCC three base pair context. We find a strong correlation between FLD and the mutation types strongly affected by parental age in the deCODE dataset of 1,536 trios.

All together, we estimate that Europeans have had the shortest historical generation time up to a 20% shorter compared to East Asians and hypothesise that the European enrichment in TCC transitions is due to differences in generation times. Our results imply that the human generation time has been dependent on environmental or cultural differences in ancient societies, perhaps related positively with population densities or grain based versus meat-based life styles.

Mutation Rate Evolution

SMBE-OR-225

Extremely High Rates and Intraspecific Variation in Microsatellite Mutation Profiles in Daphnia magna

S. Schaack[*], E. Ho[1], F. Macrae[1], L. Latta[2], M. Benner[1], C. Sun[3], D. Ebert[4]

[1]Biology, Reed College, Portland, [2]Biology, Lewis-Clark State University, Lewiston, United States, [3]Institute of Apicultural Research, Chinese Academy of Agricultural Sciences, Beijing, China, [4]Department of Environmental Sciences, University of Basel, Basel, Switzerland

Abstract: Microsatellite loci (tandem repeats of short nucleotide motifs) are highly abundant in eukaryotic genomes and are often used as genetic markers because they can exhibit variation both within and between populations. Although widely recognized for their mutability and utility, the mutation rates of microsatellites have only been empirically estimated in a few species, and have rarely been compared across genotypes and populations within a species. Here, we investigate the dynamics of microsatellite mutation over long- and short-time periods by quanitfying the starting abundance and mutation rates for microsatellite repeats for six different genotypes of *Daphnia magna*, an aquatic microcrustacean, collected from three populations (Finland, Germany, and Israel). Using whole-genome sequences of the 6 starting genotypes, 47 mutation accumulation (MA) lines, and 6 large population controls (non-MA lines), we find each genotype exhibits a distinctive microsatellite profile starting out which clusters according to the population of origin. During the period of mutation accumulation, we observe motif-specific, highly variable, and rapid microsatellite mutation rates across genotypes of *D. magna*, the average of which is order of magnitude greater than the recently reported rate observed in a single genotype of the closely-related congener, *D. pulex*. In our experiment, genotypes with more microsatellites starting out exhibit greater losses and those with fewer microsatellites starting out exhibit greater gains—a context-dependent mutation bias that has not been reported previously and which implies some repeat content in the genome may reflect some kind of equilibrium. We will comment on how genotype-specific mutation rates and spectra, in conjunction with evolutionary forces, can shape both the differential accumulation of mutations in the genome and the evolution of mutation rates over long time periods.

Novel insights into evolutionary genetics from emerging technologies

SMBE-OR-188

Revealing the organization and evolution of Drosophila centromeres with single molecule sequencing

C.-H. Chang [1], X. Wei [2], L. Hemmer [1], B. Santinello [3], A. Chavan [3], J. Palladino [3], B. G. Mellone [3], A. Larracuente [*]

[1]University of Rochester, [2]University of Rochester Medical Center, Rochester, [3]University of Connecticut, Storrs, United States

Abstract: Single molecule long-read sequencing has significantly improved our ability to assemble repetitive genomic regions. However, some genomic regions remain a major challenge. Progress towards understanding the organization of DNA sequences underlying centromeres has been slow, even in species with highly contiguous well-characterized genomes like *Drosophila melanogaster*. Centromeres have an essential function in coordinating chromosome segregation during cell divisions. Despite this conserved role, they are rich in repeats and among the most rapidly evolving regions of the genome. Drosophila centromeres are embedded deep in simple satellite repeats and have been recalcitrant to genome assembly. We used long-read sequencing from Pacific Biosciences (PacBio) to create heterochromatin-enriched genome assemblies and used these assemblies to identify the putative centromeres of each chromosome in *D. melanogaster.* We discovered that centromeres contain simple satellite DNAs interrupted by islands of complex DNA consisting of transposable elements and other AT-rich sequences. We validated our centromere assemblies using ChIP-seq with the centromere-specific histone variant, CENP-A, and fluorescence *in situ* hybridization to show that they are indeed centromeric. The centromeres of each chromosome are unique, but show similarities in composition and organization. All *D. melanogaster* centromeres contain one non-LTR retroelement in particular, *G2/Jockey-3*. Our population genetic analyses suggest that *G2/Jockey-3* is a recently active retroelement that does not specifically target centromeres, but its prevalence at centromeres suggests that it may be important. To determine if centromeric sequences are conserved between species, we used deep PacBio sequencing and our heterochromatin-enriched assembly methods in the closely related species *D. simulans*, *D. sechellia,* and *D. mauritiana.* We discovered that the simulans clade centromeres are distinct from those in *D. melanogaster*, but show similarities in composition and organization. Interestingly, *G2/Jockey-3* is also among the most CENP-A-enriched retroelements in *D. simulans*. Our study therefore reveals a striking level of conservation among otherwise rapidly evolving sequences at centromeres, suggesting that *G2/Jockey-3* may be important for centromere function and evolution. Retroelements are associated with centromeres in several species of plants, fungi, and animals, and therefore may be an important feature of centromere biology. Combining long-read sequencing with methods in molecular biology and cytogenetics thus is a promising approach to study the evolution of the most enigmatic regions of genomes.

Novel insights into evolutionary genetics from emerging technologies
SMBE-OR-188B

Population level cis-regulatory variation highlights the importance of DNA shape to transcription factor binding affinity

W. Nash [1,*], Y. Cai [1], N. Patron [1], W. Haerty [1]
[1]Earlham Institute, Norwich, United Kingdom

Abstract: Cis-regulatory variation has long been understood to play an essential role in evolutionary processes. Genetic variatiants in cis-regulatory modules are known to impact gene expression, and are suggested to be a key substrate for adaptation. Despite the importance of such variation, the study of cis-regulatory evolution has mostly been restricted to model species, with even these being represented by partially complete data sets. One key barrier to expanding such data has been the lack of a method for validating and testing predicated regulatory relationships in a quick, affordable, and scalable manner. Here, we present a novel plate-based method to test and quantify the effect of genetic variation wihtin binding sites on TF binding affinity. Our transcription factor relative affinity measurement protocol (tfRAMP) is being developed to use affordable components to allow robust testing of the effect of variants in transcription factor binding site (TFBS) sequences on the binding affinity of associated TFs. In order to generate high confidence predictions of loci at which variation may effect binding affinity, we implemented a novel bioinformatic pipeline to process and integrate publicly available information on transcription factor binding and population level genome wide variation. We applied our pipeline to *Arabidopsis thaliana*, generating and testing predictions of TFBS binding affinity. We tested both base and shape read out of TFBS and predicted the effects of SNP level variation on both of these essential components of TF binding affinity. We validated our predictions using the tfRAMP and, for variants leading to significant alteration in binding affinity, by placing them into a protoplast expression system in order to quantify their impact on gene expression levels.

Novel insights into evolutionary genetics from emerging technologies

SMBE-OR-188A

Genomics Aotearoa: Establishing high quality genomes alongside indigenous populations in New Zealand

A. Mc Cartney[*], D. Chagne [1], A. Santure [2], T. Buckley [3]

[1]Plant and Food Research, Palmerston North, [2]University of Auckland, [3]Systematics, Manaaki Whenua, Auckland, New Zealand

Abstract: Taonga species are those that have a special cultural significance amongst Māori in Aotearoa. As part of Genomics Aotearoa, a high-quality genomes project has been established alongside Māori to generate pipelines for the assembly of endemic and taonga species in New Zealand. These pipelines are specifically targeted at key species that are on the verge of extinction, treasured by Māori, key players in the primary production industry (kiwifruit), a significant threat to biosecurity within New Zealand or are considered genomically complex *i.e.* are abnormally large, have higher ploidy levels (blueberry), are highly repetitive or heterozygous. These species have been sequenced using a variety of NGS platforms, namely HiSeq/NovaSeq, ONT MinIon and Promethion, PacBio, Chromium 10X and Hi-C sequencing. The utilisation and hybridization of data from these technologies has already resulted in the generation of high quality genomes for invasive wasp species, Polynesian rat, manuka, rewarewa, 15 stick insect species, blueberry, hihi, robin and 125+ kakapo genomes to name but a few. A hybrid assembly pipeline was constructed and utilised to assemble a reference genome for the *Phasmatodea* clade, *Clitarchus hookeri* as well the low coverage assembly and phylogenetic analyses of genomes from a further 12 other stick insect species found in geographically distinct regions across Aotearoa to investigate the genomic basis of thermal tolerance and parthenogenesis. The integration of Matauranga Māori into all Genomics Aotearoa assembly and analyses pipelines as well as into the development of a distinct data repository for the genomes of taonga species will also be highlighted.

Novel insights into evolutionary genetics from emerging technologies

SMBE-OR-189

A short tale of viral evolution told using long reads

T. Sasani [1,*], K. R. Cone [1], A. R. Quinlan [1], N. C. Elde [2]

[1]Department of Human Genetics, University of Utah, [2]Department of Human Genetics, Univeristy of Utah, Salt Lake City, United States

Abstract: Poxvirus adaptation can involve combinations of recombination-driven gene copy number variation and beneficial single nucleotide variation at the same loci. But how might these distinct mechanisms of genetic diversification simultaneously facilitate adaptation to host immune defenses? To address this question, we performed experimental evolution with vaccinia virus populations harboring a single nucleotide variant (SNV) in a gene actively undergoing copy number amplification. Using long se-quencing reads from the Oxford Nanopore Technologies platform, we phased SNVs within large tan-dem gene arrays for the first time. Our analysis uncovered a mechanism of adaptive SNV homogeni-zation reminiscent of gene conversion, which is actively driven by selection. We reveal a new mecha-nism for the fluid gain of beneficial mutations in genomic regions undergoing active recombination in viruses and illustrate the value of long read sequencing technologies for investigating complex ge-nome dynamics in diverse biological systems.

Novel insights into evolutionary genetics from emerging technologies
SMBE-OR-187
Copy number variant immune genes in wild and lab zebrafish
J. Suurväli [1,*], M. Leptin [1], T. Wiehe [1]
[1]Institute for Genetics, University of Cologne, Cologne, Germany

Abstract: The nucleotide-binding domain and leucine-rich repeat containing proteins (NLRs) are a family of immune receptors crucial to key aspects of the immune response. Members of this family act as sensors of infection and cellular stress, as core components of the inflammasome complex, and as transcriptional regulators for antigen-presenting genes. While mammals and birds have only a handful of NLRs, teleost fish often have hundreds of near identical copies. Here we have used a long-read approach to sequence NLRs from wild and lab populations of the zebrafish, a model animal with its native habitat in India, Nepal and Bangladesh. The zebrafish genome contains ~400 NLR genes, many of which reside on the long arm of chromosome 4. In order to study the copy number variation of the NLRs we have adapted a protocol combining target enrichment with PacBio technology. Our custom-designed RNA baits for the NLRs and for other components of the zebrafish immune repertoire successfully capture their targets; with the PacBio technology, every copy of the repetitive exons is sequenced with the length and coverage necessary to distinguish paralogs from each other.
The dataset generated with this approach has allowed us to document high turnover rates for the zebrafish NLRs, with different copies present in different populations and even different individuals. Furthermore, we demonstrate a loss of many copies in the reference strain, an event that must have happened in the ~20 years since sequencing of this animal first began.

Origins, evolution and function of novel genes

SMBE-OR-201

De novo emergence of protein coding genes from dark genomic matter: genetic mechanisms, biophysical constraints and evolutionary dynamics.

E. Bornberg-Bauer [1,*]

[1]The Wsetfalian University of Muenster, Muenster, Germany

Abstract: Proteins are the workhorses of the cell and, over billions of years, they have evolved an amazing plethora of extremely diverse and versatile structures with equally diverse functions, both echoing the evolution of all forms of life. While many proteins result from gene duplication and subsequent divergence, newly sequenced genomes were repeatedly found to contain some 10%-30% "orphan" ORFs, which are often expressed and translated but lack a homolog in closely related outgroups. Unexpectedly, many such novel genes have emerged "de-novo", i.e. from previously non-coding DNA such as introns or intergenic regions. This is at at odds with basic biophysical principles which would stipulate that proteins walk a fine line between stability, foldability and avoiding aggregation as a prerequesite for functionality they can be selected for. Failing to do so would render them non-functional or even toxic and thus a burden to the cell such that they should be rapidly removed from the gene pool. We analysed genomic origins and structural changes of de novo proteins using high-quality transcriptomes, mapped on genomes and supported by ribosome-binding data in insects and in mmammals. First, we find that recently split lineages undergo accelerated genomic reorganisation, including the rapid gain of thousands of transcripts containing ORFs. However, the vast majority of these transcripts is rapidly lost again unless they survive a species boundary, in particular in species under strong selection pressure such as social insects. Second, many novel transcripts bind to ribosomes, making their exposure to selection very likely. Third, most de novo transcripts and random sequences alike have biophysical properties such as disorder content which are similar to well established (i.e. evolutionary old) proteins. Finally, although initially rarer than de novo transcripts, proteins which have arisen by duplication are still more frequently observed over longer time scales, probably because they have a higher likelihood to assume important functions and thus be retained.

Origins, evolution and function of novel genes

SMBE-OR-203

Selection of novel peptides conferring antibiotic resistance from a randomized sequence pool

M. Knopp [1,*], J. Gudmundsdottir, T. Nilsson, F. König, O. Warsi, F. Rajer, P. Ädelroth, D. Andersson

[1]Medical Biochemistry and Microbiology, Uppsala, Sweden

Abstract: The origin of novel genes and proteins is a fundamental question in evolutionary biology. New genes can originate from different mechanisms including horizontal gene transfer, duplication-divergence and de novo from non-coding DNA sequences. Comparative genomics has generated strong evidence for de novo emergence of genes in various organisms but experimental demonstration of this process has been limited to localized randomization in pre-existing structural scaffolds. This is bypassing the basic requirement of de novo gene emergence, i.e. lack of an ancestral gene. We constructed highly diverse plasmid libraries encoding randomly generated open reading frames and expressed them in Escherichia coli to identify peptides that could confer a beneficial and selectable phenotype in vivo. Selections on antibiotic-containing agar plates resulted in the identification of three inserts that increased aminoglycoside resistance up to 48-fold. Combining genetic and functional analyses, we show that the peptides are highly hydrophobic and that they insert into the membrane, reduce membrane potential, decrease aminoglycoside uptake and thereby confer high-level resistance. This study demonstrates that randomized DNA sequences can encode peptides that confer selective benefits, and illustrates how expression of random sequences could spark the origination of new genes.

Origins, evolution and function of novel genes

SMBE-OR-204

Many but not all lineage-specific genes can be explained by limited sensitivity of homology detection

C. Weisman [1,*], S. R. Eddy [1], A. W. Murray [2]

[1]Molecular and Cellular Biology, Howard Hughes Medical Institute & Harvard University, [2]Molecular and Cellular Biology, Harvard University, Cambridge, United States

Abstract: Genes whose homologs are detected only in a specific lineage ("lineage-specific genes," or LSGs) are pervasive: essentially every lineage, from young (like *sensu strictu* yeasts) to old (like metazoa), has its own set. Their apparent lineage-specificity is often interpreted as evidence of novelty, under the assumption that we would only fail to find homologs outside of a lineage if the gene came into existence at its base. LSGs have thus been widely interpreted as "de novo" (emerged from noncoding DNA) or radically neofunctionalized genes; implicated in lineage-specific innovations like new cell types and animal behaviors; and used to place the origin of fundamental biological structures like mesoderm.

Here we consider the alternative explanation -- an evolutionary null hypothesis -- that LSGs can be explained as genes diverging under constant evolutionary pressure whose homologs, though present in other clades, have become undetectable by homology search. We develop a simple mathematical model that predicts where in the tree of life genes evolving under this model would be undetectable. We find that this null model is sufficient to explain the lineage-specificity of a large number of LSGs in fungi and insects. We also find examples of LSGs in both clades that strongly reject this model.

Our work shows that novelty is not required to explain many LSGs, cautioning against an assumption widespread in their interpretation and characterization. It also offers a simple way of finding LSGs that are poorly explained by a null evolutionary scenario, highlighting them as interesting candidates for further study.

Origins, evolution and function of novel genes

SMBE-OR-202

The de novo emergence of adaptive membrane proteins

A.-R. Carvunis [1,*]

[1]Department of Computational and Systems Biology, University of Pittsburgh Medical School, Pittsburgh, United States

Abstract: The de novo emergence of novel protein-coding genes is thought to be facilitated by pervasive translation of intergenic transcripts, which exposes a reservoir of variable polypeptides to natural selection. How could such intergenic translation events yield polypeptides with useful biochemical capacities? To gain insights into this question, we systematically characterized how de novo emerging coding sequences impact fitness in budding yeast. Overexpression of these sequences was enriched in beneficial effects, while their disruption was generally inconsequential for fitness. We found that beneficial emerging sequences have a strong tendency to encode putative transmembrane proteins, which appears to stem from a cryptic propensity for transmembrane signals throughout the intergenic regions of the genome. These findings suggest that novel genes with useful biochemical capacities, such as transmembrane domains, tend to evolve de novo within intergenic loci that already harbored a blueprint for these capacities.

Origins, evolution and function of novel genes
SMBE-OR-205
The origin and regulation of novel genes at the single-cell level
L. Zhao*, E. Witt, N. Svetec

Abstract: Many species exhibit patterns of rapid gene expression evolution in reproductive organs, especially in testis. We and other colleagues found that testis is also a hotspot for new gene origination, including newly duplicated genes and de novo genes - genes that originated from non-coding sequences. One hypothesis is that testis provides a permissive environment for novel gene expression because of the dramatic changes of chromatin states occurring during spermatogenesis. To test this hypothesis, we performed ATAC-seq on testis of multiple strains spanning three Drosophila species. We found that a set of novel genes showing fast expression turnover is likely to be associated with fast chromatin changes. To characterize the expression patterns of novel genes, we performed single-cell RNA sequencing on the testis. Both techniques revealed and identified de novo genes and duplicated showing strong spatial and temporal expression patterns at the cell type level. These results indicate that there is a dynamic change in the expressed and translated pathways along spermatogenesis. Altogether, our results provide insight into the origination and evolution of de novo genes and duplicated genes, their regulatory circuits and their contributions to the biology of the testis.

Origins, evolution and function of novel genes

SMBE-OR-206

Creating a context to study novel gene evolution.

N. Prabh [1 2,*], C. Rödelsperger [3]

[1]Evolutionary Genetics, Max Planck Institute for Evolutionary Biology, Plön, [2]Integrative Evolutionary Biology, Max Planck Institute for Developmental Biology, [3]Integrative Evolutionary Biology, Max Planck Institute for Developmental Biology, Tübingen, Germany

Abstract: The fraction of novel genes tends to be higher in phylogenetically isolated species such as *Pristionchus pacificus*. Till recently, *Caenorhabditis elegans* was the closest neighbor of *P. pacificus* with sequenced genome data, and these two species diverged approximately 75 mya. We found that more than one-third of the *P. pacificus* genes are orphans and another one-third are duplicates of *C. elegans*, genes. On one hand, such high numbers make it imperative that these novel genes are investigated, on the other hand, the lack of genomic sampling of neighboring species obviate the possibility of an in-depth analysis. To overcome this issue, we created a unique ladder-like phylogeny around *P. pacificus* by sequencing the genomes of seven *Pristionchus* species and two outgroup species. This facilitated age estimation of taxon-restricted genes and shed some light on their chromosomal localization and evolutionary trajectories. Further, we were also able to create a high-confidence candidate set which illustrated six different mechanisms of novel gene origins in *Pristionchus* nematodes. However, even at such a fine level of taxonomic-resolution, we were unable to trace the origin of a large fraction of novel genes. This inspired us to study novel gene evolution at the microevolutionary level and thus we sequenced the genomes of seven strains of *P. pacificus*, which allowed us to trace the homology of 99% of all genes either at the species level or at the strain level. The resulting phylogenomic framework has furthered our knowledge of gene and genome evolution of *Pristionchus* nematodes. In my presentation, I will introduce our most recent findings and share the immense benefits of creating a proper phylogenetic context to investigate the evolution of novel genes.

Phylogenomics under the multispecies coalescent

SMBE-OR-085

Multispecies coalescent model with and without introgression for analysis of genomic sequence data

T. Flouris [1], X. Jiao [1], B. Rannala [2], Z. Yang [3,*]

[1]Department of Genetics, University College London, London, United Kingdom, [2]Department of Ecology & Evolution, University of California at Davis, Davis, United States, [3]University College London, London, United Kingdom

Abstract: In this talk I will give an overview of the multispecies coalescent (MSC) model and our recent extension, the multispecies-coalescent-with-introgression (MSci) model. The model is implemented in the Bayesian Markov chain Monte Carlo program bpp, and naturally accommodates deep coalescent and cross-species introgression, two challenges for inferring species phylogenies using genomic sequence data from closely related species. The full likelihood implementation accounts for the weak phylogenetic information in sequence alignment at each locus due to high sequence similarity. Application of the MSci model to the genomic sequence data from the Anopheles gambiae species complex suggest large variation in the effective introgression intensity across the genome, apparently driven by differential selection against introgressed alleles.

Phylogenomics under the multispecies coalescent

SMBE-OR-087

Maximum Likelihood Implementation of an Isolation-with-Migration Model for Three Species

T. Zhu [1,*]

[1]Academy of Mathematics and Systems Science, Chinese Academy of Sciences, Beijing, China

Abstract: We develop a maximum likelihood (ML) method for estimating migration rates between species using genomic sequence data. A species tree is used to accommodate the phylogenetic relationships among three species, allowing for migration between the two sister species, while the third species is used as an out-group. AMarkov chain characterization of the genealogical process of coalescence and migration is used to integrate out the migration histories at each locus analytically, whereas Gaussian quadrature is used to integrate over the coalescent times on each genealogical tree numerically. This is an extension of our early implementation of the symmetrical isolation-with-migration model for three species to accommodate arbitrary loci with two or three sequences per locus and to allow asymmetrical migration rates. Our implementation can accommodate tens of thousands of loci, making it feasible to analyze genome-scale data sets to test for gene flow. We calculate the posterior probabilities of gene trees at individual loci to identify genomic regions that are likely to have been transferred between species due to gene flow. We conduct a simulation study to examine the statistical properties of the likelihood ratio test for gene flow between the two in-group species and of the ML estimates of model parameters such as the migration rate. Inclusion of data from a third out-group species is found to increase dramatically the power of the test and the precision of parameter estimation. We compiled and analyzed several genomic data sets from the *Drosophila* fruit flies. Our analyses suggest no migration from *D. melanogaster* to *D. simulans*, and a significant amount of gene flow from *D. simulans* to *D. melanogaster*, at the rate of ~0.02 migrant individuals per generation. We discuss the utility of the multispecies coalescent model for species tree estimation, accounting for incomplete lineage sorting and migration.

Phylogenomics under the multispecies coalescent

SMBE-OR-088

A new statistic for detecting and quantifying the presence of admixture in species trees from genomic data using coalescent theory.

G. Guerra [1,*], R. Nielsen [1 2]

[1]Statistics, [2]Integrative Biology, University of California, Berkeley, Berkeley, United States

Abstract: A common approach to species tree estimation is the two-step approach of estimating local coalescent trees from sequence data, and inferring the phylogeny using the local trees as input data. These methods developed in recent years estimate phylogenies and the corresponding parameters (divergence times and population sizes) accounting for the presence of incomplete lineage sorting (ILS), and more recently, in the presence of mutational variance in local tree reconstruction.

The process of admixture between separated populations presents a different challenge in species tree inference, most importantly in divergence time estimation. In the presence of an admixture event, current species tree methods will systematically underestimate divergence times, and population size estimation becomes uninformative. We present an extension to our species tree inference method, COAL_PHYRE, to detect the presence of admixture in species trees, as well as determine the location on the tree. We use a novel approach of testing for variation in the empirical covariance structure of pairwise coalescence times to detect deviations from those values inferred from the multi-species coalescent. Our method takes a three-step approach: inference of a species tree under the assumption of no admixture, detection of potential admixture events, and the inference of the timing and strength of detected admixture events

Phylogenomics under the multispecies coalescent

SMBE-OR-086

The coalescent is an essential tool for delineating the 'diversity' in Biodiversity

A. D. Yoder [1,*]

[1]Biology, Duke University, Durham, NC, United States

Abstract: The majority of earth's biodiversity can be found in parts of the world that are remote and face imminent threat from destructive human activities. Thus, it can be extremely challenging for conservation biologists and others to collect the biological samples necessary for accurate measurement of species diversity and other critical parameters of biodiversity composition. In parts of the world such as Madagascar where these challenges are especially acute, investigators may be limited to sampling from very few individuals at a given locality as representatives of that population. Thus, sample size for purposes of conservation and evolutionary genetic studies, can be extremely limiting for traditional population genetic studies such as measurements of effective population size, demographic history, gene flow, divergence times, and so on. It is here that coalescent methods are empowering the study of biodiversity, both regarding species delimitation and other fundamental aspects of species evolutionary history. My talk will illustrate the strength of coalescent methods for these purposes, using genomic data from the mouse lemurs of Madagascar as a case study.

Phylogenomics under the multispecies coalescent

SMBE-OR-089

Complex patterns of genealogical discordance among jaguar, lion and leopard (Felidae, Panthera): dissecting ILS and introgression signals using whole genome sequence data

S. H. D. Santos [1], H. V. Figueiro [1], W. J. Murphy [2], E. Eizirik [1,*]

[1]School of Sciences, PUCRS, Porto Alegre, Brazil, [2]Texas A&M University, College Station, United States

Abstract: Genealogical discordance is currently recognized to be an important challenge to the reconstruction of the Tree of Life. Recent approaches such as the multispecies coalescent (MSC) have improved our ability to incorporate it into phylogenetic inference, especially in cases induced by incomplete lineage sorting (ILS). In spite of these advances, the incorporation of other factors, such as post-speciation admixture or the effects of recombination rate differences and selection across the genome, is only beginning to be pursued. The complex interactions among these factors and ILS have seldom been investigated using empirical genome-wide data. An interesting system to investigate this problem is the big cat genus *Panthera*, comprising five extant species whose phylogeny has been notoriously difficult to resolve. Our previous studies using complete genomes of this group included a single individual per species, and documented extensive genealogical discordance, especially involving the trio jaguar-lion-leopard. Here we present two novel jaguar genomes, which we used to create a 10-genome dataset (3 jaguars, 2 lions, 2 leopards, 1 tiger, 1 snow leopard, 1 domestic cat outgroup). We reconstructed the ML topology for non-overlapping windows of different sizes (50 kb, 100 kb, 1 Mb and 5 Mb), and only kept those that supported one of the three possible topologies for the trio with bootstrap values >90%. We tested the genome-wide monophyly of each focal species, and found it to be highly supported. We investigated the frequency and spatial distribution of the three topologies, and found them to be very uneven: with all window sizes, the most frequent topology (76%–95%) united lion+leopard as sister-species (topology 1), followed by lion+jaguar (topology 2: 4%–8%) and leopard+jaguar (topology 3: 1%–6%). Topology 1 was strongly predominant in regions of high recombination, whereas topologies 2 and 3 were strongly enriched in low-recombination regions. We applied the MSC approach implemented in ASTRAL-II, which supported topology 1 as representing the species tree for each separate chromosome and genome-wide. We estimated divergence times between the two sister-species defined by each topology, as well as a 'normalized age' in which the depth of this node was divided by that of the preceding (trio) node, to further correct for recombination-rate effects on substitution rate. Absolute divergence times were younger for topologies 2 and 3 (relative to topology 1), but this pattern was reversed with the 'normalized age'. Introgression analyses using DFOIL indicated pervasive historical admixture among these species, regardless of the assumed species tree. For example, if topology 1 was assumed, 7% of the genome was inferred to derive from lion-jaguar hybridization, while if topology 2 or 3 were assumed, 37% of the genome was detected as introgressed between lion and leopard. Our results indicate that topology 2 most likely reflects the original speciation sequence (species tree), given its age, frequency and recombination profile. The jaguar+leopard topology did not differ significantly in age or recombination profile, and thus likely derives from ILS. These results imply that a large proportion of the genome has been overwritten by post-speciation admixture between lion and leopard, leading to a complex mosaic whose phylogenetic resolution requires integration of different data types, which should be pursued in future developments of MSC approaches.

Phylogenomics under the multispecies coalescent

SMBE-OR-090

Quantifying Hybridization and Finding Introgressed Loci Using Gene Tree Branch Lengths

M. Miyagi [1,*], N. B. Edelman [1], A. J. Blumberg [2], J. Wakeley [1]

[1]Department of Organismic and Evolutionary Biology, Harvard University, Cambridge, [2]Department of Mathematics, The University of Texas at Austin, Austin, United States

Abstract: Gene tree and species tree discordance can either be indicative of reticulate evolution, such as introgression, or be a result of deep coalescence events (also known as incomplete lineage sorting). Differentiating between these is vital in determining if specific loci of interest are introgressed. In order to resolve these signals, we developed a framework based on the multispecies network coalescent to describe the distributions of gene tree internal branch lengths for unlinked loci in the presence or absence of introgression. Using simulations, we demonstrate that if our assumptions are met, this internal branch length distribution is sufficient for the diagnosis of the presence, direction, and timing of introgression. In addition, we show that the same framework can be used to compute the expected distributions of counts of SNPs along branches of the phylogeny in these scenarios. We implement this result in a procedure which can diagnose introgression without *a priori* knowledge of putative hybrid populations, as well as assign likelihoods of introgression on a locus-by-locus level. Finally, we use this to quantify reticulate evolution in the *Heliconius* phylogeny.

Repeats and mobile elements as drivers of innovations in protein coding genes

SMBE-OR-218

RNA and DNA editing of mobile elements by ADARS and APOBECs accelerate genome evolution.

E. Levanon [1,*]

[1]Bar Ilan University , Ramat Gan, Israel

Abstract: Genome evolution is commonly viewed as a gradual process that is driven by random mutations that accumulate over time. However, DNA- and RNA-editing enzymes have been identified that can accelerate evolution by actively modifying the genomically encoded information. The apolipoprotein B mRNA editing enzymes, catalytic polypeptide-like (APOBECs) are potent restriction factors that can inhibit retroelements by cytosine-to-uridine editing of retroelement DNA after reverse transcription. In some cases, a retroelement may successfully integrate into the genome despite being hypermutated. Such events introduce unique sequences into the genome and are thus a source of genomic innovation. adenosine deaminases that act on RNA (ADARs) catalyze adenosine-to-inosine editing in double-stranded RNA, commonly formed by oppositely oriented retroelements. The RNA editing confers plasticity to the transcriptome by generating many transcript variants from a single genomic locus. If the editing produces a beneficial variant, the genome may maintain the locus that produces the RNA-edited transcript for its novel function. Here, we discuss how these two powerful editing mechanisms, which both target inserted retroelements, facilitate expedited genome evolution.

Repeats and mobile elements as drivers of innovations in protein coding genes
SMBE-OR-216

Evolution of the Caenorhabditis germ line transcriptional network through transposable element co-option

F. N. Carelli [1,*], C. Cerrato [1], Y. Dong [1], A. Appert [1], J. Jänes [2], J. Ahringer [1]
[1]Gurdon Institute, University of Cambridge, [2]EBI, Cambridge, United Kingdom

Abstract: Transposable elements (TEs) are DNA sequences capable of inserting into new genomic locations. Although new TE integrations are usually neutral or deleterious for the host, TE amplification can disperse protein binding sites in the genome, and thus generate new regulatory elements. Distinct classes of TEs have been shown to be enriched at promoters or enhancers active in specific tissues or cell types. TE expansion has, therefore, the potential to drive the evolution of gene regulatory networks. To date, nonetheless, these observations have been mostly limited to correlation analyses or tested in cell lines.

Here we provide evidence that two classes of DNA transposons have been co-opted as new regulatory elements in *C. elegans*. We found that a pair of DNA motifs, strongly enriched in germline-specific promoters, are part of the inverted repeat sequences of CERP2 and CELE2, two inactive DNA transposons. Through reporter assays, we validated the activity of the two motifs, confirming their role in driving germline-specific regulatory elements *in vivo*. Comparative analyses revealed that CERP2-associated motifs are found across the whole *Caenorhabditis* clade but not in other nematodes, whereas the CELE2-expanded motifs are *C. elegans*-specific. By annotating regulatory elements in the sister species *C. briggsae*, we confirmed that *C. elegans*-specific repeat expansions led to the emergence of a number of germline promoters which could not be identified at the orthologous locations in *C. briggsae*. Finally, we collected initial evidence that two fast evolving chromatin-associated proteins, specific of the Rhabditomorpha, might be binding the motifs associated to CERP2/CELE2. Overall, our work strongly suggests that distinct waves of TE expansion have shaped the germline regulatory network in the *Caenorhabditis* genus.

Repeats and mobile elements as drivers of innovations in protein coding genes
SMBE-OR-219

Low Complexity Region associated variation in mRNA and protein abundance

Z. W. Dickson [1,*], G. B. Golding [1]
[1]Biology, McMaster University, Hamilton, Canada

Abstract: Intergenic regions of eukaryotic genomes are littered with volatile, highly repetitive sequences known as Low-Complexity Regions. These LCRs mutate rapidly though replication-slippage-induced copy number variation and elevated point mutation rates. With the exception of transposons, these non-coding LCRs are reasonably harmless, however their instability should make them intolerable within protein coding DNA. LCRs in amino acid sequences are often intrinsically disordered, and are capable of non-specific binding, all of this combined with high chance of mutation seems to violate our understanding of the tight linkage between stable, defined structures and function. Despite this LCRs are some of the most common motifs in eukaryotic proteins, ranging from 15% in yeast to some Plasmodium species with LCRs in 80% of their proteins. Given the prevalence of protein LCRs, the lack of defined structure and capacity for non-specific binding may indeed be features. Recent research has indicated functional roles for LCRs in regulatory networks as hub proteins, and stress responses via liquid-liquid droplet formation. Even considering these roles the presence of LCRs is still unlikely to be tolerated in highly expressed, strongly conserved proteins. We have integrated RNA-Seq data from the Genotype-Tissue Expression Project and mass spectrometry data from the Protein Abundance Database to investigate the relationship between LCRs and the abundance of the mRNA and proteins which encode and contain them. We have found that, in humans, LCRs are associated with lower than median abundance proteins ($p < 1E5$), however the mRNA which encode LCR positive proteins have higher than median abundance ($p < 1E6$). This contrast exists in both mice and humans, even when controlling for associated factors including decreased translation efficiency of LCR encoding mRNA, and more rapid degradation of LCR containing proteins. The opposing effects at the mRNA and protein level may indicate elevated gene expression as compensation for an LCR-induced reduction in steady state protein levels. We intend to further validate this by tracing the evolution of these proteins and their abundance, determining the temporal order of the sequence and expression shifts.

Repeats and mobile elements as drivers of innovations in protein coding genes

SMBE-OR-217A

Transposable elements as catalysts of cellular innovation

C. Feschotte [1,*]

[1]Cornell University, Ithaca, United States

Abstract: Most eukaryotic genomes are replete with sequences derived from selfish genetic elements, such as transposable elements and endogenous viruses. In this talk, I will present a snapshot of the myriad ways these elements have influenced, for better or worse, the biology of their hosts. I will summarize a body of work illustrating how regulatory and coding sequences prefabricated and dispersed by mobile elements have been repurposed repeatedly during evolution to foster the emergence of cellular innovations in physiology and development.

Repeats and mobile elements as drivers of innovations in protein coding genes

SMBE-OR-220

Genome-wide comparative analysis of mammalian transposable elementsthat code for viral-like proteins

M. T. Ueda [1,*], K. Kryukov [1], S. Mitsuhashi [2], H. Mitsuhashi [3], T. Imanishi [1], S. Nakagawa [1 4]

[1]Department of Molecular Life Science, Tokai University School of Medicine, Isehara, [2]Department of Human Genetics, Yokohama City University GraduateSchool of Medicine, Yokohama, [3]Department of Applied Biochemistry, School of Engineering, Tokai University, [4]Micro/NanoTechnology Center, Tokai University, Hiratsuka, Japan

Abstract: Transposable element (TE) groups of endogenous retroviruses (ERVs) and long interspersed nuclear elements (LINEs) code viral-like proteins. Most of them lost their open reading frames (ORFs), but some obtained function in the host species during evolution. However, it is still unclear what percentages of viral-like proteins derived from ERVs and LINEs, which are called endogenous viral element-derived ORFs (EVE ORFs) in this study, can be candidates for new functional genes. To understand characteristics and evolution of EVE ORFs, we examined over 600 thousand EVE ORFs in 19 mammalian genomes using public transcriptomic and epigenomic data. We found that small but certain proportions of EVE ORFs inserted relatively old lineage of host species indicating that evolutionary selection pressure may operate them. In the human and mouse genomes, fractions of EVE ORFs were significantly depleted within or downstream of transcription start sites and DNaseI hyper-sensitive sites compared to those without ORFs. On the other hand, more than thousands of EVE ORFs were detected in RNA-sequencing data in these species. These results suggest that although there is a selection to suppress the expression of EVE ORF as a whole, there are still possibilities that many unknown EVE ORFs are expressed as proteins, most of which were confirmed to be lineage specific. Furthermore, in the well-annotated human and mouse genomes, 37 multi-exon genes were identified to possess exons derived from EVEs; some of which were newly found to contain viral-like motives. These results showed that EVE ORFs have been raw materials for genes in various mammalian species during evolution.

Repeats and mobile elements as drivers of innovations in protein coding genes

SMBE-OR-221

Unique molecular evolution in teleost hatching enzyme genes – is this a new role of retrotransposon for evolution? -

T. Nagasawa*, M. Kawaguchi, T. Yano, S. Isoyama, S. Yasumasu, M. Okabe

Abstract: During the eukaryotic evolution, position(s) and number of inserted intron(s) among homologous genes are highly conserved. Retrocopy (gene duplication via spliced-mRNA) remove all-introns, while retrocopied genes lose original expression pattern because of losing their original promoter (mostly become pseudogenes). However, it is known that teleostean hatching enzyme genes (*HEs*) had lost all- (8-)introns with maintaining the expression pattern at several independent evolutionary lineages. In this study, to solve the contradiction, we compared genome sequences around *HE* genes among some teleostean species in detail. The syntenic analysis of *HE* genes revealed that (1) *HE* had translocated to different genomic loci by gene duplication, and (2) the evolutionary timing of the translocation and intron-loss were well correlated. Although the synteny of homologous genes were highly conserved in general (especially, during short evolutionary period), it was expected that the translocation had occurred several independent evolutionary lineages (at least 7-times). Surprisingly, promoter assay revealed that (3) promoter sequences of *HE*s were well conserved even after such many times of translocation. These results suggest *HE*s had retrocopied accompanying with their promoter sequences, i.e. teleostean *HE*s had experienced gene duplication by retrocopy without pseudogenigation. The gene duplication is major driving forces for molecular evolution (Ohno 1973). It seems that molecular evolution of *HE*s changes 'the theory that retrocopy, retrotransposon, randomly promotes molecular evolution depending on neighbor sequences.

The application of evolutionary approaches to the study of microbial communities and microbiomes
SMBE-OR-148
Phylogenetic analysis of genomes from metagenomes: novel species and strain-level population genetics
N. Segata [1,*]
[1]University of Trento, Povo, Italy

Abstract: Shotgun metagenomics has uncovered a substantial amount of diversity in the human microbiome, but a large fraction of the sequences in a metagenome remains uncharacterized. In my talk I will show how a phylogenetic framework applied on a large assembly-based metagenomic profiling led to the discovery of hundreds of previously unknown species and strain-level variants of known species. I will describe the catalog of >150,00 microbial genomes that we reconstructed from almost 10,000 metagenomes spanning body sites, ages, countries, lifestyles, and primate hosts from humans to apes. Phylogenetic analyses found that these genomes could be recapitulated into ~6,000 species-level genome bins (SGBs), >75% without genomes in public repositories. Analysis of phylogenies at multiple resolution levels showed that some novel candidate species are very prevalent, that a large amount of phylum-level diversity was uncovered, and that several species and subspecies and are enriched in non-Westernized populations. The new catalog permits deeper microbiome analyses and increase the average mappability of metagenomic reads from 67.76% to 87.51% in the gut (median 94.26%) and 65.14% to 82.34% in the mouth, and the newly recovered diversity range from new bacterial families to novel subspecies of already studied microbes. I will also discuss the population genomics of specific species and their evolutionary patterns.

The application of evolutionary approaches to the study of microbial communities and microbiomes
SMBE-OR-150

Horizontally transmitted symbiont populations from individual deep-sea mussels are genetically isolated

D. Romero Picazo [1], T. Dagan [1], R. Ansorge [2], J. M. Petersen [3], N. Dubilier [2], A. Kupczok [1,*]
[1]Institute for General Microbiology, Kiel University, Kiel, [2]Max Planck Institute for Marine Microbiology, Bremen, Germany, [3]Division of Microbiology and Ecosystem Science, University of Vienna, Vienna, Austria

Abstract: Eukaryotes are habitats for bacterial organisms where the host colonization and dispersal among individual hosts have consequences for the bacterial ecology and evolution. Vertical symbiont transmission leads to geographic isolation of the microbial population and consequently to genetic isolation of microbiotas from individual hosts. In contrast, the extent of geographic and genetic isolation of horizontally transmitted microbiomes is poorly characterized. The deep-sea mussel *Bathymodiolus brooksi* harbors two chemosynthetic bacterial species in its gill tissue that are horizontally acquired. Here we use high-resolution metagenomics to quantify the degree of genetic isolation among individual mussel microbiomes. The reconstruction of core genome-wide strain sequences disclosed multiple strains that group into two and four clades for both symbionts, respectively. We found variation in nucleotide diversity and strain composition between mussels and this variation is related with mussel lifespan for one symbiont. For both symbionts, high estimates of the fixation index F_{ST} and of β-diversity on the strain level revealed a high degree of genetic isolation among individual hosts.
Our results suggest that the uptake of environmental bacteria is a restricted process in *B. brooksi*, where self-infection of the gill tissue results in serial founder effects during symbiont evolution. We conclude that bacterial colonization dynamics over the host life-cycle is an important determinant of population structure and genome evolution of horizontally transmitted symbionts.

The application of evolutionary approaches to the study of microbial communities and microbiomes
SMBE-OR-151
Large scale reconstruction of over a thousand Wolbachia genomes sheds light on its co-evolution
O. Rota Stabelli*

Abstract:
Large scale reconstruction of over a thousand Wolbachia genomes sheds light on its co-evolution.

Matthias Scholz, Davide Albanese, Kieran Tuohy, Claudio Donati, Nicola Segata, Omar Rota-Stabelli

Wolbachia is an iconic example of a successful intracellular bacterium. Despite its importance as a manipulator of invertebrate biology, its evolutionary dynamics and population biology have been poorly studied from a genomic viewpoint, mainly because of a paucity of available genomes. To expand the number of Wolbachia genomes across host phylogenies, we screened over 30,000 publicly available shotgun sequencing samples (>70TB of data) from more than 500 arthropods and nematodes species using metagenomic and phylogenomic approaches. By assembling over a thousand of high quality genomes, we provide the largest collection of Wolbachia genomes to date and a substantial increase in host representation. Our Wolbachia phylogenies based on both core-genome alignment and gene content provide a robust reference for future studies, support new strains in model organisms like Drosophila and economically relevant pest species, and reveal novel affinities including recent horizontal transfers amongst distantly related hosts. We found various instances of gene function gains and losses in different super-groups, particularly in Cytoplasmic Incompatibility inducing strains. Our intra-specific Wolbachia-host cophylogenies indicate striking heterogeneity of coevolutionary dynamics and show that horizontal transfer is widespread not only at the intraspecific, but also population interspecific level. Our work reveals new patterns of Wolbachia evolution and indicates that our strain-level metagenomic analysis of host sequencing projects is an effective method to recover endosymbiont genomes for downstream analyses.

The application of evolutionary approaches to the study of microbial communities and microbiomes
SMBE-OR-149
Co-evolution within a four-species bacterial community
S. Mitri [1,*]
[1]University of Lausanne, Lausanne, Switzerland

Abstract: Evolutionary dynamics are little understood within large microbial communities. Disentangling how microbial species co-evolve over time and alter their interactions with one another needs to first be carried out on simpler ecosystems. We have focused on a synthetic bacterial community consisting of just four species involved in bioremediation. Using a well-defined medium, the four species grow in a closed system, in which we can closely follow bacterial abundances, the interactions between the four species and their genetic changes, during a long-term evolutionary experiment. For comparison with the four-species community, our experiment included two mono-culture conditions and one community of three that were all passaged in parallel. Prior to the beginning of this evolutionary experiment, we assessed that the species have positive growth effects on one another. We then ask: what will happen to these positive interactions over time? And what are the key genetic changes that will occur as bacteria adapt to the environment and to the other species? Over approximately 500 generations, we find that species abundances fluctuate following patterns that differ depending on community composition. Out of five replicate communities containing all four species, three resulted in stable co-existence. Preliminary results from metagenomic sequencing of the populations also suggest that some mutations are consistent between replicates but differ depending on which species were co-cultured. We are yet to assess how the interactions have changed and to complete our genomic analysis but expect these results to be available by the time of the conference.

The application of evolutionary approaches to the study of microbial communities and microbiomes
SMBE-OR-152

Revealing virus diversity in urban wild birds with neurological symptoms using meta-transcriptomics

W.-S. Chang [1,*], J.-S. Eden [2], K. Rose [3], M. Shi [1], E. Holmes [1]
[1]School of Life and Environmental science, the University of Sydney, Charles Perkins Centre, [2]Westmead Institute for Medical Research, Centre for Virus Research, The University of Sydney, [3]Taronga Conservation Society Australia, Australian Registry of Wildlife Health, Sydney, Australia

Abstract: **Background and objective:** Wild birds are major natural reservoirs and potential dispersers of a variety of infectious diseases. It is therefore of considerable importance to understand their viral diversity for both conservation and to determine the risks for spill-overs to humans and other wildlife. We investigated the potential infectious causes in native avian species in Australia presenting with neurological symptoms and suspected underlying virus pathogens. **Materials and Methods:** RNA from diseased birds were extracted and pooled based on tissue types and clinical manifestation for library preparation and sequencing. By using a bulk meta-transcriptomic approach (i.e. unbiased RNA-sequencing), we identified a number of novel viruses from the families Astroviridae, Picornaviridae, Polyomaviridae,Paramyxoviridae, Parvoviridae, Flaviviridae, and Circoviridae in common urban wild birds including magpies, magpie-park, pied currawongs, ravens, and rainbow lorikeets. The underlying aetiology of each case was then confirmed by PCR assays targeting the novel pathogens identified. **Results and conclusions:** Our results revealed a number of candidate viral pathogens possibly contributing to neurological pathology in black and white bird diseases and clenched claw syndrome in Australia wild birds. Importantly, the existence of a diverse virome in urban avian species highlights the importance of elucidating the ecology of wildlife pathogens in urban environments, which will become increasingly important for managing disease risks and surveillance indicators to wildlife and humans. More broadly, our work shows how meta-transcriptomics has revolutionized pathogen discovery in wildlife diseases.

The application of evolutionary approaches to the study of microbial communities and microbiomes

SMBE-OR-153

Gut microbiome composition is predictive of gastrointestinal parasites in Cameroonians

M. A. Rubel [1,*], A. A. Abbas [1], L. J. Taylor [1], A. Ranciaro [1], E. Mbunwe [1], A. J. Connell [1], C. Tanes [2], K. Bittinger [2], C. Fokunang [3], A. Njamnshi [3], F. D. Bushman [1], S. Tishkoff [1]

[1]University of Pennsylvania, [2]Children's Hospital of Philadelphia, Philadelphia, United States, [3]University of Yaounde I, Yaounde, Cameroon

Abstract: The World Health Organization has estimated that 24% of the world's population is infected with gastrointestinal parasites, and this disease burden predominantly affects developing countries. African populations provide a unique opportunity to interrogate host-microbe co-evolution and adaptive phenotypes due in part to their antiquity and phenotypic diversity. We investigated parasitemia in the context of the full microbiome for Cameroonians (N=575) by conducting marker gene sequencing, shotgun metagenomic analysis, quantification of parasite load, and enumeration of immune parameters. Populations included hunter-gatherers, pastoralists, and agropastoralists, and these results were compared to urban U.S. subjects. Microbiota and frequencies of soil-transmitted helminths varied among Cameroonians. Hunter-gatherers were the most parasitized, while agropastoralists and pastoralists showed lower parasite frequencies. Increases in parasitemia correlated with increased gut microbial diversity in Cameroonians. *Ascaris lumbricoides, Necator americanus, Trichuris trichiura*, and *Strongyloides stercoralis* helminths ("ANTS" group), significantly cooccurred, and can be predicted with ~74% accuracy based on gut microbiome composition. ANTS parasites, in turn, were associated with multiple proinflammatory cytokines, suggestive of immune responses against concurrent, different infections. Variation in the TH2 cytokine IL5 was the most predictive of microbiome composition among tested cytokines, and indicated similar taxonomic shifts as were predicted with the ANTS positivity comparisons. Undiagnosed HIV infection was detected in 27 subjects--this was not correlated with parasitemia or alterations in the microbiota, though it was weakly correlated with elevated TNFα. These data are suggestive of transkingdom interactions resulting in distinctive host immune responses, and these play out differently among the diverse populations studied.

The Causes of parallel molecular evolution

SMBE-OR-033

Predicting mutational routes to new adaptive phenotypes

P. B. B. Rainey [1,*]

[1]Max-Planck Institute for Evolutionary Biology , Plon, Germany

Abstract: The repeatability of molecular evolution in both laboratory and wild populations gives reason to think that evolution follows certain rules. Work with experimental bacterial populations shows that the genotype-to-phenotype map (defined by the functional and regulatory interactions among genes and gene products) provides basis for predicting evolutionary change. This is primarily through effects on the capacity of loci to translate mutation into phenotypic variation. I will describe recent work that attempts to ascertain the limits on bottom-up evolutionary forecasting. It begins with mathematical descriptions of three regulatory pathways that underpin evolution of the "wrinkly spreader" (WS) phenotype. The models predict the likelihood that evolution follows each of the pathways and the expected mutational targets. I will then describe experimental studies that test key predictions. I will conclude on a mix of optimism and caution: optimism from the fact that evolutionary change can be forecast, but caution because of current inability to a priori predict locus-specific mutational biases and fitness effects.

The Causes of parallel molecular evolution

SMBE-OR-035

Parallel evolution of fungicide resistance in lab and field

N. Hawkins [1,*], B. Fraaije [1]

[1]Biological Chemistry and Crop Protection, Rothamsted Research, Harpenden, United Kingdom

Abstract: When a new pesticide or antimicrobial drug is introduced, the predictability of evolution has practical implications. Knowing in advance what mutations are likely to emerge can inform resistance management strategies, and enable pre-emptive design of molecular diagnostics. Resistance mutations can be predicted by experimental evolution in the laboratory, and by mutations found in the highest-risk pathogens later emerging in slower-evolving species. However, the mutational pathways to fungicide resistance have proven more repeatable for some fungicides than others.

We conducted a meta-analysis comparing resistance mutations that were identified in laboratory mutagenesis studies with those that evolved in the field for the MBCs, an early group of single-site-inhibitor fungicides. Whilst pathogens in the lab and field are all under strong directional selection for resistance, only a small subset of lab mutations were subsequently found in field isolates, but the same few field mutations were found across multiple pathogen species, suggesting common functional constraints and fitness penalties under field conditions. Experimental evolution of resistance to the newer SDHI fungicides has provided further evidence of fitness penalties in some lab-evolved mutations, with the most resistant mutants selected against at lower doses.

For DMI fungicides, the evolutionary pathways are more complex, with some fungi accumulating multiple mutations in the target-site-encoding gene. Some mutations have evolved many times, in different plant pathogens, clinical pathogens and even non-target saprophytic fungi, whereas others are limited to one or a few related species. We are using functional genetic approaches, including homologous recombination with site-directed mutagenesis, to quantify the individual effects and epistatic interactions of the mutations, both on fungicide resistance and on fitness in the absence of fungicide selection. This has revealed functional constraints limiting some mutations to a certain genetic background, making their repeated parallel evolution less likely.

The Causes of parallel molecular evolution

SMBE-OR-037

Transition bias influences the evolution of antibiotic resistance in Mycobacterium tuberculosis

J. L. Payne [1,*], F. Menardo [2], A. Trauner [2], S. Borrell [2], S. M. Gygli [2], C. Loiseau [2], S. Gagneux [2], A. Hall [1]

[1]ETH Zurich, Zurich, [2]SwissTPH, Basel, Switzerland

Abstract: Transition bias, an overabundance of transitions relative to transversions, has been widely reported among studies of the rates and spectra of spontaneous mutations. However, demonstrating the role of transition bias in adaptive evolution remains challenging, and in particular it is unclear whether such biases direct the evolution of bacterial pathogens adapting to treatment. We addressed this challenge by analyzing adaptive antibiotic-resistance mutations in the major human pathogen *Mycobacterium tuberculosis*. We found strong evidence for transition bias in two independently curated datasets comprising 152 and 208 antibiotic resistance mutations. This was true at the level of mutational paths (distinct, adaptive DNA sequence changes) and events (individual instances of the adaptive DNA sequence changes), and across different genes and gene promoters conferring resistance to a diversity of antibiotics. It was also true for mutations that do not code for amino acid changes (in gene promoters and the 16S ribosomal RNA gene, *rrs*), and for mutations that are synonymous to each other and are therefore likely to have similar fitness effects, suggesting that transition bias can be caused by a bias in mutation supply. These results point to a central role for transition bias in determining which mutations drive adaptive antibiotic resistance evolution in a key pathogen.

The Causes of parallel molecular evolution

SMBE-OR-034

Population size and the repeatability of antibiotic resistance evolution

J. Krug [1,*]

[1]Institute for Biological Physics, University of Cologne, Koeln, Germany

Abstract: The likelihood of parallel evolution in asexual populations is governed by a complex interplay of factors, which include the distribution of fitness effects, mutational bias and population size. The talk will review the underlying theoretical concepts in the context of microbial evolution experiments using the bacterium Escherichia coli challenged by the antibiotic cefotaxime. Particular emphasis will be placed on how clonal interference mediates the tradeoff between mutation and selection, leading to the dominance of distinct classes of resistance mutations in populations of different size.

The Causes of parallel molecular evolution

SMBE-OR-038

Butterfly mimicry evolved via parallel evolution of ancient, pleiotropic enhancers

J. J. Lewis [1,*], C. G. Danko [1], R. D. Reed [1]

[1]Cornell University, Ithaca, United States

Abstract: Evolution of complex phenotypes is frequently driven by a few genes of large effect, yet the genetic mechanisms and evolutionary processes underlying this phenomenon are poorly understood. This is particularly true for parallel evolution of complex traits, for which there are currently very few case studies. Color pattern mimicry in *Heliconius* butterflies provides a key example of convergent trait evolution via selection on a few large effect genes. To understand how a single gene can drive the evolution of complex new traits in multiple species, we used a combination of ATAC-seq, ChIP-seq, Hi-C, and CRISPR to functionally characterized five enhancers of the color pattern gene *optix* in *Heliconius* butterflies. Contrary to current models of rapid and modular enhancer evolution, we found that enhancers associated with wing pattern evolution are largely ancestral, pleiotropic, functionally interdependent, and have introgressed between populations. Remarkably, many of the same enhancers are also associated with color pattern adaptation in distantly related co-mimics, suggesting that parallel co-evolution of regulatory elements facilitated color pattern mimicry. Our results provide a counterpoint to prevailing expectations of modular regulatory evolution and raise the possibility that fragile, pleiotropic enhancer networks may be important drivers of parallel morphological evolution.

The Causes of parallel molecular evolution

SMBE-OR-36

Detecting convergent amino acid evolution in the presence of mutational biases

C. Rey [1], V. Lanore [2,*], P. Veber [2], L. Guéguen [3], N. Lartillot [2], M. Sémon [1], B. Boussau [2]

[1]ENS de Lyon, Lyon, [2]CNRS, [3]Université Claude Bernard Lyon 1, Villeurbanne, France

Abstract: The evolutionary genomics community has taken an interest in locating so-called “convergent sites”, that is, amino acid positions in the genome linked with convergent phenotypic changes. We have recently conducted a review where we evaluated the capacity of 7 methods from the literature to detect convergent sites in simulated alignments [1]. Our simulations used a mutation-selection model where the convergent phenotype is associated with a change of the direction of selection—in the form of a different amino acid fitness profile—and optionally with a change of the effective population size. These simulations did not take into account other phenomena—such as GC-biased gene conversion (gBGC) or CpG hypermutation—which could hamper convergence detection, either by hiding the signal of legitimate convergent sites or by producing non-convergent parallel substitutions leading to false positives. Some of the most advanced detection methods, based on codon models that distinguish mutations at the DNA level from selection at the amino acid level, may be less sensitive to those biases, but this has never been evaluated on simulations.

Therefore, we have developed more complex simulators which allow us to take into account mutational biases such as gBGC and CpG hypermutation. This extension of our method evaluation pipeline allows us to assess the capacity of such biases to produce convergent-looking substitutions, as detected by existing convergence detection methods. This is an important question to address in order to guide future method developments and to assess our ability to detect sites underlying convergent phenotypes. I will present results on simulated data and will look back on real data results with an updated knowledge of the methods’ weaknesses.

[1] Carine Rey, Vincent Lanore, Philippe Veber, Laurent Gueguen, Nicolas Lartillot, Marie Semon, and Bastien Boussau. "Detecting convergent adaptive amino acid evolution." *Philosophical Transactions of the Royal Society B.* (2019). 10.1098/rstb.2018-0234

The Coalescent in the Era of Population-Scale Genomics
SMBE-OR-117
HYBRID SIMULATIONS SOLVE BIASES IN THE COALESCENT
J. Kelleher*

Are you a member of SMBE?: No
Poster Submission: Coalescent simulations are an extremely useful and powerful part of the population genetics toolkit. However, such simulations are predicated on a number of simplifying assumptions, in particular, that the region of genome being simulated is short and the sample size is small relative to the effective population size. The effect of these assumptions must be reevaluated in the light of exponentially growing empirical datasets and recent computational advances enabling simulations of millions of whole genomes. We show that coalescent simulations of long regions of the genome exhibit large biases in identity-by-descent, long-range linkage disequilibrium, and ancestry patterns, particularly when sample size is large. The solution to this problem is to simulate the recent past with a model that explicitly takes diploid inheritance patterns into account; for the ancient past, the coalescent is an excellent approximation that can be simulated very efficiently. We discuss an extension to msprime that allows simulating a discrete time Wright-Fisher model backwards in time, which provides much more realistic patterns of inheritance in the recent past. More generally, we show that arbitrarily complex patterns of selection and demography can be simulated for the recent past using SLiM, with the ancient past simulated under the coalescent using msprime. This hybrid approach combines the strengths of both models, and is enabled by the tree sequence data structure and tskit software library.

The Coalescent in the Era of Population-Scale Genomics
SMBE-OR-112

Ancestral Haplotype Reconstruction in Endogamous Populations using Identity-by-Descent

K. Finke [1], M. Kourakos [1], R. Kember [2], M. Bucan [2], S. Mathieson [3,*]
[1]Computer Science, Swarthmore College, Swarthmore, [2]Genetics, University of Pennsylvania, Philadelphia, [3]Computer Science, Haverford College, Haverford, United States

Abstract: As more and more present-day humans are sequenced, close relationships between individuals will need to be modeled as part of ancestral inference. Here we develop a novel algorithm for ancestral haplotype reconstruction, given sequence data from living individuals and an extended pedigree. Unlike previous approaches, our method does not assume a standard pedigree structure and is able to accommodate hundreds of individuals. We apply our method to an Old-Order Amish population from Lancaster, Pennsylvania, including 784 individuals from the past 10 generations, 394 with whole genome sequence data. Nearly one third of the individuals in the sequenced pedigree have some form of mood disorder, and our aim is to understand the genetic basis of this condition through tracking haplotype transmission over time. To accomplish this, we reconstruct the haplotypes of the unsequenced individuals, up through the founders who came to the United States in 1744.

Our method works by first identifying segments of Identity-by-Descent (IBD) shared between two or more sequenced individuals. Then, possible sources ("roots") for each IBD segment are identified. Sources are weighted probabilistically using the local recombination rate and the number of generations the IBD segment must have survived. Next, we cluster and stitch together IBD segments that have been rooted in a single individual to form putative haplotypes. During conflict resolution, we remove IBD segments that conflict with strong clusters and up-weight the probability they are rooted in other individuals. We repeat this process until all IBD segments have been rooted. We validate our method by first using coalescent simulations to create a founder population. Then we sample individuals from this group and run their sequences through our full pedigree structure.

Our method was developed for endogamous populations with complex pedigree structures, but can be used on any extensive pedigree with the recent generations sequenced or genotyped. The results allow us to compute polygenic risk scores of the reconstructed individuals - or, since we often know their disease status, to augment GWAS or pedigree mapping analysis. We are also in a position to study haplotype diversity across generations, including deviations from neutrality such as transmission distortion. This type of practical ancestral reconstruction will become more common and necessary to understand rare or complex heritable diseases in extended families.

The Coalescent in the Era of Population-Scale Genomics

SMBE-OR-113

Population histories of the United States revealed through fine-scale migration and haplotype analysis

C. L. Dai [1], M. M. Vazifeh [2], C.-H. Yeang [3], M. G. Vilar [4], M. J. Daly [5 6 7 8], C. Ratti [2], A. R. Martin [6 7 8,*]

[1]Department of Electrical Engineering and Computer Science, [2]Senseable City Lab, Massachusetts Institute of Technology, Cambridge, MA, United States, [3]Institute of Statistical Science, Academia Sinica, Nankang, Taipei, Taiwan, [4]Genographic Project, National Geographic Society, Washington D.C., United States, [5]Institute for Molecular Medicine Finland (FIMM), University of Helsinki, Helsinki, Finland, [6]Analytic and Translational Genetics Unit, Massachusetts General Hospital, Boston, MA, [7]Program in Medical and Population Genetics, [8]Stanley Center for Psychiatric Research, Broad Institute of MIT and Harvard, Cambridge, MA, United States

Abstract: The population of the United States is shaped by centuries of migration, isolation, growth, and admixture between populations of global origins. Here, we assemble a comprehensive view of recent population history by studying the ancestry and population structure of over 32,000 individuals in the US from the National Geographic Genographic Project using genetic, ancestral birth origin, and geographic data. We identify migration routes and barriers that reflect historical demographic events. We also uncover the spatial patterns of relatedness in subpopulations through the combination of haplotype clustering, ancestral birth origin analysis, and local ancestry inference. These patterns include substantial structure and heterogeneity in Hispanics/Latinos, isolation-by-distance in African Americans, elevated levels of relatedness and homozygosity in Asian immigrants, and fine-scale structure in European descents. Furthermore, quantification of familial birthplaces recapitulates historical immigration waves at high resolution. Taken together, our results provide detailed insights into the genetic structure and demographic history of the diverse US population.

The Coalescent in the Era of Population-Scale Genomics
SMBE-OR-116
Identity inference of genomic data using long-range familial searches
T. Shor [1,*], Y. Erlich [1], I. Pe'er [1], S. Carmi [1]
[1]MyHeritage, Or Yehuda, Israel

Abstract: Consumer genomics databases have reached the scale of millions of individuals. Recently, law enforcement authorities have exploited some of these databases to identify suspects via distant familial relatives. Using genomic data of 1.28 million individuals tested with consumer genomics, we investigated the power of this technique. We project that about 60% of the searches for individuals of European descent will result in a third-cousin or closer match, which theoretically allows their identification using demographic identifiers. Moreover, the technique could implicate nearly any U.S. individual of European descent in the near future. We demonstrate that the technique can also identify research participants of a public sequencing project. On the basis of these results, we propose a potential mitigation strategy and policy implications for human subject research.

The Coalescent in the Era of Population-Scale Genomics
SMBE-OR-114
A fast genome chopper to detect strong species decline.
E. Kerdoncuff [1,*], A. Lambert [1], G. Achaz [1]
[1]College de France, Paris, France

Abstract: Only 5% of described species have a conservation status. Methods used to assess conservation status cannot be generalized to all species. Using coalescent theory, we developed a new method to study demography based on the length of compatible blocks along the genome, i.e. blocks of nucleotides within which recombination events are not detectable. From whole-genome data of multiple individuals in a population, we can chop a chromosome into compatible blocks in seconds. Lengths of compatible blocks depend on the frequency of recombination events which is influenced by the ancestral history of the population. Using the distribution of block lengths, we can discriminate a constant population and a declining one. This method can infer a very recent decline of a population from DNA sequences. It could be a new tool to assess conservation status in a wide range of species.

The Coalescent in the Era of Population-Scale Genomics

SMBE-OR-115

Relatives from the Iron Age: Combining Ancient and Modern Genomes in the Genealogy

R. Hui [1,*], C. Scheib [2], S. Inskip [1], C. Cessford [3], J. Robb [4], T. Kivisild [5]

[1]McDonald Institute for Archaeological Research, University of Cambridge, Cambridge, United Kingdom, [2]Estonian Biocentre, University of Tartu, Tartu, Estonia, [3]Cambridge Archaeological Unit, [4]Depeartment of Archaeology, University of Cambridge, Cambridge, United Kingdom, [5]Department of Human Genetics, Katholieke Universiteit Leuven, Leuven, Belgium

Abstract: Parallel to the explosive growth of modern human genomes being sequenced from across the world, development in the field of ancient DNA also brings an increasing number of genomes from past populations. Analysis on ancient genomes can benefit from large modern reference panels to mitigate limitations in data quality; at the same time, the temporal dimension adds new excitement to the genetic history of populations and even individuals. Here we study the relatedness and population history from remains in an Iron Age settlement in Duxford, England. Large modern reference panel enabled us to accurately impute common variants even in extremely low-coverage genomes (~0.05x). We report pairs of close relatives, including a family trio, supported by identical-by-descent segments using the imputed genomes. However, difficulties in imputing rare variants greatly limits the resolution of demographic inferences in the recent past. We therefore also developed a “threading” algorithm that uses allele frequencies to add low-coverage historical genomes on top of a modern coalescent tree. Combined with Relate, a fast algorithm to estimate local genealogies across the genome for large numbers of samples, this method captures temporal changes in coalescent rates between historical genomes and modern populations. We applied the algorithm to newly-sequenced Duxford individuals and other individuals from the Neolithic and Anglo-Saxon period from nearby sites to show changes in ancestral makeup. Our results demonstrate that combining ancient and modern genomes in the same framework can help to clarify recent population history. At the same time, expanding the inference toolbox for extremely low-coverage genomes also helps to integrate aDNA ananlysis into bioarchaeological research to make full use of the historical and archaeological context.

The Ecological Genomics of Extremophile Eukaryotes
SMBE-OR-091
Plant Adaptation to Local Soil Composition
U. Kraemer*

Abstract: Among the sister species of Arabidopsis thaliana, the outcrossing, perennial, stoloniferous diploid Arabidopsis halleri has an exceptionally broad edaphic range that requires locally effective nutrient balancing. In addition, A. halleri groups among 700 known metal hyperaccumulator taxa, which in their natural habitat accumulate a heavy metal or metalloid at leaf concentrations >10-fold above critical toxicity thresholds of ordinary plants. We consider hyperaccumulation an analytically accessible trait exemplifying the co-option of minerals, i.e. their use for non-nutrition biological functions. A. halleri is a pseudo-metallophyte common at sites highly contaminated with the heavy metals zinc, cadmium and lead, but the majority of its populations exist on pristine uncontaminated soils. Species-wide hyperaccumulation of Zn (>3,000 – 54,000 g g-1 dry biomass) and the geographically confined hyperaccumulation of Cd (>100 – 3,600 g g-1 dry biomass) in A. halleri are known to occur on both metalliferous and non-metalliferous soils. We study A. halleri as a model species to uncover the genetic basis and evolutionary dynamics of two types of adaptation to local soil composition, nutrient balancing and mineral co-option. To this end, we have established a biodiversity resource of ca. 850 field-collected A. halleri individuals from Central and Eastern Europe alongside a documentation of their individual soil micro-environments of origin1. Based on genome-wide sequence data, we combine marker association, genome scan, genetic linkage mapping, targeted analysis of molecular functions and transcriptomic approaches. The status of results will be presented. Our work will provide insights into the repeated evolution of plant traits.

The Ecological Genomics of Extremophile Eukaryotes

SMBE-OR-093

Polar Porifera: Selection and Novelty in Antarctic Sponge Gene Complements

N. Kenny [1,*], A. Riesgo [1]

[1]Life Sciences, The Natural History Museum, London, London, United Kingdom

Abstract: Animals in the Antarctic have adapted to some of the most challenging conditions found anywhere on Earth. Temperatures ranging between 0 and -1.8°C and food supplies which fluctuate widely render their survival difficult. Nevertheless, species have found the means to thrive in such conditions. Sponges are keystone members of Antarctic ecosystems, but our knowledge of how they endure these temperatures is limited at best, especially at a molecular level.

We are investigating the mechanisms by which sponges have adapted to such extreme environments by contrasting congeneric species pairs adapted to vastly differing thermal environments. These aims are being addressed using transcriptomic and genomic sequences from 10 species in the order Halichondrida and genera *Mycale* and *Phorbas*. These are abundant in the Antarctic, Caribbean and Mediterranean. We are supplementing our "omic" work with targeted *in situ* and functional experiments, as well as RADseq of differing populations spanning thermal conditions found along the Western Antarctic Peninsula and beyond.

A number of tests (particularly in Hyphy/CODEML) have identified genes with multiple lines of evidence for positive selection. 10 genes with congruent results from all tests are of particular interest, including eukaryotic initiation factors and SPRY domain containing proteins, as well as a number of phylogenetically well-conserved "housekeeping" genes, which have been chosen for particular investigation. We have also analyzed differential gene expression and content. We can infer which genes are vital in cold conditions, and when adaptive molecular mechanisms have been used broadly, convergently, or in vastly varying ways across sponge phylogeny.

The Ecological Genomics of Extremophile Eukaryotes

SMBE-OR-094

The genome of a subterrestrial nematode reveals an evolutionary strategy for adaptation to heat

K. Walters-Conte*

Abstract: The nematode *Halicephalobus mephisto* was originally discovered inhabiting a deep terrestrial aquifer 1.3 km underground. *H. mephisto* can thrive under conditions of abiotic stress including heat and minimal oxygen, where it feeds on a community of both chemolithotrophic and heterotrophic prokaryotes in an unusual ecosystem isolated from the surface biosphere. Here we report the comprehensive genome and transcriptome of this organism, identifying a signature of adaptation: an expanded repertoire of 70 kilodalton heat-shock proteins (Hsp70) and AIG1 proteins. We find that positive selection has driven the expansion of Hsp70 genes, which are also transcriptionally induced upon growth under heat stress. We further show that AIG1 may have been acquired by horizontal gene transfer (HGT) from a rhizobial fungus. Over one-third of the genes of H. mephisto are novel, highlighting the divergence of this nematode from other sequenced organisms. This work sheds light on the genomic strategies of adaptation to heat in the first complete subterrestrial eukaryotic genome.

The Ecological Genomics of Extremophile Eukaryotes
SMBE-OR-092
PARALLEL ADAPTATION TO EXTREME ALTITUDE: RECONSTRUCTING THE GENETIC STEPS TOWARD ADAPTIVE PEAKS
A. Hancock [1,*]
[1]MAX-PLANCK-GESELLSCHAFT, Munich, Germany

Are you a member of SMBE?: No
Poster Submission: High altitude exerts strong selection pressures on plants through low temperatures, high UV radiation and low partial pressure of CO2. One of the most striking phenotypic responses to high altitude is a reduction in height. Co-variation between altitude and height was noted in classic studies of high-altitude ecotypes and clines. Subsequent work over the years has shown that reduced height is common across a broad range of alpine species and height reduction been associated with increased cold tolerance. We are studying Arabidopsis thaliana populations from altitudinal gradients in the East African Sky Islands, which exhibit remarkable variation in height. In these populations, we find that while a large number of loci contribute to height variation, the majority of the variation can be explained by a very small number of loci of large effect. By analyzing the spatial distribution of the functional alleles and their coalescence times, we provide insights into the detailed history of adaptation.

The Ecological Genomics of Extremophile Eukaryotes

SMBE-OR-095

Searching for patterns of Arctic adaptation in the genome of Draba nivalis (Brassicaceae)

M. Nowak [1,*], S. Birkeland [1], A. K. Brysting [2], L. Gustafsson [1], C. Brochmann [1]

[1]Natural History Museum, [2]Centre for Ecological and Evolutionary Synthesis, University of Oslo, Oslo, Norway

Abstract: The Arctic accounts for approximately 20% of Earth's land surface, and contains some of the most extreme terrestrial environments found anywhere on the planet. In Arctic habitats, plants and animals must endure months of darkness coupled with frigid temperatures during the winter, followed by a brief summer growing season characterized by unpredictable temperatures and nearly 24-hour sunlight. Previous work on the ecophysiology of Arctic plants show that these species are capable of tolerating high UV light exposure, drought, and extreme cold stress. While the physiological adaptations of Arctic plants have been relatively well documented, we still lack detailed information about the genetic basis of these adaptations. To gain an understanding of the genetic foundations of Arctic adaptation in plants, we produced a chromosome-scale genome assembly of *Draba nivalis* (Brassicaceae), a perennial Arctic specialist species with a circumpolar distribution. The results of several comparative genomic analyses suggest that the *D. nivalis* genome contains several significantly expanded gene families that are involved in specific types of stress tolerance consistent with adaptation to the harsh Arctic habitat, including temperature stress, desiccation stress, metabolism, and oxidative stress (related to high UV light exposure). Our results also show that a significant portion of the gene duplication events leading to the expansion of *D. nivalis* gene families putatively involved in Arctic adaptation appear to be driven by transposable element activity. As the first Arctic specialist species with a fully sequenced genome, we believe *D. nivalis* holds great promise for continued development as a model system for studying the genetics of Arctic plant adaptation.

The Ecological Genomics of Extremophile Eukaryotes

SMBE-OR-096

The canalizing role of gene body methylation in long-term stress adaptation

Y. Wang [1], A. Dai [1], Y. Chen [1], T. Tang [1,*]

[1]SUN YAT-SEN UNIVERSITY, GUANGZHOU, China

Abstract: Whether induced epigenetic changes contribute to long-term adaptation remains controversial.We jointly analyzed genome-wide dynamics of gene body methylation (gbM) in response to environmental stress and during mangrove evolutionary adaptation to coastal environments.Wefind accelerated gains of gbM in the mangrove lineages. A large fraction of the acquisition events underwent convergent evolution, strongly implying they are adaptive. Genes that convergently gained gbM during evolution are more likely to become methylated in response to salt stress in species where they are normally not marked. Stress-induced and evolutionarily convergent gains of gbM both correlate with reduction in expression variation, in accordance with robustness of mangrove transcriptomes under salt stress. Moreover, convergent gbM evolution is uncoupled with convergent sequence evolution. Our findings suggest that transgenerational inheritance of acquired gbM helps environmental canalization of gene expression, facilitating long-term stress adaptation of mangroves in the face of a severe reduction in genetic diversity.

The evolution of senescence: from theory to molecular data

SMBE-OR-021

Evolutionary repeatability of increased transposable element load in Drosophila populations selected for longevity

D. K. Fabian*, J. M. Thornton

Abstract: Transposable elements (TEs) often inflict numerous negative effects on health and fitness as they replicate by integrating into new regions of the host genome. Even though organisms utilize powerful mechanisms to demobilize TEs, transposons become increasingly derepressed during ageing. The rising TE activity causes genomic instability and was suggested to be involved in age-dependent neurodegenerative diseases and the determination of lifespan. It is therefore conceivable that long-lived individuals have improved TE silencing mechanisms and consequently fewer genomic insertions relative to their shorter-lived counterparts. Here, we test this hypothesis by quantifying genome-wide TE insertions in populations of *Drosophila melanogaster* selected for longevity through late-life reproduction for 50-170 generations from three independent studies. Surprisingly, we found that within studies, 54% to 74% of the ~120 present TE families were significantly more abundant in long-lived populations compared to non-selected controls, while only 11% to 26% showed a reduction. Moreover, among the increased TE families, 30 exhibited parallel expansions in selected populations of the different studies. The TE accumulation in long-lived flies therefore demonstrates a striking example of evolutionary repeatability with a to date unknown role in the evolution of lifespan. We hypothesize that selection for late-life reproduction increases the opportunity of TEs to replicate in the germline and be passed on to the next generation. Yet, longevity is retained despite an increased amount of TE insertions in selected flies, suggesting that they evolved tolerance to high TE loads or, alternatively, that the abundance of TE insertions has little impact on lifespan *per se*.

The evolution of senescence: from theory to molecular data

SMBE-OR-023

Relaxed selection limits lifespan by increasing mutation load in annual African killifishes

R. Cui [1,*], T. Medeiros [1], D. Willemsen [1], L. N. M. Iasi [1], G. E. Collier [2], M. Graef [1 3], M. Reichard [4], D. R. Valenzano [1 3]

[1]Max Planck Institute for Biology of Ageing, Cologne, Germany, [2]Biological Science, University of Tulsa, Tulsa, United States, [3]CECAD, University of Cologne, Cologne, Germany, [4]Institute of Vertebrate Biology, The Czech Academy of Sciences, Brno, Czech Republic

Abstract: To uncover the selective forces shaping life history trait evolution across species, we investigate the genomic basis underlying the repeated adaptations to seasonal habitat desiccation in African killifishes. By performing de novo genome assemblies for 4 species and resequencing additional samples covering all major lineages of the family Nothobranchiidae, we identify the genetic variants associated with positive and relaxed purifying selection in 45 killifish species and 231 wild individuals distributed throughout sub-Saharan Africa. In annual species, genetic drift led to the expansion of nuclear and mitochondrial genomes and caused the accumulation of deleterious genetic variants in key life-history modulating genes, such as *mtor*, *insr*, *ampk* and *foxo3 and polg.* Approximately one sixth of the protein coding genes are relaxed in the annual genus *Nothobranchius*. Using a yeast erythromycin assay, we functionally validated *polg* mutations in annual species that increased mitochondrial mutational rate. By re-analyzing published human genetic data, we find that aging-related genes are also preferentially affected by relaxed selection in human populations. Our findings provide empirical support for the evolution of limited lifespan by means of repeated weakening of natural selection due to population bottlenecks and to the weakened purifying selection in genes expressed in late life. We find that relaxation of purifying selection prominently shapes genomes and is a prime candidate force molding the evolution of lifespan and the distribution of genetic variants associated with late-onset diseases in different species.

The evolution of senescence: from theory to molecular data

SMBE-OR-022

The molecular basis of extraordinary longevity in bats

E. Teeling [1,*]

[1]School of Biology and Environmental Science, University College Dublin, Dublin, Ireland

Abstract: Of all mammals, bat possess the most unique and peculiar adaptations that render them as excellent models to investigate the mechanisms of extended longevity and potentially halted senescence. Indeed, they are the longest-lived mammals relative to their body size, with the oldest bat caught being >41 years old, living approx. 8 times longer than expected.Bats defy the 'rate-of-living' theories that propose a positive correlation between body size and longevity as they use twice the energy as other species of considerable size, but live far longer. The mechanisms that bats use to avoid the negative physiological effects of their heightened metabolism and deal with an increased production of deleterious Reactive Oxygen Species (ROS) is not known, however it is suggested that they either prevent or repair ROS damage. Bats also appear to have resistance to many viral diseases such as rabies, SARS and Ebola and have been shown to be reservoir species for a huge diversity of newly discovered viruses. Thissuggests that their innate immunity is different to other mammals, perhaps playing a role in their unexpected longevity. Here the potential genomic basis for their rare immunity and exceptional longevity is explored across multiple bat genomes and divergent 'ageing' related markers (e.g. microbiome, telomeres, mitochondria, cellular dynamics). A novel blood-based population-level transcriptomics approach is used to explore the molecular changes that occur in an ageing wild population of bats to uncover how bats 'age' so slowly compared with other mammals and to further validate *in silico*functional predictions. These findings provide a deeper understanding of the causal mechanisms of ageing, potentially uncovering the key molecular pathways that could be eventually modified to halt, alleviate and perhaps even reverse this process in humans.

The evolution of senescence: from theory to molecular data

SMBE-OR-024

An insight into the evolution of ageing in mammals: a cross-species phylogenetic genome-phenome approach

X. Farré [1,*], G. Muntané [1 2], A. Navarro [1 3 4]

[1]Institute of Evolutionary Biology (UPF-CSIC), Department of Experimental and Health Sciences, Universitat Pompeu Fabra, Barcelona, [2]Hospital Universitari Institut Pere Mata, IISPV, Universitat Rovira i Virgili, Biomedical Network Research Centre on Mental Health (CIBERSAM), Reus, [3]Centre for Genomic Regulation (CRG), The Barcelona Institute of Science and Technology, [4]Institució Catalana de Recerca i Estudis Avançats (ICREA), Barcelona, Spain

Abstract: Humanity is inevitably growing older and, thus, understanding the genetics of ageing is one of the most pressing challenges worldwide. However, the nature of ageing (or senescence) remains unclear, despite having been deeply studied for decades.

While most studies on aging have focused on individual species, comparative genomics offers a promising avenue of advance. Recent studies of genomic, transcriptomic, and metabolomic data across species with different lifespans are unveiling molecular signatures associated with longevity.

We used the power of cross-species phylogenetic comparative methods to examine the relationship between genomic variation and maximum lifespan (MLS) across 62 mammal species. We report (1) signals of convergent molecular evolution linked to parallel changes in MLS; and (2) genes in which the rates of protein evolution correlate with MLS. Our observations allow identifying a set of ageing-related genes in categories such as inflammatory response, blood coagulation, and immune system pathways, which have been previously linked with longevity using other experimental methods.

Additionally, we identified a set of genes presenting signs of coevolution with longevity. By leveraging the results from a GWAS in parental lifespan, we assessed the contribution of our geneset to current human longevity variation. Our findings shed light into the differences in longevity patterns across mammals and afford lists of candidate genes for follow-up studies.

The evolution of senescence: from theory to molecular data

SMBE-OR-025

Biodiversity of ageing: old evolutionary mechanisms and new genomic approaches

J. P. Magalhaes*

Abstract: Given the extraordinary diversity of life on earth, it is not surprising to observe an immense range of ageing phenotypes. Even amongst animals, some species exhibit an exceptionally quick degeneration while others appear not to age at all. In mammals, the ageing phenotype is remarkably similar, though the pace of ageing can be extremely variable. In this talk, I will present evolutionary theories that aim to explain the biodiversity of ageing as well as possible mechanistic explanations for species differences in ageing. Many traits have been analyzed for correlations with species longevity and our AnAge database of ageing and longevity in animals (http://genomics.senescence.info/species/) is one powerful tool for such comparative approaches. Moreover, recent advances in genome sequencing allow genome-wide cross-species comparisons which open new avenues for unraveling the genetic basis of species differences in ageing. I will discuss novel comparative genomics methods to identify genomic features and specific genes and pathways associated with the evolution of longevity. Furthermore, I will discuss our experiences in *de novo* genome sequencing of long-lived species using next-generation sequencing platforms. The molecular and genetic basis of species differences in ageing remains a major mystery but one that if solved would provide important insights about the roots of the ageing process and human age-related diseases.

The molecular basis of major transitions in evolution
SMBE-OR-003A
COOPERATION AND MAJOR EVOLUTIONARY TRANSITIONS: THE CASE OF EUKARYOGENESIS
P. Lopez-Garcia [1,*]
[1]Ecologie Systématique Evolution, CNRS – Université , Paris, France

Are you a member of SMBE?: Yes
Poster Submission: Major evolutionary transitions represent punctuated increases in average biological complexity. While competition is fundamental to promote biological diversification, cooperation has been causal in major biological evolutionary transitions, from the emergence of life to the multiple origins of complex multicellularity. I will illustrate this using the example of eukaryogenesis. Historically unique, the origin of the eukaryotic cell was elusive for a long time. Although the endosymbiotic origin of mitochondria and chloroplasts from, respectively, Alphaproteobacteria and Cyanobacteria was accepted since the late 1970s, models proposing the symbiotic origin of eukaryotes directly from bacterial and archaeal ancestors were considered heretic. In recent years, the exploration of microbial diversity in natural environments combined with methodological advances making it possible to reconstruct genomes of uncultured microbes from metagenomes have brought new light into eukaryogenesis by allowing the discovery of deep-branching archaea that share more and more similar genes with eukaryotes than other archaea. This strongly supports cooperative, symbiogenetic models for the origin of the eukaryotic cell, whereby higher complexity evolved from the physical integration of prokaryotic cells and extensive gene and genome shuffling. Nonetheless, current symbiogenetic models differ with regard to the tempo and mode of evolution of key eukaryotic traits and many open questions remain. Comparative genomic and phylogenomic studies in light of microbial ecology should contribute to build a plausible detailed evolutionary process leading to the emergence of eukaryotic cells.

The molecular basis of major transitions in evolution

SMBE-OR-002

Testing correlations in morphological and molecular evolution: a meta-analysis

J. Heath [1], J. Keating [1], R. Garwood [1], R. S. Sansom [1,*]

[1]University of Manchester, Manchester, United Kingdom

Abstract: Genomic data have been extremely powerful in unlocking the nature of major evolutionary transitions. However, selection acts directly on the phenotype and only indirectly on the genotype. The interplay between morphological and molecular characters are therefore important considerations in unlocking evolutionary dynamics. Here we take a meta-analysis approach to investigate the interplay between between morphological and molecular characters. Analysis of 15 total-evidence datasets comprising morphological and molecular partitions from different animal clades finds only very weak correlation between morphological and molecular evolution in terms of branch lengths, and most datasets show no correlation at all. We also compare a subset of those data in terms of relative morphological and molecular evolutionary rates in time-calibrate phylogenies. Autocorrelation was eliminated as a factor by considering the pairwise differences in branch length and inter-branch distance within datasets. As such, our meta-analyses suggests that the apparent disconnect between genotypic and phenotypic change might be a general pattern for animals and their phylogenetic data, and this has important implications for studies aiming to reconstructing the mode and tempo of evolutionary changes and transitions from only one class of data.

The molecular basis of major transitions in evolution

SMBE-OR-001

Phylogenomic analyses reveal the genomic novelties accompanying the major transitions of plant life

A. M. C. Bowles [1,*]

[1]University of Essex, Colchester, United Kingdom

Abstract: Identifying the genomic changes that have accompanied the origin of distinct plant groups is key to unravelling the molecular basis of biological innovations. In the last decade, the quantity and quality of complete genomes has dramatically increased, allowing for large-scale phylogenomic comparisons. Approaches that utilise a phylogenetic framework to perform comparative genomics will inform our understanding of the influences of speciation and duplication on gene family innovation, expansion and reduction. Using a phylogenomics pipeline incorporating data for 208 species, evolutionarily conserved gene innovations across the plant phylogeny were identified. Analyses have revealed large numbers of core gene novelties in the ancestors of Embryophyta and Streptophyta, likely involved in the transition of plants from water to land and the evolution of multicellularity. Gene Ontology analysis has identified the functional changes that have accompanied major plant developments. These results will help us to understand the influence of genomic innovations on plant diversification.

The molecular basis of major transitions in evolution

SMBE-OR-005

The role of gene remodeling in animal evolution – patterns of emergence of gene fusion events and placement on the Animal Tree of Life

P. Mulhair [1,*], R. Moran [2], J. Pathmanathan [3], C. Creevey [4], B. Constantinides [1], I. Carr [5], D. Pisani [6], E. Bapteste [7], J. McInerney [8], M. O'Connell [1 8]

[1]School of Biology, University of Leeds, Leeds, United Kingdom, [2]Dublin City University, Dublin, Ireland, [3]Biological Sciences, Columbia University , New York, United States, [4]School of Biological Sciences, Queen's University Belfast, Belfast, [5]Faculty of Medicine and Health, University of Leeds, Leeds, [6]School of Earth Sciences, University of Bristol, Bristol, United Kingdom, [7]CNRS, Muséum national d'Histoire naturelle, Paris, France, [8]School of Life Sciences, University of Nottingham, Nottingham, United Kingdom

Abstract: Novel gene genesis by gene duplication is well studied and is known to play a major role in creating novelty in gene content and function. Comparatively we know a lot less about recombinogenic processes that form remodeled genes. Gene fusions are an example of gene remodeling where two or more distinct protein-coding regions are physically remodelled and combined to form a single open reading frame. These remodelling events require a network-based approach to capture their unique “sharing” properties. We wished to determine what the overall contribution of gene remodelling is to the emergence of novel gene families in the metazoan group. We have assembled a dataset of 1.2 million protein coding genes from 63 high quality metazoan genomes. Applying a sequence similarity network-based approach we have identified 13,632 gene families that have emerged by gene fusion and that are distributed across the Metazoan clade. To mitigate against assembly error in our estimates, we performed a metadata analysis of transcriptome data and confirmed the expression of the break-points of gene fusion, therefore providing evidence that these genes are indeed expressed as a single transcript. We mapped the patterns of emergence of these confirmed gene fusion families and found large numbers that emerged at specific nodes of major transitions including nodes that define the origin of Eumetazoa, Bilateria, Chordata and Euteleostomi. These fusion gene families are enriched for cellular processing and signaling functions. Our results indicate that gene remodeling may have played a key role in major transitions in animal life.

The molecular basis of major transitions in evolution

SMBE-OR-006

Origin of vertebrates at the light of the hagfish genome

J. Pascual-Anaya [1,*], D. Yu [2], Y. Ren [3 4], Y. Li [3], Z. Yong [2], W. Wang [3 4], S. Kuratani [1 5]

[1]Evolutionary Morphology Laboratory, RIKEN Cluster for Pioneering Research , Kobe, Japan, [2]Institute of Zoology, Chinese Academy of Sciences, Beijing, [3]Center for Ecological and Environmental Sciences, Northwestern Polytechnical University, Xi'an, [4]Kunming Institute of Zoology, Chinese Academy of Sciences, Kunming, China, [5]Laboratory for Evolutionary Morphology, RIKEN Center for Biosystems Dynamics Research, Kobe, Japan

Abstract: Comparative genomics analyses between jawed and jawless vertebrates are crucial to understand the ancestral vertebrate genome. Here, I will introduce the efforts done by the inshore hagfish genome consortium to generate a draft genome sequence of the hagfish *Eptatretus burgeri*. First, I will present the state of our reference assembly, made from germ line tissues, and with a scaffold N50 of over 2.6 Mbp. The annotation of this genome was done by the Ensembl's Vertebrate Annotation team of the EBI (European Bioinformatics Institute) and it is one of the most complete among jawless vertebrate genomes published so far. By sequencing at high coverage the genome of 8 somatic tissues, we identify genes that are lost from these tissues during development, including for instance *SPOP*, whose lamprey counterpart is also lost through programmed genomic rearrangement, demonstrating that the mechanisms underlying this genome reorganization are ancestral to cyclostomes. Next, we will present our analysis on chromatin accessibility during hagfish embryology, revealing the regulatory landscape of hagfish development. Last, we investigate whether macro- and micro-synteny changes are conserved or not between cyclostomes and gnathostomes, allowing us to present an up-to-date vision of the still unresolved timing of the 2R-hypothesis around early divergence events of vertebrate evolution.

The molecular basis of major transitions in evolution

SMBE-OR-003

From social to social parasite - the unusual genetic architecture of thelytokous reproduction in honeybees

E. Stolle [1,*], D. Aumer [1], M. Allsopp [2], F. Mumoki [3], C. Pirk [3], R. F. Moritz [1]

[1]Institute of Biology, Martin-Luther-University Halle-Wittenberg, Halle (Saale), Germany, [2]Honey bee Research Section, ARC Plant Protection Research Institute, Stellenbosch, [3]Department of Zoology and Entomology, University of Pretoria, Pretoria, South Africa

Abstract: The evolution of altruism in complex insect societies is arguably one of the major transitions in evolution. While eusocial species became incredible successful, some species and lineages exploit these societies. The workers of the South African Cape honeybee (Apis mellifera capensis) can reverse to selfish behavior by becoming social parasites and parthenogenetically producing female offspring (thelytoky). Although investigated for decades, the genomic basis of this significant change in behavior and associated traits remained elusive.

Using a joint genetic mapping and population genomics approach, we uncover the genetic architecture underlying thelytoky in the Cape honeybee. We found a single coding nucleotide polymorphism in a gene of unknown function at the mapped thelytoky locus to be associated with the iconic thelytokous phenotype. The locus is partially homozygous as part of a strong selective sweep together with the neighboring ecdysis triggering hormone receptor. The locus also is partially heterozygous with a non-recombining haplotype including the neighboring ebony homolog. These results unravel the genomic basis of a loss of social behavior in Cape honeybees and suggest that the unusual genomic architecture represents a balanced lethal/detrimental allele system. Such a system can plausibly explain why thelytoky and its associated traits are specific to the Cape honeybee and cannot easily spread into neighboring honeybee subspecies.

The molecular basis of major transitions in evolution

SMBE-OR-007

Organ complexity and cell history: a case of eye evolution in Cnidaria

N. Picciani [1,*], T. Oakley [1]

[1]Ecology, Evolution, and Marine Biology, UC Santa Barbara, Santa Barbara, United States

Abstract: Animal eyes vary considerably in complexity, strongly impacting the lifestyle of organisms, and are thus valuable for understanding major transitions in trait evolution. While eyes evolved many times in bilaterian animals with elaborate nervous systems, image-forming and simpler eyes also exist in cnidarians, which are ancient non-bilaterians with neural nets and regions with condensed neurons to process information. How often eyes of varying complexity, including image-forming eyes, arose in animals with such simple neural circuitry remained obscure. In this talk, we will present large-scale molecular phylogenies of Cnidaria and their photosensitive proteins (opsins) to show that cnidarian eyes originated at least eight times, with complex, lensed-eyes having a history separate from other eye types. Overall, our results showed eyes evolved repeatedly from ancestral photoreceptor cells in non-bilaterian animals with simple nervous systems, co-opting existing precursors, similar to what occurred in Bilateria. Now, we are leveraging single cell analysis tools to measure gene expression across photoreceptor cells from jellyfish with eyes that originated separately to understand their evolutionary history and interplay with eye origins.

The molecular evolution of cancer

SMBE-OR-051

Mutations beget more mutations – The evolution of mutation rate and the runaway accumulation

Y. Ruan [1,*], C.-I. Wu [1 2 3]

[1]State Key Laboratory of Biocontrol, School of Life Sciences, Sun Yat-sen University, Guangzhou, [2]CAS Key Laboratory of Genomic and Precision Medicine, Beijing Institute of Genomics, Chinese Academy of Sciences, Beijing, China, [3]Department of Ecology and Evolution, University of Chicago, Chicago, United States

Abstract: There is a sizable literature on the evolution of the mutation rate itself. Since the genomic data reveal the mutation burden among samples to vary by several hundred fold, the prevalence of a mutator effect in somatic tissues seems incontrovertible. However, given that mutations in DNA repair genes are not disproportionately represented in cancers, it is unclear how the high mutation rate is acquired. We now propose a runaway model whereby the accumulation of mutator mutations forms a positive-feedback loop. In this loop, any mutator mutation would increase the probability of acquiring the next mutator, thus triggering a run-away escalation in the mutation rate. The model is applicable to the mutational process in both the germline and soma. We show that the process could operate, however, only when mutator mutations are common, although their effects can be weak. With the runaway process, the mutation accumulation would increase with age at a pace faster than an exponential rate. For the germline, the runaway process can be triggered by a baseline mutation rate that has become only twice bigger. For the soma, cancer risk would increase very rapidly once the mutator accumulation gains the momentum, manifested in the long tail of the distribution of the mutation burden. The momentum starts to build up early; thus, even young adults may have a non-negligible chance of developing cancer. In conclusion, germline vs. soma and cancer vs. non-cancer cases form a continuum in the distribution of mutation accumulation underpinned by the runaway process.

The molecular evolution of cancer
SMBE-OR-050
Evolution of drug resistance in chronic myeloid leukaemia: can drug scheduling help?
J. Lindström [1], A. de Wijn [2], R. Friedman [1,*]
[1]Chemistry and Biomedical Science, Linnaeus University, Kalmar, Sweden, [2]Mechanical and Industrial Engineering, Norwegian University of Science and Technology, Trondheim, Norway

Abstract:
Background

Targeted therapies, which directly target molecular pathways critical to tumours instead of rapidly dividing cells in general, have in many cases improved survival significantly when compared to cytotoxic chemotherapy or radiation. Unfortunately, a recurring difficulty with targeted therapies is the occurrence of resistance. A single nucleotide variance may be enough to cause resistance. Even if such a change occurs in a single cancer-cell it brings about a fitness advantage that can generate a cell lineage which oucompetes the rest of the tumour. Such a lineage has therefore a higher probability of becoming a major fraction of the tumour cells. This clonal evolution among tumour cells allows resistance-giving traits to propagate in the population. Once a large proportion of the tumour becomes resistant, the success of continued treatment is unlikely. In many cases, the fitness advantage which allows resistant clones to expand exists only during treatment. Thus, by altering the treatment protocol, it is possible that we could steer the evolution of drug resistance.

Chronic myelooid leukaemia is treated by drugs aimed at a single molecular target, the Abl1 protein. Several drugs exist that inhibit Abl1, with partially-overlapping resistance profiles. Patients usually receive one of three available drugs, and switch to another one if drug resistance is observed. One of the mutations, T315I, is especially problematic. There is only one drug that can target cells carrying the T315I variant, and this drug is associated with potential toxicity. Moreover, once the tumour becomes resistant again, treatment is very difficult as no targeted therapies can help.

Approach

Given that drugs against CML have different resistance profiles, we wished to study whether it could be possible to postpone the emergence of uncontrolled drug resistance by applying different treatment protocols. In particular, we examined a rotation between two drugs. To this end, we implemented a simulation framework. The simulations were based on available resistance mutation growth data and the sequence of the gene. The evolution of a relatively stable pool of cancer stem cells was modelled as a stochastic process, with the growth of cells expressing a tumourigenic protein (Abl1) and any emerging mutants determined principally by the drugs used in the therapy.

Results

The results show that there can be some benefit to a rotation therapy even if the mutation status is unknown, depending on the drug and resistance profile. Furthermore, an interplay between growth inhibition and selection effects generated a non-linear relationship between drug doses and the risk of developing resistance. Overall, we could conclude that drug rotation therapy might be able to delay the onset of resistance in CML patients without an ongoing observation of mutation status. Moreover, the simulations give credence to the suggestion that lower drug concentrations may achieve better results following major molecular response in CML.

The molecular evolution of cancer
SMBE-OR-45B
NEUTRAL EVOLUTION IN CANCER
A. Sottoriva [1,*], M. J. Williams [1], B. Werner [1], C. Barnes [1], T. Graham [1]
[1]The Institute of Cancer Research, London, United Kingdom

Are you a member of SMBE?: No
Poster Submission: The temporal dynamics of cancer evolution remain elusive, because it is impractical to longitudinally ob-serve cancers unperturbed by treatment. Consequently, our knowledge of how cancers grow largely derives from inferences made from a single point in time – the endpoint in the cancer's evolution, when it is re-moved from the body and studied in the laboratory. Fortuitously however, the cancer genome, by virtue of ongoing mutations that uniquely mark clonal lineages within the tumour, provides a rich, yet surreptitious, record of cancer development. I will describe how a cancer's genome can be analysed to reveal its evolu-tionary history and show how the neutral evolution paradigm can explain a substantial amount of genetic patterns routinely observed in cancers. The neutral evolution framework, supported by decades of popula-tion genetics work, also provides the theoretical foundation to understand clonal selection in cancer.

The molecular evolution of cancer

SMBE-OR-048

Modeling the molecular evolution of cancer with scaled selection coefficients and epistasis

J. P. Townsend[*], V. L. Cannataro [1 2], S. G. Gaffney [3]

[1]Biostatistics, Yale University, New Haven, [2]Biology, Emmanuel College, Boston, [3]Yale University, New Haven, United States

Abstract: Since the advent of whole-exome and whole-genome sequencing of tumor tissues, studies of somatic mutations have revealed the underlying genetic architecture of cancer, producing ordered lists of significantly mutated genes that imply their relative importance to tumorigenesis and cancer development. Two quantifications have ordered the importance of discovered cancer "driver" genes: the prevalence of gene mutation among tumor tissues sequenced from that tumor type, and the statistical significance (P value) of the disproportionality of mutation frequency. However, neither of these metrics provide insight into the evolutionary process. Building on classical population genetics and molecular evolutionary theory, we have developed a data-driven gene- and site-specific molecular evolutionary model that provides a comprehensive means to evaluate how somatic mutations interact to influence cancer progression in multiple cancer types. Our approach can be projected to provide a genotype-based fitness landscape for tumorigenesis. It quantifies how mutation and selection lead to stage-specific progression and metastasis in each type of cancer, in particular revealing the epistatic effects between known somatic mutations and providing probabilistic trajectories of tumor evolution associated with cancer stage progression and therapeutic treatment. Predictions from the model enable better basic research prioritization, faster therapeutic discovery, efficient clinical trial design, and inform medical decision-making by precision medicine tumor boards.

The molecular evolution of cancer

SMBE-OR-047

The evolution of cancer microRNAs is driven by regulatory network constraints

N. Akhundova [1], M. Helmy [1], A. Marco [1,*]

[1]School of Biological Sciences, University of Essex, Colchester, United Kingdom

Abstract: Genes associated to cancer are traditionally classified as either oncogenes or tumor-suppressor (TS) genes. MicroRNAs, short RNA regulatory molecules, are also known to behave as either oncogenes (the so-called oncomiRs) or as TS microRNAs. When and how microRNAs became important regulatory elements within the cancer-associated gene network is still unknown. We hypothesize that TS microRNAs generally repress the activity of oncogenes while oncomiRs down-regulate TS genes. We formalized this hypothesis into a dynamic model which predicted that, as transcriptional regulatory networks increase in complexity, novel oncomiRs are more likely to emerge than TS microRNAs. Using comparative genomics we confirmed that oncomiRs are evolutionarily younger than TS microRNAs. Specifically, both the recent emergence of de novo sequences and the duplication of existing microRNAs was associated to oncomiRs. We also showed that microRNA/target interactions are evolutionary more recent between oncomiRs and TS genes than between TS microRNAs and oncogenes, further supporting the predictions from our model. In conclusion, we provide evidence that constraints in gene regulatory networks determine that novel oncomiRs are more likely to emerge during evolution than TS microRNAs.

Resistance evolution in real-time

SMBE-OR-019

Evolution-at-large: Field epidemiology reveals novel insights from a national-scale selection experiment.

P. Neve*

Abstract: Weedy plants rarely grab the headlines, yet globally weeds result in more lost food production than both insect pests and plant pathogens. There is currently a global epidemic of evolved resistance to herbicides, exemplified in the UK by widespread multiple resistance in the grass weed, Alopecurus myosuroides (blackgrass). Previous studies have used (i) theoretical models and (ii) experimental evolution to establish that mixtures, rotations and/or sequences of multiple pesticide modes of action can significantly slow the evolution of resistance. However, most models assume that populations evolve discrete or 'specialist' mechanisms of resistance, meaning that control strategies based on heterogeneity of selection only break down when multiple, independent resistance mechanisms simultaneously evolve. In 2014, we established a national network of 71 farms to monitor blackgrass populations, including assays to characterise the frequency and molecular genetic basis of herbicide resistance. Uniquely, we were also able collect historical field management data (from 2005 - 2014), including data on herbicide use, providing details of the selection regime to which each of our blackgrass populations had been exposed. We found that the resistance epidemic in UK blackgrass is underpinned by two co-existing and interacting mechanisms – specialist target site resistance and a more generalist mechanism of resistance, based on enhanced herbicide metabolism, which confers broad spectrum resistance to a range of herbicides. We used statistical models to show that selection regimes characterised by more diverse herbicide use did indeed select against target site-based resistance as predicted by models. However, more diverse herbicide selection regimes preferentially selected for more generalist resistance mechanisms. The results demonstrate that optimal resistance management is contingent on the mechanism and genetics of resistance traits and that herbicide (pesticide, antimicrobial etc.) diversity may not universally slow evolution of resistance. These results are relevant to all systems where generalist resistance mechanisms can evolve.

Resistance evolution in real-time

SMBE-OR-017A

Adaptive modelling of antibiotic resistance through genomics-based surveillance and machine learning

N. E. Wheeler [1,*], K. C. Ma [2], A. Hicks [2], L. Jenniches [3], Y. Grad [2], L. Barquist [3], S. Harris [4], D. Aanensen [1]

[1]Centre for Genomic Pathogen Surveillance, Wellcome Sanger Institute, Hinxton, United Kingdom, [2]Department of Immunology and Infectious Diseases, Harvard T.H. Chan School of Public Health, Boston, United States, [3]Helmholtz Institute for RNA-based Infection Research, Wurzburg, Germany, [4]Wellcome Sanger Institute, Hinxton, United Kingdom

Abstract: With the increasing use of whole-genome sequencing in public health microbiology laboratories and hospitals, prediction of antibiotic resistance using WGS data offers a promising approach for rapidly informing clinical and public health decision making. Machine learning algorithms have been proposed as a way to revolutionise our understanding of genotype-to-phenotype relationships and dynamically model and predict antibiotic resistance as WGS data is collected. A number of promising machine learning algorithms for predicting antibiotic resistance in pathogens of high clinical significance have been developed in recent years. However, evaluation of the sources of error or bias in these models is lacking. In particular, the influence of population structure in training datasets on the patterns learned by machine learning algorithms has not yet been illustrated and is not well appreciated. We compare machine learning methods for predicting resistance using several examples that vary in resistance mechanism and degree of population structure, and identify the best approaches for building models that generalise well beyond training data.

Resistance evolution in real-time

SMBE-OR-017

Genetics of the evolution of a male-killing suppressor in the butterfly Hypolimnas bolina

L. Reynolds [1,*], E. Hornett [1], G. Hurst [1]

[1]Evolution, Ecology, and Behaviour, University of Liverpool, Liverpool, United Kingdom

Abstract: The evolution of resistance is often observed in real time in short-lived organisms e.g. evolution of antibiotic resistance in bacteria. It is rare for resistance evolution to be seen in real time in organisms with longer generational times, and even rarer for it to be observed naturally, rather than in response to anthropogenic influences. The *Hypolimnas bolina* - *Wolbachia* study system is therefore unusual in that we have witnessed resistance evolution in a wild population. Populations of the butterfly *H. bolina* are infected with a male-killing strain of the maternally-inherited endosymbiont, Wolbachia. In response to infection, we observed the spread of a suppressor of male-killing in Samoan *H. bolina* between 2001 and 2005. Here, we examine the pattern of spread and genetic basis of suppression. Our results reveal a 20cM wide selective sweep associated with change at a single locus in the *H. bolina* genome. Our ongoing research seeks to understand the nature of this mutation, and why there was a long evolutionary lag before it arose and spread.

Resistance evolution in real-time

SMBE-OR-018

Real-time spread and evolution of plasmid-mediated carbapenem resistance in a hospital

J. De La Fuente Hidalgo[*], R. León-Sampedro [1], A. Alonso-Del Valle, J. Rodríguez-Beltrán [1], C. Delavega-Losada [1], C. Díaz-Agero [1], T. Crellen [2], P. Musicha [2], M. Hernández-García [1], P. Ruiz-Garbajosa [1], N. López-Fresneña [1], R. Cantón [1], B. S. Cooper [2], Á. San Millán [1]

[1]Instituto Ramón y Cajal de Investigación Sanitaria (IRYCIS), Madrid, Spain, [2]Oxford University, MORU,, Bangkok, Thailand

Abstract: Conjugative plasmids play an essential role in the horizontal spread of antimicrobial resistance genes among bacteria. Epidemiological data reveal that particular associations between plasmids and bacterial clones become especially successful, creating multi-drug resistant bacteria that spread in clinical settings. Good examples of these associations are *Klebsiella pneumoniae* clones carrying the carbapenemase-coding plasmid pOXA-48. Although little is known about the factors underlying the success of these associations in clinical settings, *in vitro* evidence reveals that the fitness costs produced by plasmids in their bacterial hosts, and the subsequent compensatory evolution, determine the fait of plasmid-carrying clones in bacterial populations.

Here, we analysed the evolutionary dynamics of plasmid pOXA-48 in a clinical setting by using an enterobacterial collection recovered from hospitalized patients as part of the project R-GNOSIS (http://www.r-gnosis.eu). We used PacBio and Ilumina HiSeq-4000 technologies to perform plasmid and whole genome sequencing of all pOXA-48-carrying isolates (244, from 135 patients) detected in our hospital during a 5-year period. Then, combining genomic and epidemiological data with mathematical modeling, we analyzed the *in vivo* spread and evolution of pOXA-48 in the hospital and the evolution of pOXA-48-carrying enterobacteria in the gut microbiome of hospitalized patients. Our results showed that specific clones of *K. pneumoniae* (mainly ST11) are responsible for the between- and within-patient dissemination of pOXA-48 in the hospital. Moreover, we were able to reconstruct the between-patient transmission routes of plasmid-carrying clones. Finally, we detected and characterized several putative compensatory mutations accumulated in pOXA-48 over time, which are mainly located on conjugative genes.

Understanding the genomics of climate change response

SMBE-OR-130

The role of asexual evolution and epigenetic changes in the climate change responses of corals

I. Baums [1,*]

[1]The Pennsylvania State University, University Park, United States

Abstract: Corals built the three-dimensional structure of one of the most diverse ecosystems of the plant. Yet, their ecological role is threatened by climate change because small increases in ocean temperature cause dysbiosis of the coral animal and its photosynthetic algal symbionts. Despite sharp population declines, large standing genetic and phenotypic variation remains in even the most threatened reef builders and genomic tools are increasingly available to study this variation as a clue to how coral may respond to climate change. Genomic data have clarified the contribution of genetic diversity of host, algal and prokaryotic partners to temperature stress in great detail. The most puzzling findings are cases of phenotypic heterogeneity in stress response among isogenic host-symbiont modules. Several mechanisms can cause such heterogeneity and of these, data is accumulating on somatic mutations and differential methylation in marine invertebrates. The first comprehensive datasets on variation in DNA methylation in natural stands of reef-building corals allowed us to attribute some of the variation in intra-genet stress response to epigenetic changes but the data points to other, as yet unexplored mechanisms that cause intra-genet variation in corals such as microenvironmental variation, stochastic gene expression, and somatic mutations. In fact, we now have evidence for inheritance of somatic mutations via the larval stage in coral genets that can attain ages of over 1000 years. These observations open a new chapter in considering the role of asexual evolution and epigenetic changes in the response of marine foundation fauna to climate change.

Understanding the genomics of climate change response

SMBE-OR-132

Intensified selection and elevated mutation in wild barley over 28 years of global warming

Y.-B. Fu [1,*]

[1]Plant Gene Resources of Canada, Agriculture and Agri-Food Canada Saskatoon RDC, Saskatoon, Canada

Abstract: Many studies have investigated the threats of climate changes on wild plants, but few have investigated the genetic responses of plant natural populations under threats. We characterized the genetic responses of 10 wild barley (*Hordeum spontaneum* K. Koch; WB) populations in Israel, sampling them in 1980 and again in 2008, through an exome capture analysis. Sequencing 48 WB samples representing two sampling years generated six million SNPs and SNP annotations identified 12,926 and 13,361 deleterious SNPs for the 1980 and 2008 samples, respectively. As expected, WB samples displayed intensified selective sweeps and elevated deleterious mutations across seven chromosomes over 28 years of global warming. On average, the 2008 samples had lower individual and population mutational burdens, but acquired more adaptive mutations than the 1980 samples. These findings illustrate the feasibility of characterizing genetic responses in plant natural populations under threats and are useful for understanding the evolutionary dynamics of crop wild relatives in threatened natural populations.

Understanding the genomics of climate change response

SMBE-OR-133

Population genomics of Ice Age gray wolves

A. Bergström [1], U. Talon [2], L. Frantz [3], M.-H. S. Sinding [4], E. Ersmark [5], V. Schuenemann [6], T. Davy [1], L. Girdland-Flink [7], S. Gopalakrishnan [4], J. Ramos-Madrigal [4], T. Feuerborn [4 8], S. Münzel [9], S. Pfrengle [8], N. Conard [8], C. Carøe [4], J. Haile [10], A. Linderholm [10], D. Stanton [5], N. Benecke [11], H. Bocherens [8 12], R. F. Carden [13], S. Fedorov [14], M. Lázničková-Galetová [15], M. Germonpré [16], V. V. Ivanova [17], R. P. Jennings [7], A. K. Kasparov [18], I. V. Kirrilova [19], Y. Kuzmin [20], P. Nikolskiy [21], E. Y. Pavlova [22], V. V. Pitulko [18], M. Sablin [23], F. K. Shidlovskiy [19], J. Storå [24], J. A. Leonard [25], A. Götherström [24], M. Hofreiter [2], M. T. P. Gilbert [4], G. Larson [10], J. Krause [8 26], L. Dalén [5], P. Skoglund [1,*]

[1]Francis Crick Institute, London, United Kingdom, [2]University of Potsdam, Potsdam, Germany, [3]Queen Mary University, London, United Kingdom, [4]University of Copenhagen, Copenhagen, Denmark, [5]Swedish Museum of Natural History, Stockholm, Sweden, [6]University of Zürich, Zürich, Switzerland, [7]Liverpool John Moores University, Liverpool, United Kingdom, [8]University of Tuebingen, [9]Institute for Archaeological Sciences, Archaeozoology, University of Tuebingen, Tuebingen, Germany, [10]University of Oxford, Oxford, United Kingdom, [11]German Archaeological Institute, Berlin, [12]Senckenberg Center for Human Evolution and Palaeoenvironment (s-HEP), Tübingen, Germany, [13]University College Dublin, Dublin, Ireland, [14]North-Eastern Federal University, Yakutsk, Russian Federation, [15]Moravian Museum, Brno, Czech Republic, [16]OD Earth and History of Life, Royal Belgian Institute of Natural Sciences, Brussels, Belgium, [17]FSBI VNIIOkeangeologia, Saint-Petersburg, [18]Institute for the History of Material Culture, Russian Academy of Sciences, St. Petersburg, [19]Ice Age Museum, Moscow, [20]Institute of Geology and Mineralogy. Siberian Branch of the Russian Academy of Sciences, Novosibirsk, [21]Geological Institute, Russian Academy of Sciences, Moscow, [22]Arctic & Antarctic Research Institute, Saint-Petersburg, [23]Zoological Institute of the Russian Academy of Sciences, St. Petersburg, Russian Federation, [24]Stockholm University, Stockholm, Sweden, [25]Estación Biológica de Doñana, Sevilla, Spain, [26]Max Planck Institute for the Science of the Human Past, Jena, Germany

Abstract: The gray wolf is distributed across diverse environments in Eurasia and North America, and persisted as a species through dramatic periods of climate change, including the time around the Last Glacial Maximum when many other megafaunal populations went extinct. However, the fossil record has suggested the existence of distinct wolf ecomorph populations that may have gone extinct. To understand gray wolf population dynamics, history, and genomic adaptations, we sequenced 48 ancient wolf genomes from Europe, Siberia and Alaska spanning the last 50,000 years, to a median 1-fold depth. We find that all present-day Eurasian wolves trace their ancestry to a founding population emerging after the Last Glacial Maximum (LGM), with direct evidence for replacement of the Ice Age populations in both western and eastern Eurasia. Domestic dogs also trace their ancestry to the Eurasian wolf population emerging after the LGM, and we observe the first appearance of differentiation between the dog and present-day gray wolf lineages shortly after this time. We identify genomic regions that were highly differentiated in the Ice Age Eurasian wolf lineages, that may have had adaptations to the cold climates and prey niche of that period. Finally we use the deep time resolution of the dataset to reconstruct allele frequency trajectories through time, to identify genomic regions that may have conferred adaptations to the warming climate. This dataset thus provides the first direct genomic time transect of a natural mammal population and its evolution throughout the changing environmental conditions of the last 50,000 years, persisting through the last glacial maximum in the face of local extinctions and giving rise to a key domesticated animal.

Understanding the genomics of climate change response

SMBE-OR-131

Understanding insect evolutionary responses to shifting climatic gradients

R. Dudaniec [1,*]

[1]Macquarie University, Sydney, Australia

Abstract: Wide-ranging insects that occupy diverse environmental conditions are ideal for understanding how and why local adaptation develops, particularly within the dynamic conditions present at shifting range margins. Importantly, neutral genetic structure across large areas can lead to false positive detection of selection signatures, while high gene flow is thought to limit the capacity of species to locally adapt. These considerations make simultaneous analyses of neutral landscape genetic structure and adaptive genetic processes valuable for disentangling common and unique factors that affect how species are responding to changing environments. Here I discuss recent approaches using Restriction site-Associated DNA sequencing data (RADseq) to examine the evolutionary responses of wide-ranging insects to contrasting environmental gradients. With concurrent analyses of landscape genetic connectivity and loci exhibiting signatures of environmental selection, I present case studies on two insect species that are subject to effects of climate change across wide latitudinal gradients, including: 1) range expanding damselflies in Europe (Ischnura elegans) and 2) pest grasshoppers within agricultural Australia (Phualacridium vittatum). These studies demonstrate interactions between neutral genetic processes and the environmental and morphological variables that contribute towards observed patterns of selection. I further discuss the functional relevance of annotated candidate genes involved in local environmental adaptation. Finally I explore the potential for interpreting and applying results from these studies to make predictions about species' evolutionary responses under ongoing climate change.

Understanding the genomics of climate change response

SMBE-OR-134

Characterizing the molecular basis of desiccation resistance in Drosophila melanogaster European populations

V. Horvath [1,*], J. Salces-Ortiz [1], G. E. Rech [1], S. Guirao-Rico [1], J. Gonzalez [1]

[1]Institute of Evolutionary Biology (CSIC-UPF), Barcelona, Spain

Abstract: Climate change is altering precipitation and water availability thus exposing species and ecosystems to increasing harnessing conditions such as desiccation. Adaptation to dry environments is a complex trait and little is known about the molecular changes underlying this process. *Drosophila melanogaster* is a great model to study the response of populations to rapidly changing environments because of its subtropical origin and recent and fast worldwide spreading. In this study, we subjected to desiccation stress 74 natural *Drosophila* European strains, collected from three different climate zones. We found, that the level of desiccation tolerance of these strains correlates with the climate zone in which they were collected. Indeed, some of the variability can be explained by the effect of altitude, evaporation and annual precipitation. We also performed RNA-seq analysis in six of the most and least tolerant strains. The results show that our dataset is enriched in genes related to metabolic pathways such as cuticular wax and fatty acid metabolic processes. Protein-protein interaction network analysis allowed us to identify the genes most likely to be playing a central role in desiccation tolerance. In addition, we found 80 transposable elements nearby these genes, which are very powerful mutagens and have been proved to cause adaptive mutations. Our results help to understand the relation between environmental changes and adaptive mechanisms, which is critical for predicting climate change impacts and to successfully conserve biodiversity for agriculture and human health-related reasons.

Understanding the genomics of climate change response

SMBE-OR-135

The genomic architecture of variation in thermal responses of Atlantic cod at multiple spatial scales

R. Oomen [1,*], S. Jentoft, H. Knutsen, E. M. Olsen, E. Juliussen, B. Star, J. Hutchings

[1]Centre for Ecological and Evolutionary Synthesis, Department of Biosciences, University of Oslo, Oslo, Norway

Abstract: Genetic and genomic architecture play a key role in the origin and maintenance of local adaptation. However, the genomic basis of adaptation with gene flow remains poorly understood. The question of whether different genomic architectures underlie adaptation at different spatial scales with varying levels of gene flow has not been empirically explored. The genomic basis of adaptation in thermal responses is of particular interest for predicting population responses to climate change. Integrating a common garden experiment across a range of temperatures with large-scale RNA sequencing and SNP analysis of Atlantic cod (*Gadus morhua*) larvae of wild origin, we find variation in gene expression means and thermal plasticities, growth reaction norms, and survival consistent with divergence at macro- and micro- geographic spatial scales along coastal Norway. Thermal responses were uniquely impacted by: 1) parental pre-spawning environments, 2) coexisting ecotypes with low genome-wide divergence, and 3) three chromosomal inversions. At a macrogeographic scale characterized by moderate gene flow, low-level differentiation located throughout the genome contributes most to local adaptation. At a microgeographic scale with high gene flow, structural variants contribute most to local adaptation. We present the first experimental evidence for the putative adaptive functions of these inversions that are polymorphic throughout the species range. Our findings help us to understand how climate change will affect population dynamics and distributions of cod and other high gene flow species and support the development of a new management strategy for coastal cod in Norway.

Using Ancient DNA to Study Natural Selction: New Models and New Data

SMBE-OR-027

Inferring the selection history of Europe over the last 10,000 years using a novel statistical approach

L. N. Ormond [1,*], J. Bloecher [2], K. Kirsanow [2], Z. Hofmanova [3], K. Wang [4], J. Mendoza [5], S. Figarska [6 7], J. Burger [2], G. Hellenthal [8]

[1]Department of Genetics, Evolution and Environment, UCL Genetics Institute, London, United Kingdom, [2]Palaeogenetics Group, Johannes Gutenberg-University Mainz, Mainz, Germany, [3]Department of Biology, University of Fribourg, Fribourg, Switzerland, [4]Department of Archaeogenetics, Max Planck Institute for the Science of Human History, Jena, Germany, [5]Human Evolutionary Genetics Unit, Institut Pasteur, Paris, France, [6]Department of Medicine, Stanford University School of Medicine, [7]Stanford Cardiovascular Institute, Stanford, United States, [8]UCL Genetics Institute, Department of Genetics, Evolution and Environment, London, United Kingdom

Abstract: We describe an efficient new Bayesian statistical model to identify selection in admixed populations using allele counts from Single Nucleotide Polymorphisms (SNPs) in low and/or high coverage ancient and/or modern genome data. This novel approach accounts for demography and variation in coverage across loci and samples, and also infers proportions of ancestry relating populations (e.g. due to admixture) and levels of genetic differentiation among groups. The program can provide both selection probabilities for individual SNPs and/or jointly test sets of SNPs (*e.g.*in pathways) for selection effects. We demonstrate the model's utility through simulations, and we showcase its ability to identify previous targets of selection using DNA from prehistoric and modern humans. We apply our method to a large dataset of ancient and modern genomes, including 75 previously unpublished data from ancient genomes, that span hunter-gatherer and Neolithic early farmer populations from 12,000 years BC through until the late Bronze Age. We identify the time periods over which individual SNPs have experienced selection, and we assess the evidence for selection in sets of SNPs associated with diet, immunity, skin pigmentation and the metabolic syndrome, illustrating the evolutionary constraints on populations at critical periods throughout the history of Europe.

Using Ancient DNA to Study Natural Selction: New Models and New Data

SMBE-OR-028

Evolving in a bottleneck: tracking natural selection of a fungal pathogen over the domestication and dispersal of maize

N. Wales [1,*], M. D. Martin [2], B. K. Blackman [3]

[1]Department of Archaeology, University of York, York, United Kingdom, [2]NTNU University Museum, Norwegian University of Science and Technology, Trondheim, Norway, [3]Department of Plant & Microbial Biology, University of California, Berkeley, Berkeley, United States

Abstract: While many researchers investigate domestication to infer how plant and animal species have been impacted by human selection, it is increasingly recognized that endophytes and obligate parasites of domesticated species can also undergo rapid evolution in response to changes in host morphology and introduction to new environments. This form of natural selection can be studied across wide temporal scales using ancient and historic samples, insofar as genetic traces of the symbiotic organisms are preserved. In this project we are exploring how metagenomes preserved on herbarium specimens can reveal coevolutionary processes between hosts and pathogens.

Ustilago maydis, the fungus responsible for corn smut disease, presents an ideal case study to understand how crop pathogens react to domestication of the host and rapid dispersal to new geographic regions. *U. maydis* is well represented in herbarium collections due to its agricultural importance, as it consistently decreased maize yields before the breeding of resistant cultivars in the early-1900s. It is currently unknown how the fungus adapted from its likely origins in subtropical Mexico to the cold winters of high latitude regions in the USA and South America. Moreover, we currently do not understand if the pathogen's virulence changed with the development of new varieties of maize. To shed new light on these questions, we characterized low to medium depth (1–25×) genomes from 28 historic teliospore specimens. In addition, we established the first reference panel of high-coverage whole genomes from *U. maydis* collected throughout the Americas. The estimated ages of geographic lineages and signatures of selective sweeps provide important insights on how this pathogen was transported to distant regions and responded to new conditions. Our results show that phylogeographic structure was maintained in different locations across the Americas from the late 19th century to today, with little evidence for replacement even in our globally connected agricultural system, a finding that may help assess the risk of future pathogen dispersals.

Using Ancient DNA to Study Natural Selction: New Models and New Data
SMBE-OR-029
Using patterns of introgressed archaic ancestry to detect adaptive evolution in modern humans
B. Shapiro[*], N. Schaefer[1], R. E. Green[1]
[1]University of California Santa Cruz, Santa Cruz, United States

Abstract: An ancestral recombination graph (ARG) is a precise description of how every genome is related to every other genome across all variable positions in the genome. In this way, an ARG is a complete description of the relatedness of individuals and can be used to ask fundamental questions about demography and selection. We present a new heuristic, parsimonious, ARG inference algorithm called SARGE (Speedy Ancestral Recombination Graph Estimator), and use it to map segments of archaic ancestry across a panel of 279 modern human genomes from the Simons Genome Diversity Panel. We locate high-frequency Neanderthal and Denisovan variants present in moderns humans including sub-Saharan Africans, and show that Denisovan-like ancestry in modern humans may have arisen from multiple instances of admixture. Finally, we catalog genomic regions for which modern human and archaic hominin lineages are completely sorted and discuss how these may have played a role in the evolution of the modern human lineage.

Using Ancient DNA to Study Natural Selction: New Models and New Data
SMBE-OR-030
Demography-free detection of selective sweeps in low-coverage ancient genomes
T. Davy [1,*], P. Skoglund [1]
[1]The Francis Crick Institution, London, United Kingdom

Abstract: Ancient genome studies of selective sweeps has so far mostly focused on the evolution of present-day populations, since DNA degradation complicates the direct study of adaptation in ancient populations. We compiled 263 published ancient shotgun genomic sequences of at least 1-fold sequencing depth, excluding SNP panels subject to ascertainment biases. To utilise signals of diploid genotypes from low-coverage genomes (~1-10X), we use maximum likelihood methods to propagate uncertainty in the genotype calls, thereby estimating cluster membership components and ancestral per-SNP allele frequencies for genic regions using *NGSadmix*. This circumvents arbitrary population labels, and grants increased power by harnessing admixed individuals, such the Middle Neolithic descendants of Mesolithic and Early Neolithic populations in Europe. We searched for outlier regions of excess allele frequency differentiation in windows of equal numbers of SNPs in several ancestries, including Eurasian Hunter-Gatherer populations, northern and southern Native Americans, Neolithic Farmers, and Steppe populations. We identify outliers with a genome-wide null distribution, computing our window-based statistic on distant, independent genomic loci. This approximated null distribution allows us to compute p-values for all loci. We observe several candidate selective sweep regions that exceed the standard genome-wide significance threshold corrected for multiple testing ($P < 10\text{-}6$). This includes recapitulation of previously described signals around the *HLA,* and *EDAR* loci, alongside new candidate selective sweep regions in other ancient populations. Our study provides a generalized framework for studying selective sweeps in low-coverage genomic data, with robust outlier detection independent of demographic history.

Using Ancient DNA to Study Natural Selction: New Models and New Data
SMBE-OR-031
The Neanderthal Protein Structure Atlas: Comprehensive modeling of Neanderthal 3D structures
D. Rinker, G. Sliwoski, J. A. Capra*

Abstract:

Whole exome and genome sequences from several Neanderthal and Denisovan individuals have identified hundreds protein-coding variants that are specific to archaic hominins or to anatomically modern humans (AMHs). Comparisons of the allele frequencies, heterozygosities, and computational variant effect predictions suggest that patterns of coding variation in Neanderthals were substantially different than in AMH populations, likely due to their smaller effective population sizes. As a result of Neanderthals' very low genetic diversity, they likely had more deleterious missense variants than AMHs. However, previous studies have based their conclusions on computational methods for inferring missense variant consequences that have a number of shortcomings, most notably that they are trained on biased sets of human mutations and very frequently disagree with one another.

The dramatic growth of available protein 3D structural models of human proteins provides an opportunity to directly model the effects of non-synonymous variants specific to Neanderthal and AMH populations on protein structure and function. Because this approach is based on biophysical modeling of the 3D structure of the protein, it is less subject to the shortcomings of previous comparisons. Experimentally derived protein structures are available for nearly 25% of human proteins, and structural models based on homologous structures are available for more than 80% of human proteins, though these computationally derived models are of variable quality and coverage.

To directly model the structural effects of non-synonymous variants specific to Neanderthals or to AMHs, we comprehensively mapped all such variants into all available protein 3D structural models. This enabled the quantification of the effect of these variants on protein structure, stability, and interactions. We highlight several cases in which these variants cause substantial perturbations to protein structure. Next, we modeled the structural effects of protein-coding variants present in modern human populations due to Neanderthal introgression. These analyses enable comparison with the effects of non-introgressed variants, and they suggest molecular hypothesis about how specific introgressed coding variants associated with organism-level phenotypes in AMHs influence protein function. Finally, we also place the Neanderthal- and AMH-specific variants in the spatial context of all known benign and disease-associated human variants. Overall, this synthesis provides a framework for generating and evaluating hypotheses about the mechanistic molecular effects of missense variants relevant to Neanderthal- and AMH-specific biology.

Open Symposium Abstracts

The molecular evolution of cancer

SMBE-OR-045

Early origin and fast metastatic dissemination in colorectal cancer

J. M. Alves [1 2 3,*], S. Prado-Lopez [1 2 3], J. Cameselle-Teijeiro [4 5], D. Posada [1 2 3]

[1]Department of Biochemistry, Genetics and Immunology, [2]Biomedical Research Center (CINBIO), Universidad de Vigo, [3]Galicia Sur Health Research Institute, Vigo, [4]Department of Pathology, Clinical University Hospital, [5]Galician Healthcare Service (SERGAS), Santiago de Compostela, Spain

Abstract: Colorectal cancer (CRC) is the third most commonly diagnosed cancer worldwide, with metastatic disease representing the primary cause of mortality. While previous studies have provided significant insights into different aspects of metastatic dissemination, it is currently unclear how, and when, CRC cells start spreading to surrounding and distant tissues. Here we illustrate how a Bayesian phylogenetic/biogeographic framework has much power in order to elucidate the evolutionary history of CRC within a patient, leveraging whole exome sequencing data obtained from multiple primary and metastatic locations. Our results support an early and very rapid monoclonal metastatic progression, occurring only two years after tumor initiation, which resulted in a fairly rapid expansion of the tumor cell population in parallel at both primary and metastatic lesions. Moreover, despite the overall negative effect of spatial distance on the dispersal ability of malignant cells, our biogeographic analysis suggests an hematogenous metastatic spread, with the primary tumor directly seeding liver metastases without an apparent early involvement of the lymphatic system.

In conclusion, our work provides a finer picture of the clonal dynamics over time and space in CRC, and represents the first attempt to clarify important properties of cancer evolution from a solid model-based biogeographical framework.

The molecular evolution of cancer

SMBE-OR-046

Evolution and genetic prediction of prostate cancer risks in African men

J. Lachance [1,*]

[1]Georgia Institute of Technology, Atlanta, United States

Abstract:

Prostate cancer is a highly heritable disease that disproportionately affects men of African descent. What are the evolutionary causes of this health inequity? Focusing on germline mutations, we integrated results from genome-wide association studies with individual-level genetic data from 45 African and 19 non-African populations. Tests of natural selection were also used to assess why some SNPs have large allele frequency differences across populations. We found that genetic predictions of prostate cancer risks are highest for West African men and lowest for East Asian men. These differences are due to neutral and selective processes, including the out-of-Africa bottleneck and genetic hitchhiking. A small number of loci appear to drive elevated prostate cancer risks in men of African descent including rs9632117, rs6983267, rs10896449, rs10993994, and rs817826. Although most prostate cancer-associated loci are evolving neutrally, there are multiple instances where alleles have hitchhiked to high frequencies with linked adaptive alleles. For example, a protective allele at 2q37 appears to have risen to high frequency in Europe due to selection acting on pigmentation.

Despite the above discoveries, the vast majority of known disease associations have been detected in individuals who have European ancestry. Furthermore, most genome-wide association studies have used genotype arrays that are hindered by SNP ascertainment bias. These factors cause genetic predictions of prostate cancer risk to be mis-estimated across global populations. To overcome these disparities in genomic medicine, we have developed a custom genotyping array that is optimized for detecting prostate cancer associations in sub-Saharan African populations. The MADCaP (Men of African Descent Carcinoma of the Prostate) Array contains more than 1.5 million markers and an imputation backbone that successfully tags 94% of common genetic variants in African populations (MAF > 0.05, r^2 threshold = 0.8). To aid in fine-mapping, the MADCaP Array also has a high density of markers in genomic regions surrounding known cancer associations (including 8q24). Using the MADCaP Array, we conducted a pilot study of 843 prostate cancer cases and controls. These samples were collected at seven study sites in Ghana, Nigeria, Senegal, and South Africa. This novel dataset enabled us to identify non-African GWAS results that replicate well in African populations. We also assessed how well polygenic risk scores are able to distinguish between prostate cancer cases and controls in African populations.

The molecular evolution of cancer

SMBE-OR-049

Large-scale DNA-based tracking of the evolution of copy number alterations during xenograft engraftment and passaging

X. Woo [1], A. Srivastava [1], Z.-M. Zhao [1], M. Lloyd [2], Y.-S. Suh [1 3], C. Lee [1 4], C. Bult [2], J.-I. Kim [3], H.-K. Yang [3], B. E. Welm [5], A. L. Welm [5], S. Scherer [5], M. P. Klein [6], E. Cortes-Sanchez [5], M. Bailey [5], M. T. Lewis [7], J. Roth [8], B. Fang [8], F. Meric-Bernstam [8], M. Davies [8], M. Herlyn [9], V. Rebecca [9], A. Kossenkov [9], P. Wickramasinghe [9], L. Ding [10], R. Govindan [10], J. Mashl [10], H. Sun [10], S. Davies [10], Y. Evrard [11], R. Patidar [11], L. Chen [11], E. Medico [12], J. Giordano [12], C. Isella [12], J. Rosains [13], C. Frech [13], J. Randjelovic [14], R. Jeon [13], D. Dean [13], J. Chuang [1 15,*]

[1]THE JACKSON LABORATORY FOR GENOMIC MEDICINE, Farmington, [2]The Jackson Laboratory, Bar Harbor, United States, [3]Seoul National University, [4]Ewha Woman's University, Seoul, Korea, Republic Of, [5]Huntsman Cancer Institute, [6]University of Utah, Salt Lake City, [7]Baylor College of Medicine, [8]MD Anderson Cancer Center, Houston, [9]The Wistar Institute, Philadelphia, [10]Washington University, St. Louis, [11]Frederick National Laboratory for Cancer Research, Frederick, United States, [12]Universita di Torino, Turin, Italy, [13]Seven Bridges Genomics, Charlestown, United States, [14]Seven Bridges Genomics, Belgrade, Serbia, [15]University of Connecticut Health Center, Farmington, United States

Abstract: Patient-derived xenografts (PDX) are in vivo models of human cancer in which a resected tumor is engrafted into a mouse for expansion and therapeutic testing. Numerous studies have shown that PDX models retain the therapeutically relevant genomic aberrations of the patient tumors and are able to mimic the treatment responses observed in patients. However, there have also been reports that additional evolution relevant to treatment response can occur during the PDX passaging of the tumor. Currently, the tracking of genomic alterations that occur in xenografts has been limited, with prior studies focusing on small sets or on RNA data. We assembled over 800 PDX samples and their originating patient tumors from more than 300 PDX models, across a variety of tumor types as part of the U.S. National Cancer Institute PDXNet (PDX Development and Trial Centers Research Network) Consortium. We estimated copy number alterations (CNA) profiles from data collected under multiple platforms (SNP arrays, whole-exome sequencing, and/or RNA-sequencing). Through matched comparisons across samples with multiple datatypes, we observed that the accuracy of CNA inference is highest from SNP arrays, followed by whole-exome sequencing, with RNA-sequencing the least accurate. Similarly, the resolution of CNA segments derived from SNP array data is the highest, while it is more difficult to detect focal CNAs from whole-exome or RNA sequencing data. We observed high similarities between each PDX and its corresponding patient tumor, as well as between PDXs derived from the same model but separated by passaging. Finally, we analyzed genes with CNAs in each PDX to determine if they were affected by systematic evolutionary pressures from the mouse environment, while also using cell line drug sensitivity data to evaluate their functional impact on drug response. Our in-depth tracking of CNAs throughout PDX engraftment and passaging elucidates how PDX models are representative of the patient tumors and their suitability for pre-clinical drug testing.

Evolutionary Insights from Metagenonics

SMBE-OR-171

GraphToTree: a novel approach based on sequence similarity networks to infer phylogenies for large scale data

G. Bernard [1,*], T. Lazard [1], M. C. Macey [2], M. Fox-Powell [3], C. R. Cousins [3], K. Olsson-Francis [2], P. Lopez [1], E. Bapteste [1]

[1]Sorbonne Universités, Paris, France, [2]The Open University, Milton Keynes, [3]University of St Andrews, St Andrews, United Kingdom

Abstract: Alternative approaches to the MSA-based methods, such as sequence similarity networks (SSN) and alignment-free (AF) methods, have been increasingly used in evolutionary analyses to cope with the deluge of data generated by high-throughput sequencing. These latter approaches are faster and more scalable than their MSA-based counterpart, and can be applied to a broader range of data (sequencing reads, metagenomic data, etc). Here we present a novel approach combining the SSN and the AF methods to quickly identify gene/proteins of interest in metagenomic data, to remove potential problematic sequences/taxa and to infer proxies of phylogenies when the data are too large or divergent to perform a MSA. This approach is as accurate as the MSA-based approach across a range of evolutionary scenario while being faster and more scalable. Moreover, we also introduce a novel method to reconstruct a phylogenetic tree, directly from a SSN, largely unaffected by the long branch attraction (LBA) problem. By considering an SSN corresponding to a gene family, we use a recursive grouping approach to infer the underlying phylogenetic tree of that gene family. This novel method, although less accurate than the classical methods, is more robust to the LBA than the MSA-based and AF methods in all the scenarios tested. Finally, we will present a very large phylogenetic tree of more than a hundred thousand environmental 16S genes (from Colour Peak in the Canadian high Artic) inferred using this novel approach, to get a global picture of the phylogenetic diversity within an extreme environment.

Evolutionary Insights from Metagenonics

SMBE-OR-173

Real-time analyses of extracellular DNA and horizontal gene transfer in microbial communities: biofilms as a model for pangenome evolution via nanopore sequencing.

J. Parker*

Abstract: Biofilms - (single- or multispecies communities of sessile microbial organisms) are present in all major environments on earth including marine, terrestrial and aquatic systems. They are also found ubiquitously throughout human-controlled environments including water treatment systems, bioreactors, hospital surfaces, and medical implants, as well as on mucosae of human patients in a range of conditions including cyctic fibrosis (CF) and urinary tract infections (UTI). The biofilm phenotype is throught to be a major driver of harms including antimicrobial resistance (AMR) and tolerance in human and animal hosts, billion-dollar losses in industrial systems, and fouling in shipping. Unusually, biofilms contain substantial amounts of extracellular DNA (here, 'ecDNA'). Unlike environmental DNA (eDNA), which is largely released into the environment by stochastic processes of excretion, death, and degredation, ecDNA appears to have a functional role in stabilising the biofilm structure. It may also mediate horizontal gene transfer (HGT). This could includes the potential for genotypic carry-over, where genetic material can be transferred through time between previous and future inhabitants of the biofilm matrix even if the biofilm has been sterilized in between.

Given the polyspecies nature of the biofilm community and the potential for HGT, in the presence of strong shaping selection pressures such as AMR, biofilms can constitute a particularly interesting system for investigating microbial evolution through the model of pangenome composition and evolution. The presence of ecDNA-mediated HGT in biofilms is potentially of significance for hospital-acquired infection and AMR control, and in clinical settings the speed of time-to-information (from sample collection, through DNA sequencing and analysis, to presentation of clinically actionable results) is critical.

We will present results from clinical metagenome samples sequenced using nanopore sequencers, derived from CF and UTI patients with a variety of clinical presentations and biofilm phenotypes, to determine the biofilm pangenome, backed up by culture and microscopy results. We will detail progress made in identifying HGT modes and rates, and examine the 'carry-over' hypothesis for ecDNA. Finally we will report progress in developing real-time identification and analysis pipelines for clinical decision-making at University Hospitals Southampton, and suggest priority areas for future work.

Although biofilms are of major clinical and industrial importance, their significance to microbial ecology and evolution has only recently been appreciated. We hope that interdisciplinary communication between biofilms microbiologists and evolutionary biologists will grow once biofilms' unique status as a system for microbial evolution is appreciated.

Origins, evolution and function of novel genes
SMBE-OR-207

De novo gene evolution in maize

U. Singh [1], Z. Arendese [1], M. Hufford [1], J. Walley [1], E. S. Wurtele [1,*]
[1]Iowa State University, Ames, United States

Abstract: The premise that new genes can arise from non-genic DNA sequences, borne out from massive sequencing data, sharply contrasts with the long- accepted view that novel gene functions primarily arise from a slow process of accumulated mutations and rearrangements of already-established genes. We hypothesize that a major role of de novo genes is to enable evolutionary adaptation to new environments. Here we focus on the de novo genes from 25 lines of Zea mays (the NAM lines) that are representative of the remarkable diversity of this organism. Evidence of translation is provided from proteomics data. We infer de novo genes and gene candidates that are common to all 25 lines, and those that are unique particular clades, and use RNA-Seq data from >4000 samples from diverse environmental, genetic, and developmental conditions to identify the contexts under which each gene is expressed. We use micro-synteny across the maize lines to infer the potential evolutionary processes that likely gave rise to each de novo gene. For the analyses, we use three software tools recently developed in our group: 1) Phylostratr, R-software for automated phylostratigraphic analysis, 2) fagin, R-software for syntenic inference, and 3) MetaOmGraph (MOG), a Java tool for interactive statistical analysis and visualization of high-throughput data (tested for 10,000 RNA-Seq runs x 100,000 transcripts). All are free, publicly-available software that can be applied to study of any organism. to particular clades.

Origins, evolution and function of novel genes
SMBE-OR-208
The role of novel genes in the origins of the Animal and Plant Kingdoms
J. Paps*

Abstract: The history of life on Earth comprises major evolutionary transitions, the emergence of new features that dramatically changed the biology of organisms and the planet. Understanding the genomic basis of these events, such as the origin of animals or plants, is one of the major challenges of modern evolutionary biology. Using extensive genome comparisons, we infer the gene complement of the genomes of the ancestors of major groups or organisms.
These analyses uncover an unprecedented increase in the extent of genomic novelty during the origin of the Animal Kingdom and land plants, and identify new genes essential to their biology. However, we also show a previously unsuspected role for gene loss which has shaped the genomes of metazoans after their emergence. Together, these results shed new light on the role that the dynamic nature of genomes played during these transitions.

Novel insights into evolutionary genetics from emerging technologies
SMBE-OR-241A
Toward a complete cell genealogy tree of complex organisms
K. Liu*, C. Ye, X. He

Abstract: Mapping the cell genealogy of a complex multicellular organism relies on somatic genetic alterations accumulated from zygote to adult. Given the huge cell number of an individual, available methods can only track a tiny fraction of the developmental lineages which provide a low-resolution cell genealogy tree for organisms such as fly, zebrafish, and mouse. Combined rational design and directed evolution, we here developed a highly efficient genealogy mapping system which comprises an optimized mutation-generating protein that targets a designed DNA fragment as single cell readout. We applied the system to *Drosophila melanogaster* and observed increasing mutations on a three kilobase-pair readout sequence from stages of embryo, larva, pupa, to adult. There are on average 12 substitution mutations accumulated on the readout sequence in early-adult cells. Using the mutations we constructed two single-cell-resolution developmental trees each comprising over 10,000 cell lineages from diverse tissues and supported by strong bootstrapping values. Analysis of the trees revealed a simple statistical model that involves no specialized oligopotent intermediates for describing the global pattern of cell fate commitments from multipotency to unipotency.

Novel insights into evolutionary genetics from emerging technologies
SMBE-OR-241B

Nanopore sequencing of the Acanthamoeba castellanii genome: toward a model for eukaryote lateral gene transfer

M. J. Colp [1,*], B. A. Curtis [1], J. M. Archibald [1]
[1]Biochemistry & Molecular Biology, Dalhousie University, Halifax, Canada

Abstract: Lateral gene transfer (LGT) is well recognized as an important evolutionary force in prokaryotes. Antibiotic resistance and virulence genes have spread among bacterial taxa by LGT, and understanding its frequency and mechanisms can provide insight into the evolution of bacterial pathogenicity. Although LGT has also been proposed to play a role in eukaryotic evolution, the evidence comes largely from comparative genomic investigations; there is little or no experimental evidence for the occurrence, frequency, and mode(s) of eukaryotic LGT. To better understand the importance of lateral gene transfer in eukaryotic organisms, the amoebozoan protist *Acanthamoeba castellanii* is being developed as a model organism. First, we have used Oxford Nanopore sequencing to produce a high quality reference genome sequence to serve as a baseline for studying genome evolution. This long-read technology has dramatically improved the contiguity of the 44.9 Mbp *A. castellanii* genome (N50=1.0 Mbp) relative to the publicly available reference sequences. The genomes of genetically transformed amoebae have also been sequenced in order to better understand transgene integration in this system, and suggest that foreign genetic material integrates preferentially in sub-telomeric regions. Finally, experiments designed to replicate lateral gene transfer in the laboratory have been initiated in order to experimentally characterize the process. All things considered, nanopore long-read sequencing technology is a powerful tool for studying eukaryote genome evolution.

Open Symposium
SMBE-OR-244
The population genomic legacy of the second plague pandemic in Trondheim, Norway

S. Gopalakrishnan [1,*], I. Lundstrøm [1], S. S. Ebenesersdottir [2], L. van Dorp [3], P. Luisi [4], S. Piedrahita Abella [1], M. Martin [5], G. Turner-Walker [6], I. Moltke [1], M. Avila-Arcos [7], J. Samaniego Castruita [1], J. Schraiber [8], S. Rasmussen [1], D. Juan [9], P. Gelabert [9], L. Masatti [1], T. de-Dios [9], A. Margaryan [1], A. Fotakis [1], M. Orbegozo [1], Å. Vågene [1], A. Christophersen [5], H. Stenøien [5], F. Vieira [1], S. Liu [10], T. Günther [11], T. Kivisild [12], B. Skar [5], M. Sandoval Velasco [1], N. Wales [13], H. Schroeder [1], P. Campos [1], V. Guðmundsdóttir [2], T. Sicheritz-Ponten [1], B. Petersen [1], J. Halgunset [14], T. F. Higham [15], E. Willerslev [1], J. Krause [16], F. Balloux [3], T. Marques-Bonet [9], C. Lalueza-Fox [9], K. Stefansson [2], R. Nielsen [17], A. Helgasson [2], M. T. P. Gilbert [1 5]
[1]University of Copenhagen, Copenhagen, Denmark, [2]deCODE genetics, Reykjavik, Iceland, [3]UCL Genetics Institute, London, United Kingdom, [4]Universidad de Buenos Aires, Buenos Aires, Argentina, [5]Norwegian University of Science and Technology, Trondheim, Norway, [6]National Yunlin University of Science & Technology, Yun-Lin province, Taiwan, [7]Universidad Nacional Autonoma de Mexico, Queretaro, Mexico, [8]Temple University, Philadelphia, United States, [9]Institute of Evolutionary Biology, PRBB, Barcelona, Spain, [10]China National GeneBank, Shenzhen, China, [11]Uppsala University, Uppsala, Sweden, [12]KU Leuven, Leuven, Belgium, [13]University of York, York, United Kingdom, [14]St Olavs Hospital, Trondheim, Norway, [15]University of Oxford, Oxford, United Kingdom, [16]Max Planck Institute for the Science of Human History, Jena, Germany, [17]University of California, Berkeley, Berkeley, United States

Abstract: Human populations have been shaped by past catastrophes, some of which may have left long-lasting signatures in our genomes. The second plague pandemic represented the most important demographic collapse in historical Europe, with mortality rates estimated to possibly >50% of the population. It is widely assumed this must have significantly affected the genetic makeup of the populations. We explored its consequences on the Norwegian city of Trondheim, which like most European cities suffered a pandemic induced bottleneck, with estimates ranging from 50-70% reduction in census population size attributable to the second plague pandemic.
To explore the genomic effects of the second plague pandemic in Trondheim, we collected 54 genomes spanning 3 time periods, specifically 11 samples from before 1349 AD, 13 samples from the 16-18th centuries, and 30 samples from modern Trondheim. Using whole genome sequencing on these samples, we examined the global genome-wide changes in the ancestry composition of the population, and identified regions of the genome which showed radical changes in allele freuquencies through the period of the second plague pandemic.
Despite the relatively small sample sizes of our dataset, our results validate with the hypothesis that the second plague pandemic played a significant role in shaping the genomic diversity of Trondheim - something expected given the large estimated change in population size that occurred during this period. In particular, our data is compatible with a major decrease in migration from previous population sources (e.g. the British Isles) during this time, consistent with the decrease in Trondheim's political and economic importance. The average proportion of ancestry derived from British Isles decreases from ~30-40% in the pre-1349 samples, to under 5% in the post-1349 samples.
Additionally, we find allele frequency changes in genes that can be plausibly linked to resistance against bacterial infections. We note that it is impossible to pinpoint the pathogen *Yersinia pestis as the driver of these changes* - in theory other pathogens or even non pathogenic factors may have lead to the observed decreases in population diversity and shifts in allele frequencies at key genomic loci. Ultimately this leads us to propose that genetically encoded immune defense and ability to deal with septicemic states may have been a relevant factor for surviving bubonic plague.

Novel insights into evolutionary genetics from emerging technologies

SMBE-OR-244A

Population continuity of a Pre-Hispanic Civilization from Central Mexico

V. Villa-Islas [1,*], M. Sandoval-Velasco [2], E. Mejía Pérez-Campos [3], A. Herrera-Muñóz [3], G. Zepeda García-Moreno [4], K. Sandoval-Mendoza [5], M. A. Nieves-Colón [5 6], M. C. Ávila-Arcos [7]

[1]International Laboratory for Human Genome Research, National Autonomous University of Mexico, Querétaro, Mexico, [2]Center for GeoGenetics, Natural History Museum of Denmark, University of Copenhagen, Copenhagen, Denmark, [3]National Institute of Anthropology and History, Querétaro, [4]National Institute of Anthropology and History, Guanajuato, [5]National Laboratory of Genomics for Biodiversity (LANGEBIO-UGA), Centro de Investigación y de Estudios Avanzados del Instituto Politécnico Nacional, Irapuato, Mexico, [6]School of Human Evolution and Social Change, Arizona State University, Tempe, United States, [7]International Laboratory for Human Genome Research, National Autonomous University of Mexico, , Querétaro, Mexico

Abstract: Toluquilla and Ranas are two archeological sites located at Sierra Gorda, Querétaro, Central Mexico. Interesting aspects about the culture that inhabited these sites are a larger occupation span compared to other major Pre-Columbian societies (300 BCE to 1,500 CE) and that they were located at the limits of two large cultural regions: Aridoamerica and Mesoamerica. At the end of the XI century, the climate underwent drastic fluctuations affecting the agricultural resources and inducing a shift in the perimeters of these cultural regions, leading to a hypothesized population replacement of the Sierra Gorda inhabitants by hunter-gatherers from Aridoamerica. However, Toluquilla site displays cultural continuity suggesting that this climatic phenomenon did not affect this side of Sierra Gorda with the same harshness compared to the central highlands. To assess both scenarios and gain insights into the population history and fine-scale regional demographic events of the inhabitants of Sierra Gorda region, we generated paleogenomic data from human remains of different time points spanning the time of the alleged replacement. We recovered complete mitochondrial genomes and low-coverage nuclear data for 9 individuals. When comparing these to paleogenomic data of 4 individuals from a Pre-Hispanic site at the neighboring State of Guanajuato, as well as to genome-wide data from present-day Native Mexicans, we identified a clear genetic affinity to present-day populations from Central-West Mexico and similar genetic structure for both time periods. Suggesting a genetic continuity between both periods leading to interesting insights about adaptation strategies by these population in response to drastic climate shifts.

Novel insights into evolutionary genetics from emerging technologies
SMBE-OR-244B

Mitochondrial mutational spectrum as an universal marker of cellular longevity

K. Popadin [1,*], A. Mikhaylova [2], A. Mikhaylova [3], K. Ushakova [3], E. Tretyakov [4], A. Iurchenko [5], M. Zazhytska [1], D. Knorre [6], Z. Fleischmann [7], S. Annis [7], M. Franco [7], K. Wasko [7], W. Kunz [8], I. Mazunin [3], S. Nikolaev [5], A. Reymond [1], K. Khrapko [7], K. Gunbin [9]

[1]Center for Integrative Genomics, University of Lausanne, Lausanne, Switzerland, [2]Center for Mitochondrial Functional Genomics, Immanuel Kant Baltic Federal University, [3]Center for Mitochondrial Functional Genomics, Immanuel Kant Baltic Federal University, Kaliningrad, Russian Federation, [4]medical university of vienna, Vienna, Austria, [5]Gustave Roussy Cancer Center, Paris, France, [6]The A.N. Belozersky Institute Of Physico-Chemical Biology, Moscow, Russian Federation, [7]Northeastern University, Boston, United States, [8]University of Bonn Medical Center, Bonn, Germany, [9]The Institute of Cytology and Genetics of the SB RAS, Novosibirsk, Russian Federation

Abstract: It has been shown recently that mtDNA somatic variants due to their high mutation rate and high copy number per cell are very informative to trace cellular lineages in our body (Ludwig et al. 2019). Here we extend this statement and demonstrate that mtDNA variants can be interpreted not only as neutral markers of cell divisions but the relative frequency of different mtDNA substitutions (i.e. mtDNA mutational spectrum) can inform us about important biological properties of these cells such as longevity.

First, using a collection of more than 2000 somatic mtDNA mutations from 25 human tissues (GTEx, Ludwig et al. 2019) we derived mutational spectrum for each tissue and tested if there is an association between mutation spectrum and tissue-specific cell longevity collected from literature (Tomasetti and Vogelstein 2015). Ratio of the most common transitions (T>C/G>A, light chain notation) demonstrated significant positive correlation with cell longevity (spearman's rho = 0.5791803, p = 0.0024, N = 25). This correlation was robust to effect of variant allele frequencies (VAF) and expression level of mitochondrial genes.

Second, to prove our results on independent dataset we analyzed more than 7000 somatic mtDNA mutations from human cancers (ICGC & TCGA: Yuan et al. 2017). During tumorigenesis cells are becoming more fast-dividing and thus we can use two predictors of the cancer cell longevity: (i) cell longevity of the ancestral tissue and (ii) the stage of tumorigenesis, approximated by the VAF (the lower the VAF, the more recent mutation, the faster cell divisions). Using multiple logistic regression we estimated a probability of T>C versus G>A as a function of the ancestral tissue type and VAF and obtained expected decrease in this probability with both: decrease in cell longevity (from tissue to tissue) and decrease in VAF.

Third, we hypothesized that, if universal, the discovered mutation bias may shape mutational spectra of different species with short- ('mice') and long- ('elephants') lived oocytes. Based on neutral mtDNA polymorphisms we reconstructed mutational spectra for 650 mammalian species and obtained that the best predictor of the species-specific generation length is the fraction of T>C substitutions (SMBE 2019 abstract: Mikhailova et al "Mitochondrial mutational spectrum in vertebrates is shaped by generation time") and exactly this process shapes the neutral nucleotide content of the whole mitochondrial genomes of short- versus long-lived mammals.

Altogether we demonstrated that mtDNA mutational spectrum is changing between species, tissues and during tumorigenesis and on all these scales T>C/G>A correlates with longevity. Taking into account extremely low VAFs of analyzed heteroplasmic mtDNA mutations in both human datasets (with median VAF < 5%) and presumed absence of selection on synonymous four-fold degenerate sites in mtDNA of mammals (Faith and Pollock 2003) we don't consider potential effect of selection and propose, that the observed changes in mutational spectrum are driven by mutagenesis. To improve the existed model of mtDNA mutagenesis (Faith and Pollock 2003) we propose that cell longevity might be associated either with mtDNA replication time or strength of chemical mutagens (aerobic conditions, pH) or mtDNA replication type (asynchronous vs, strand-coupled).

The described properties of mtDNA mutational spectrum can be used as a marker of cell longevity in single-cell, cancer and tissue-specific analyses

Novel insights into evolutionary genetics from emerging technologies
SMBE-OR-244C

Assessing the influence of regulatory landscape complexity on enhancer turnover and gene expression using structural variation

M. L. Benton [1,*], L. Han [2], CommonMind Consortium, D. M. Ruderfer [1 2 3], J. A. Capra [1 4]
[1]Department of Biomedical Informatics, Vanderbilt University, [2]Department of Medicine, [3]Department of Psychiatry, Vanderbilt University Medical Center, [4]Departments of Biological Sciences and Computer Science, Vanderbilt University, Nashville, United States

Abstract: Disrupted regulation of gene expression substantially contributes to the risk of many complex diseases. The vast majority of loci associated with disease in genome-wide association studies are in non-coding regions of the human genome, and many contribute to disease risk by disrupting or altering gene regulatory elements such as enhancers. Although there are numerous examples of enhancers that are mutated in disease, predicting whether mutations in a given enhancer will influence phenotype is still a difficult task. Current strategies for interpreting enhancer variation consider enhancers in isolation, despite evidence from mammals and insects that redundancy in enhancer landscapes buffers the phenotypic effects of enhancer loss on the expression of important genes. This project seeks to leverage novel whole genome sequencing data with functional genomic and evolutionary characterization to study the influence of structural variation on enhancer function within the broader enhancer landscape.

We hypothesized that the evolutionary history and complexity of the enhancer landscape of a gene influences its robustness to variation. We can test this using liver and brain enhancers identified across multiple mammalian species to quantify the relationship between evolutionary conservation of activity and landscape complexity. We found that gains of enhancer activity are more likely to occur in complex enhancer landscapes with higher gene density. Both the number of genes and the complexity of the landscape contribute significantly to the number of predicted enhancer gains. This model suggests that the gain of new enhancers is more likely to occur in regions with a larger number of gene targets and a greater amount of existing regulatory activity.

To test the utility of these relationships between regulatory landscape and robustness of the effects of enhancer mutations, we analyzed whole genome sequencing data and RNA-seq expression data from the cerebral cortex across 537 individuals. To focus on clear instances of enhancer gain and loss, we used structural variant (SV) data to identify the gain and loss of 79,056 cerebral cortex enhancers between individuals due to duplications and deletions of enhancer sequence. We found that deleting enhancers significantly decreased the levels of expression of their predicted target genes. The size of the effect was modulated by the number of enhancers deleted. We observed similar effects on gene expression when regulatory elements involved in maintenance of the 3D architecture of the genome, such as CTCF binding sites, were deleted. These results combined suggest that the regulation of genes by enhancers is influenced by both the complexity of the regulatory landscape in which they occur and the potential effects on the 3D interaction landscape of the region. This work integrates information about the genomic composition of enhancer landscapes across species and large-scale structural variation data linked to gene expression to provide insight into the functional impacts of enhancer gain and loss. Further study of the effect of enhancer alteration within the broader regulatory landscape will facilitate better interpretation of non-coding variants and perturbations to the gene regulatory architecture.

SMBE Editors symposium
SMBE-OR-245
A new approach to studying regulatory evolution
W.-H. Li [1,*]
[1]Department of Ecology and Evolution, University of Chicago, Chicago, United States

Abstract: In 1975 Mary-Claire King and Allan Wilson proposed that regulatory mutations, not protein sequence changes, account for the biological differences between humans and chimpanzees because their protein sequences are too similar. Ever since then regulatory evolution has been an interesting subject to evolutionists. In the past, however, it is difficult to obtain experimental data for studying regulatory evolution. Fortunately, recent advances in transcriptomics and cistromics have made it considerably simpler. In particular, the "DNA affinity purification sequencing (DAP-seq)" technique, an in vitro technique, is faster and cheaper technique than ChIP-seq, providing an affordable means to determine transcription factor (TF) binding sites (TFBSs). Moreover, my lab has developed computational methods for predicting TFBSs, for inferring the key DNA binding residues in a TF, and for constructing gene regulatory networks (GRNs). A combination of these computational methods and DAP-seq provides a powerful means to study regulatory differences between individuals or species. I shall present this new approach to the study of regulatory evolution.

Inside Africa: Uncovering patterns of human genetic diversity

SMBE-OR-246

Regulatory variation and selection in traditional populations of Southeast Asia

K. Bobowik [1,*], I. G. Romero [2], M. Cox [3], N. Banovich [4], H. Natri [4]

[1]Biosciences, University of Melbourne, Brunswick, [2]Biosciences, University of Melbourne, Parkville, Australia, [3]School of Fundamental Sciences, Massey, Palmerston North, New Zealand, [4]Integrated Cancer Genomics, Translational Genomics Research Institute, Phoenix, United States

Abstract: Our understanding of function within the human genome has been largely limited to populations of European ancestry. In order to better characterise human functional diversity, this study aims to analyse regulatory variation and differentially methylated regions in three isolated populations within the Indonesian archipelago. No study has yet analysed regulatory variation in Indonesia and the small, isolated nature of these populations makes them ideal to identify variants that have risen to higher frequencies. Furthermore, these populations face unique environmental pressures, particularly to malaria which is endemic in the region. Studying these populations can therefore give insight into different immune pathways that have arisen due to differing population histories and environmental pressures. We use methylation and transcriptomic data from over 100 individuals to explore the relationship between DNA methylation and gene expression levels between Indonesian islands. We identify differentially expressed genes and differentially methylated regions that correlate with changes in environmental variables, as well as pathways involved in the innate immune response.

Origins, evolution and function of novel genes

SMBE-OR-247

Genome plasticity in Papillomaviruses and de novo emergence of E5 oncogenes

A. Willemsen [1,*], M. Félez-Sánchez [2], I. G. Bravo [1]

[1]MIVEGEC, Centre National de la Recherche Scientique (CNRS), Montpellier, France, [2]Infections and Cancer Laboratory, Catalan Institute of Oncology (ICO), Barcelona, Spain

Abstract: While most papillomaviruses (PVs) are part of a healthy skin microbiota, only a handful of PVs are associated to malignant transformations linked to the specific activities of the *E5*, *E6* and *E7* oncogenes. The functions and origin of *E5* remain to be elucidated. We have computationally assessed whether the *E5* ORFs have a common origin and whether they display the properties of a genuine gene. Our results suggest that at least four events lead to the presence of a long non-coding DNA stretch between the early and late viral gene cassettes. In three of these events, the novel regions evolved coding capacity, becoming the extant *E5* ORFs. We then focused on the evolution of the *E5* genes in *AlphaPVs* infecting primates. The sharp match between the type of *E5* protein encoded in *AlphaPVs* and the infection phenotype supports the role of E5 in the differential oncogenic potential of these PVs. In our analyses, the best-supported scenario is that the five types of extant E5 proteins within the *AlphaPV* genomes do not have a common ancestor. Our evolutionary interpretation is that an originally non-coding region entered the genome of the ancestral *AlphaPVs*. This genetic novelty allowed to explore novel transcription potential, triggering an adaptive radiation that yielded three main viral lineages encoding for different E5 proteins, displaying distinct infection phenotypes. Overall, our results provide an evolutionary scenario for the *de novo* emergence of viral genes and illustrate the impact of such genotypic novelty in the phenotypic diversity of the viral infections.

Inside Africa: Uncovering patterns of human genetic diversity
SMBE-OR-248
Whole genome sequence analysis of the Aeta, a First Sundaland people from the Philippines
T. A. Jinam [1,*], N. Saitou [1]
[1]Population Genetics Laboratory, NATIONAL INSTITUTE OF GENETICS, Mishima, Japan

Abstract: The First Sundaland people (negritos) of Southeast Asia encompasses various groups from Peninsular Malaysia, Andaman Islands and the Philippines who share common phenotypes including dark skin, frizzy hair, and short stature. We sequenced the whole genomes of 10 Aeta individuals (Philippine negritos), whose samples were collected in 1985 by Keiichi Omoto, at high coverage. Comparison with other published genomes reveal that these Aeta individuals had substantially higher proportion of archaic human introgressed regions. Their effective population size showed signals of gradual decrease until approximately 5 thousand years ago (kya), and they split from other Southeast Asians approximately 20 kya. A more in-depth analysis of these genomes will hopefully shed more light into the complexities of human movements out of Africa.

Evolution of proteins and molecular pathways/networks under higher level selective effects
SMBE-OR-249

Rewiring of Molecular Networks Following Loss of a Subfunctionalized Ohnolog Node

D. Gray [1,*], J. Coulombe [1], S. Zachariah [1], V. Saxena [2], M. Ekker [2]
[1]Centre for Cancer Therapeutics, Ottawa Hospital Research Institute, [2]Department of Biology, University of Ottawa, Ottawa, Canada

Abstract: USP4 and USP15 are deubiquitinating enzymes encoded by genes that we have previously shown to be ohnologs arising from whole genome duplication events early in vertebrate evolution. USP4 and USP15 are known to regulate the stability of critical substrates in multiple molecular pathways directing cell differentiation, cell fate, DNA repair, and innate immunity. At the sequence level the enzymes remain quite similar but there has been some subfunctionalization over the course of evolutionary history such that the two enzymes operate in the same pathways but have distinct substrates in each pathway. With multiple interaction partners USP4 and USP15 can be considered network nodes in these pathways; given the nature of the pathways it seems likely that selection occurs at multiple levels (in the case of innate immunity one could, for example, argue for group selection through herd immunity). To determine the extent of functional redundancy in USP4 and USP15 we have analyzed progeny from genetic crosses of mice in which these genes have been inactivated. We have established that embryos null for both genes die in midgestation from an apparent failure of hematopoiesis, which occurs in the liver at this developmental stage (embryos with the compound null genotype have remnant livers containing few hematopoietic cells). Fetal hematopoiesis is reliant on signalling through the Wnt/β-catenin and TGFβ pathways, both of which are replete with USP4 and USP15 substrates; we are currently analyzing components of these pathways to determine how their levels correlate with mouse genotypes. Phenotypes have been revealed in mice null for only one of the two genes (for example increased susceptibility to virus infection in USP4 null mice), but the viability of such mice indicates that despite subfunctionalization when one ohnolog is inactivated there can be at least partial compensation by the other. This seems remarkable given the multiplicity of pathways in which these enzymes operate. It is all the more remarkable that after 500 million years of subfunctionalization there would be no network disruption and selective cost associated with the loss of one of the ohnologs , but there are species in which gene inactivation has occurred through disruption of the coding sequence of USP4 or USP15. Using CRISPR technology we have completely excised the USP4 or USP15 genes of zebrafish, a species in which USP15 appears to have been naturally inactivated (a very recent event in teleost evolution). By comparing the status of network components in zebrafish and mice we hope to determine how network rewiring can occur when an ohnolog is lost either naturally (as in zebrafish USP15) or artificially (as in our engineered strains).

Evolution of phenotypes: understanding diversity and the role of plasticity in adaptation to new environments
SMBE-OR-250

Ancient balancing selection maintains incompatible versions of a conserved metabolic pathway in yeast

J. Boocock [1,*], J. Bloom [1], M. Sadhu [2], L. Kruglyak [1]
[1]Department of Human Genetics, University of California, Los Angeles, Los Angeles, [2]Genetic Disease Research Branch, National Human Genome Research Institute, NIH, Bethesda, Maryland, United States

Abstract: Accumulation of genetic incompatibilities can lead to speciation, but there are few detailed examples of this process in action. We discovered that the classical galactose pathway of the yeast *Saccharomyces cerevisiae* exists in two incompatible states maintained by ancient balancing selection. In a mapping study of growth on galactose, we identified a genetic interaction among three unlinked loci in crosses involving the strain CBS2888. All three loci contained components of the galactose pathway: the galactose transporter (*GAL2*), the GAL cluster (*GAL1*, *GAL10*, *GAL7*) encoding the enzymes which convert galactose into glucose-6-phosphate, and phosphoglucomutase 1 (*PGM1*), an enzyme that converts glucose-6-phosphate to glucose-1-phosphate—the substrate for glycolysis. We engineered strains carrying all eight allelic combinations at these loci and showed that the non-CBS2888 version of *PGM1* is incompatible with CBS2888 alleles of the other genes. The CBS2888 alleles appear to have diverged from those of most other *S. cerevisiae* strains before the birth of the *Saccharomyces sensu stricto* species cluster ~10-20 MYA. Globally, the CBS2888 alleles are found in isolates from galactose-rich environments, such as cheese, kefir, and milk. Population genetics analysis of linked neutral sites revealed a strong signature of ancient balancing selection. Strains with the CBS2888 alleles grew faster in galactose, but slower in glucose, revealing a tradeoff on which this balancing selection may act. Our work shows that balancing selection of alternative versions of a metabolic process can maintain incompatible unlinked alleles for millions of years.

Insights from fitness landscapes into evolutionary pathways
SMBE-OR-251
Environmental pleiotropy and demographic history direct adaptation under antibiotic selection
D. Gifford[*], R. Krašovec[1], C. G. Knight[1]
[1]University of Manchester, Manchester, United Kingdom

Abstract: Evolutionary rescue following environmental change requires mutations permitting population growth in the new environment. If change is severe enough to prevent most of the population reproducing, rescue becomes reliant on mutations already present. If change is sustained, the fitness landscapes of both environments, and how they are associated—termed 'environmental pleiotropy'—may determine which alleles are ultimately favoured. A population's demographic history—its size over time—influences the variation present. Although demographic history is known to affect the probability of evolutionary rescue, how it interacts with environmental pleiotropy during severe and sustained environmental change remains unexplored. Here, we demonstrate how these factors interact during antibiotic resistance evolution, a key example of evolutionary rescue fuelled by pre-existing mutations with pleiotropic fitness effects. We combine published data with novel simulations to characterise environmental pleiotropy and its effects on resistance evolution under different demographic histories. Comparisons among resistance alleles typically revealed no correlation for fitness—i.e., neutral pleiotropy—above and below the sensitive strain's minimum inhibitory concentration. Resistance allele frequency following experimental evolution showed opposing correlations with their fitness effects in the presence and absence of antibiotic. Simulations demonstrated that effects of environmental pleiotropy on allele frequencies depended on demographic history. At the population level, the major influence of environmental pleiotropy was on mean fitness, rather than the probability of evolutionary rescue or diversity. Our work suggests that determining both environmental pleiotropy and demographic history is critical for predicting resistance evolution, and we discuss the practicalities of this during in vivo evolution.

Insights from fitness landscapes into evolutionary pathways

SMBE-OR-252

Mechanistic Constraints on Genotype-Phenotype Mapping

M. Lagator [1,*], T. Paixao [2], J. P. Bollback [3], C. C. Guet [2]

[1]University of Manchester, Manchester, United Kingdom, [2]IST Austria, Vienna, Austria, [3]University of Liverpool, Liverpool, United Kingdom

Abstract: Existing molecular mechanisms in cells constrain the evolutionary trajectories accessible to a population under selection. One way in which molecular mechanisms constrain evolutionary pathways is by shaping the patterns of epistasis (interactions between mutations). Epistatic interactions determine the shape of the adaptive landscape, and in doing so impact how rapidly and effectively a population evolves by defining how genotype maps onto the phenotype. Yet, in spite of the importance of epistasis for determining evolutionary outcomes, we lack the ability to predict interactions between mutation *de novo*, and are therefore limited to observing and documenting epistasis through labor-intensive experiments. I will show that studying how molecular mechanisms constrain mutational effects enables predictive understanding of epistasis. Specifically, I will focus on how the mechanisms of protein-DNA binding shape the patterns of epistasis between mutations in prokaryotic promoters. A general thermodynamic framework for gene regulation, based on a biophysical understanding of protein-DNA binding, accurately predicts the sign of epistasis in a canonical *cis*-regulatory element consisting of overlapping RNA polymerase and repressor binding sites. Sign and magnitude of individual mutation effects are sufficient to predict the sign of epistasis and its environmental dependence. These results demonstrate how understanding the constraints imposed by molecular mechanisms allow predictions of complex phenomena (such as epistasis) from first principles.

The Causes of parallel molecular evolution

SMBE-OR-253

A high-resolution view of repeatability in asexual populations evolved in a fluctuating environment using lineage tracking

A. Rego-Costa [1,*], A. Nguyen Ba [1], I. Cvijović [2 3 4], M. Desai [1 3 4 5]

[1]Department of Organismic and Evolutionary Biology, [2]Graduate Program in Systems Biology, [3]NSF-Simons Center for Mathematical and Statistical Analysis of Biology, [4]Quantitative Biology Initiative, [5]Department of Physics, Harvard University, Cambridge MA 02138, United States

Abstract: Parallel evolution is constrained by the interplay of selection and drift, which determine the fate of mutations already segregating in a population. While these dynamics are well understood for a single selected mutation, we lack a statistical description in the biologically-relevant case of genetically diverse populations. Moreover, the low frequency-resolution of previous experimental methods prevented appropriate accounting of selected rare mutations. Here we used a new barcode-based long-term lineage tracking method to observe the fate of naturally accumulated segregating beneficial mutations in replicate asexually evolving yeast populations. A clonal source population was first evolved for 500 generations to accumulate standing genetic variation in an environment fluctuating between YPD (rich medium) and YPDA (YPD, acetic acid). It was then split into 24 replicate populations that were evolved for another 400 generations in either of three different passage conditions: constant YPD, constant YPDA, and fluctuating YPD/YPDA. While mutation frequency trajectories were very repeatable in constant YPD, outcomes diverged within the other conditions. Pleiotropic effects are key: while all populations shared the same mutations, subtle differences in fitness distributions among passage conditions led to the fate of individual mutations to depend more or less on that of the others. Most notably, fluctuations in the establishment time of rare large-effect mutations have a disproportionate effect on the repeatability of the population-wide dynamics. Our experimental work exposes the fine-scale dynamics of parallel evolution in the case of shared ancestral mutations.

The Causes of parallel molecular evolution

SMBE-OR-254

A strong selection, strong mutation regime drives predictable copy number variant dynamics in the early stages of adaptive evolution

G. Avecilla [1,*], S. Lauer, P. Spealman, D. Gresham

[1]New York University, New York, United States

Abstract: Copy number variation (CNV) is a common but complex class of mutations with important roles in evolution, adaptation, and disease. CNVs mutate at a different rate than single nucleotide variants (SNVs), and how this contributes to evolutionary dynamics in unique ways is largely unknown. We studied the general amino acid permease *GAP1* in *Saccharomyces cerevisiae,* using a fluorescent CNV reporter that allows examination of CNV dynamics in evolving populations. We constructed a barcoded lineage tracking library in a *GAP1* CNV reporter strain, and performed experimental evolution in glutamine-limited chemostats. We found that CNVs are repeatedly selected and rise to high frequency in early evolution, and these dynamics are driven by hundreds of CNV lineages. We isolated *GAP1* CNV clones, and found that the majority of these clones have CNVs which are formed by non-allelic homologous recombination between two long terminal repeats which flank *GAP1*. Additionally we observed some CNVs that vary in size and copy number, with structures that form as a result of diverse mechanisms. We hypothesized that the observed predictable increase in the CNV subpopulation results from a high mutation rate of *GAP1* CNVs, primarily driven by NAHR. We tested this by simulating evolution experiments, and find the best fit mutation rates and distribution of fitness effects of *GAP1* CNVs and SNVs. We found that these parameters explain the observed CNV dynamics: a predictable, deterministic initial increase in CNV frequency due to high mutation rate, followed by stochastic dynamics characterized by clonal interference with rare high fitness SNVs.

The Causes of parallel molecular evolution

SMBE-OR-255

Parallel within-host evolution of RNA viruses during persistent infection and transmission

L. Ferretti [1,*]

[1]Big Data Institute, University of Oxford, Oxford, United Kingdom

Abstract: Rapidly evolving RNA viruses represent a natural model system to understand evolutionary biases and constraints. Due to a combination of high mutation rates, effective population sizes and short genomes, many populations of RNA viruses explore a large fraction of the local genotype space, therefore reducing the impact of mutational stochasticity. Their evolution is characterised by large effective selection coefficients which increases evolutionary predictability, but also by ever-changing fitness landscapes. These features makes them excellent candidates for studies of evolvability and evolutionary constraints. This is especially true for viruses leading to persistent infections.

In persistent or chronic infections, a viral population goes through cycles of infection, within-host evolution and transmission, with each cycle lasting much longer than a viral generation. Adaptation during transmission and intra-host growth occurs repeatedly, allowing for studies of parallelisms among multiple transmission and infection events. Understanding the constraints on such evolutionary processes for viral pathogens has relevant implications for public health.

Constraints in evolution are shaped by multiple forces - mutation, selection, epistasis and recombination - as well as by features of the infection cycle of the virus. We will present results on major human and animal pathogens that show evidence of parallel evolution from different forces at different steps along the life cycle of these viruses. Phylodynamic reconstruction of between- and within-host evolution of HIV across multiple transmission pairs reveals common trends in the mutational and selective features of successfully transmitted viral lineages and clarifies the extent of evolution for intra-host constraints on evolvability. We also take advantage of multiple infections and superinfections as natural experiments to explore constraints from competition and ecological niches in the intra-host dynamics of HIV. Finally, quasi-species structure, artificial inoculation experiments, deep sequencing of different animals and tissues and theoretical approaches are combined to reveal the extent of parallelism in the evolution of Foot-and-Mouth Disease Virus swarms persistently carried by African Buffaloes.

The Causes of parallel molecular evolution
SMBE-OR-256
The mosaic evolution of bacterial life cycles
J. Van Gestel*, M. Ackermann, A. Wagner

Abstract: Many bacteria are exposed to ever changing environments, to which they respond by adopting various lifestyles. Cells can for example disperse by swimming, form colonies, or become dormant. Here, we use a novel computational approach to study lifestyle transitions in the context of a bacterium's life cycle and, subsequently, examine how these transitions affect evolution. We employ machine learning to automatically reconstruct the complete life cycle of Bacillus subtilis from hundreds of previously acquired gene expression profiles. This yields a uniquely detailed timeline that shows how regulation changes in time and reveals the strong modular organization of the life cycle. By analyzing over 380 Bacillales genomes, we show that life cycle modularity gives rise to mosaic evolution, in which certain life stages – such as motility and sporulation – are conserved and lost as discrete units. We postulate that habitat changes, which make these life stages obsolete or detrimental, can explain this mosaic pattern. Indeed, when evolving eight distinct Bacillales strains and species in the laboratory, we observe rapid and parallel losses of the sporulation life stage across species. We conclude that a life cycle perspective is pivotal for understanding the causes and consequences of modularity in microbial evolution.

Barriers and drivers of evolutionary innovation by horizontal gene transfer
SMBE-OR-257
Diversity and evolution of extracellular polysaccharides in enteric bacteria
R. Mostowy*, F. Lassalle, K. Wyres, R. Wick, K. Holt

Abstract: Bacterial species are often surrounded by extracellular polysaccharides (EPs), like capsules and lipopolysaccharides. These diverse molecules play many important roles in helping bacteria adapt to and persist in different environments. Recent advances in our understanding of how these structures are built and synthesised have elucidated the diversity-generating potential of the EP synthesis genetic loci. However, our understanding of how EP loci evolve across different bacterial species, genera and families is still far from complete.
Here, we tackled this question by analysing 27,365 genomes of the bacterial family Enterobacteriaceae, including *Escherichia coli*, *Salmonella enterica* and *Klebsiella pneumoniae*. Using comparative genomics and rapid distance-estimation techniques, we analysed the diversity and distribution of genes in two major EP locus regions in 45 bacterial genera: *cps*, typically synthesizing group I and IV capsules (usually comprising *galF/gnd/ugd* genes each), and *kps*, typically synthesizing group II and III capsules (usually comprising *kpsT/kpsT/kpsC/kpsS* genes each). We identified 2442 distinct locus variants in the *cps* region and 296 variants in the *kps* region. The EP loci were highly plastic and dynamic genetic entities, exhibiting high rates of horizontal DNA transfer. First, locus sharing was observed not just between different bacterial lineages within a species, but also between species and genera. While many of these instances were a result of recent locus exchange via recombination (700 unique instances of exchange between enteric species in *cps* and 19 cases in *kps*), there were multiple examples of selective locus maintenance through vertical inheritance, presumably driven by ecological factors. Second, genes identified in EP loci were under weaker purifying selection than other bacterial genes (higher proportion of non-synonymous substitutions in locus regions compared to other genes, $p<10^{-16}$). Third, evolution of EP loci was driven by gene gain/loss dynamics. Individual EP locus genes were highly mobile, emphasising the `plug-and-play' nature of EP loci, with some gene families more likely to co-evolve together (e.g., housekeeping locus genes or biosynthetic gene clusters). Finally, the EP locus diversity was not correlated with the diversity of genomic backbones and instead was best predicted by the number of sampled isolates. We can thus anticipate large, yet unobserved repertoires of novel EP diversity in enteric bacteria.
Overall, these results demonstrate the enormous adaptive potential of bacterial EPs by their ability to rapidly generate new types, enabling adaptation to novel ecological niches via horizontal transfer of DNA.

Barriers and drivers of evolutionary innovation by horizontal gene transfer

SMBE-OR-258

Emergence of plasmid stability under non-selective conditions maintains antibiotic resistance

T. Wein [1,*], N. F. Hülter [1], I. Mizrahi [2], T. Dagan [1]

[1]Institute of Microbiology, Kiel University, Kiel , Germany, [2]The Department of Life Sciences & the National Institute for Biotechnology in the Negev, Ben-Gurion University of the Negev, Beer Sheva , Israel

Abstract: Plasmid acquisition is an important mechanism of rapid adaptation and niche expansion in prokaryotes. Positive selection for plasmid-encoded function is a major driver of plasmid evolution within populations while plasmids that do not confer a selective advantage are considered costly and expected to go extinct. Yet, plasmids are ubiquitous in nature and their persistence remains an evolutionary paradox. Here we demonstrate that non-mobile plasmids persist over evolutionary timescales without selection for the plasmid function. Evolving a minimal plasmid encoding for antibiotics resistance in *Escherichia coli* we discover that plasmid stability emerged in the absence of antibiotics, and that plasmid loss was determined by transcription-replication conflicts induced by the resistance gene of the plasmid. We further find that environmental conditions modulated these conflicts and led to plasmid resistance persistence. Silencing the transcriptional load of the resistance gene resulted in stable plasmids that became fixed in the bacterial lineage. Thus, the emergence of plasmid stability under non-selective conditions provides an evolutionary explanation for the ubiquity of plasmids in nature and may further explain the persistence of antibiotic resistance in microbial populations.

Biochemistry, epistasis and the evolutionary process

SMBE-OR-259

Adaptive changes in hemoglobin function in high-altitude Tibetan canids were derived via successive lateral transfers involving interparalog gene conversion and introgressive hybridization

A. Signore [1,*], Y.-Z. Yang [2], Q.-Y. Yang [2], G. Qin [2], H. Moriyama [1], R.-L. Ge [2], J. Storz [1]

[1]School of Biological Sciences, University of Nebraska, Lincoln, United States, [2]Research Center for High Altitude Medicine, Qinghai University, Xining, China

Abstract: A key question in evolutionary biology concerns the relative importance of different sources of adaptive genetic variation, such as *de novo* mutations, standing variation, and introgressive hybridization. A corollary question concerns how allelic variants derived from these different sources may influence the molecular basis of phenotypic adaptation. Here we use a protein-engineering approach to examine the phenotypic effect of putatively adaptive hemoglobin (Hb) mutations in the high-altitude Tibetan wolf that were selectively introgressed into the Tibetan mastiff, a high-altitude dog breed that is renowned for its hypoxia tolerance. Experiments revealed that the introgressed coding variants confer an increased Hb-O_2 affinity in conjunction with an enhanced Bohr effect. We also document that affinity-enhancing mutations in the β-globin gene of Tibetan wolf were originally derived via interparalog gene conversion from a tandemly linked β-globin pseudogene. Thus, affinity-enhancing mutations were introduced into the β-globin gene of Tibetan wolf via one form of intragenomic lateral transfer (ectopic gene conversion) and were subsequently introduced into the Tibetan mastiff genome via a second form of lateral transfer (introgression). Site-directed mutagenesis experiments revealed that the increased Hb–O_2 affinity requires a specific two-site combination of amino acid replacements, suggesting that the molecular underpinnings of Hb adaptation in Tibetan mastiff (involving mutations that arose in a nonexpressed gene and which originally fixed in Tibetan wolf) may be qualitatively distinct from functionally similar changes in protein function that could have evolved via sequential fixation of *de novo* mutations during the breed's relatively short duration of residency at high altitude.

Biochemistry, epistasis and the evolutionary process

SMBE-OR-260

The accuracy, bias, and precision of ancestral protein resurrection methods

D. Theobald [1,*], B. Beckett [1], M. Sennett [1]

[1]Biochemistry, Brandeis University, Waltham, United States

Abstract: Ancestral sequence reconstruction (ASR) provides unique experimental knowledge of the physical, chemical, and biological characteristics of both ancient and modern proteins — knowledge that is unattainable by other strategies. Widespread use of ASR has been enabled by the incredible growth of our sequence databases, the development of rigorous model-based phylogenetic methods, and the recent innovations in likelihood and Bayesian statistics. Despite these advances, the accuracy, precision, and bias of resurrected ancestral sequences is currently unknown. Are the most probable ancestral sequences systematically biased to have anomalous biophysical properties? How well do the biochemical properties of resurrected proteins recapitulate the properties of the true ancestral biomolecules? Which evolutionary models provide the most accurate ancestral reconstructions? These questions are extremely difficult to answer definitively because the real ancestral proteins are generally lost to history.

Even with a correct evolutionary model, systematic methodological biases can influence the observed properties of ASR sequences. In common practice, the uncertainty in the ancestral sequence is ignored, and the single most probable (SMP) sequence is chosen as the definitive ancestor and studied experimentally. It is known theoretically that the SMP ancestral sequence may result in a protein with extremely biased biophysical properties that are not representative of the ancestral protein. To give one well-known example of a potential bias, ancestrally resurrected proteins are often much more thermostable than their modern descendants, and it is currently controversial whether the observed high thermostability is a bona fide property of ancient proteins or rather is a methodological artifact.

ASR has been criticized because of the inherent difficulty in rigorously validating and benchmarking its methods. In order to know if ASR methods are reliable, we have developed a general method to benchmark ASR inferences of real biological systems against the properties of known biomolecules. We have reconstructed nodes in the phylogeny of Apicomplexan malate and lactate dehydrogenases, characterized their biophysical properties (molecular weight, extinction coefficient at 280nm, kcat, Km, log(kcat/KM), Tm, and ΔH of folding), and evaluated their bias, precision, and accuracy against known standards. Preliminary results indicate for all reconstructed nodes, there is no significant bias in any global property of the SMP reconstruction. Our data confirm the use of the single most probable ancestor as an accurate representative of the true ancestor at any given ancestral node, within the precision of the methodology as determined from sampled sequences.

Genetic conflicts in molecular evolution

SMBE-OR-261

Sex-ratio meiotic drive shapes the evolution of the Y chromosome in Drosophila simulans.

Q. Helleu [1], C. Courret [2,*], D. Ogereau [2], K. Burnham [3], S. Aulard [2], N. Chaminade [2], C. Montchamp-Moreau [2]

[1]Department of Ecology and Evolution, University of Lausanne, Lausanne, Switzerland, [2]CNRS, Gif-sur-Yvette, France, [3]Wellcome Centre for Human Genetics,, University of Oxford, Oxford, United Kingdom

Abstract: Meiotic drivers are selfish genetic elements that promote their own transmission into the gametes, thus triggering intragenomic conflicts. In the Paris *sex-ratio* system of *Drosophila simulans*, X-linked drivers prevent the segregation of the heterochromatic Y chromosome during meiosis II, and hence the production of Y-bearing sperm. These drivers, which cause a bias toward females in the progeny of carrier males, are currently spreading across the species range. Natural selection, which tends to restore equal sex ratio, has favored the emergence of resistant Y chromosomes and autosomal suppressors, thus making the distorters cryptic or nearly in the wild. We present here recent findings on drive resistance.

We investigated gene variation among 351 Y chromosomes from 29 population samples collected over more than 20 years and showing a wide continuum of phenotypes, from sensitivity to complete resistance. We identified only three haplotypes. One of them is associated with resistance and proved able to replace sensitive Ys within an handful of years in populations invaded by the drivers, showing that intragenomic conflicts can drive astonishingly rapidly the evolution of Y chromosomes. *In situ* hybridization with satellite DNA probes revealed extensive structural variation, suggesting that repeated sequences are rapidly evolving and may account for the continuum of resistant phenotypes.

Genome-wide methods for detecting selection

SMBE-OR-262

A new test for balancing selection using linkage disequilibrium, applied in humans and mosquitoes to scan for malaria-relevant genes

J. Tennessen [1,*], D. Neafsey [1], M. Duraisingh [1]

[1]Harvard T.H. Chan School of Public Health, Boston, United States

Abstract: Genome-wide scans for selection are now routine in many species, but most successful implementations have been for positive selection rather than balancing selection. Existing methods to detect balancing selection often perform poorly or give conflicting results. Nevertheless, balancing selection is likely to be common, and inferring it with population genomic data is a promising way to identify phenotypically-relevant adaptive genetic diversity. Balancing selection can yield a striking pattern of two or more exceptionally divergent haplotypes at a locus. Kelly's ZnS statistic is designed to pinpoint such clusters of adjacent polymorphisms in high linkage disequilibrium. However, scans for balancing selection rarely incorporate linkage disequilibrium, in part because recombination, rare variants, and phasing ambiguity can reduce the power of statistics like ZnS. Here, we modify the ZnS framework to detect narrow genomic regions showing both high linkage disequilibrium and high nucleotide diversity relative to interspecies divergence. This improved method performs well both in simulations and for known empirical balancing selection targets. We apply this method to large, phased population genomic datasets in both humans and mosquitoes, with a focus on identifying polymorphisms that could impact resistance to infection by malaria parasites. By refining established tests for balancing selection, and ground-truthing them with empirically-validated cases of adaptive polymorphisms, we can distinguish functionally important genetic variants from selectively neutral ones.

Genome-wide methods for detecting selection

SMBE-OR-263

MLAdapt - a machine learning-based method for detecting genome-wide adaptive introgression

X. Zhang [1,*], B. Kim [2], A. Durvasula [1], K. Lohmueller [1]

[1]Ecology and Evolutionary Biology, University of California, Los Angeles, Los Angeles, [2]Biology, Stanford University, Palo Alto, United States

Abstract: Adaptive introgression is known to play an important role in facilitating adaptation in a wide range of organisms including humans. Many current approaches for detecting adaptive introgression rely upon the "outlier approach", where outliers from scans for positive selection are intersected with putatively introgressed genomic regions. This approach is vulnerable to a high Type II error rate, as the power of different methods to infer selection or introgression can vary, particularly at different time scales. In addition, population genetic processes unrelated to the gene flow or selection of a beneficial variant, such as background selection or heterosis, may create similar genomic signals as adaptive introgression.

In this study, we present *MLAdapt* - a likelihood-free machine learning approach to identifying genomic regions that are likely to be adaptively introgressed. Based on a random forest regression model, this method jointly considers a large number of summary statistics to detect genomic regions that are likely adaptively introgressed. To further increase power, our training procedure incorporates genomic features such as variation in recombination and mutation rate, the distribution of functional elements, and the distribution of fitness effects, resulting in localized models across the genome. Compared to other existing methods, our approach takes into account the presence of deleterious variants and complex demography, and we demonstrate the power of our method at detecting adaptive introgression by applying it to modern human genome data.

Genome-wide methods for detecting selection

SMBE-OR-264

DFEnitely different: Genome-wide characterization of differences in mutation fitness effects between populations

A. L. Fortier [1], A. J. Coffman [2], T. J. Struck [3], J. L. Burguete [4], A. P. Ragsdale [5], P. H. Hsieh [6], R. N. Gutenkunst [3,*]

[1]Stanford University, Stanford, [2]University of Pennsylvania, Philadelphia, [3]University of Arizona, Tucson, United States, [4]National Autonomous University of Mexico, Mexico City, Mexico, [5]McGill University, Montreal, Canada, [6]University of Washington, Seattle, United States

Abstract: The fitness effect of a mutation may differ between populations, depending on environmental and genetic context. To quantify genomic patterns of such differences, we extended the concept of a distribution of fitness effects (DFE) to a joint DFE between populations. To infer the joint DFE, we fit parametric models that included demographic history to genomic data summarized in the joint allele frequency spectrum. We applied this framework to African and European populations of both Drosophila and humans, finding that mutation fitness effects are much more similar between populations of humans than Drosophila. Among gene sets in Drosophila, genes involved in immunity and stress response showed low similarity of fitness effects, whereas genes involved in reproduction showed high similarity. Our results represent the first genome-scale quantification of mutation fitness effect differences between populations and point toward gene functions that are more likely to experience divergent selection.

Mutation Rate Evolution
SMBE-OR-265
Evolution of germline mutation rate in great apes
M. E. Goldberg [1,*], K. Harris [1]
[1]Department of Genome Sciences, University of Washington, Seattle, United States

Abstract: Although the germline mutation rate is classically regarded as a fixed parameter of the evolutionary process, recent studies of human and ape genetic variation have shown that the mutation rate and spectrum can evolve rapidly. The relative mutation rates of different three-base-pair genomic motifs differ significantly among great ape species, suggesting the recent fixation of unknown modifiers of DNA replication fidelity. To shed light on what these modifiers might be, we measured the relative mutabilities of all three-base-pair motifs in specific compartments of the genome (such as endogenous retroviruses and late-replicating regions) that we expect to be targeted by known mutational processes. Using genetic diversity data from 79 great apes, we measured the covariation of mutational spectra covary between compartments and species, finding evidence of compartment-specific mutational processes that are robustly conserved across the ape phylogeny. These compartment-specific signatures layer with species-specific signatures to create rich mutational portraits: for example, orangutan endogenous retroviruses contain an identifiable mixture of an orangutan-specific signature and a signature that we hypothesize is due to hydroxymethylation of retrovirus-derived DNA. Strikingly, western chimpanzees have a different mutation spectrum from other subspecies of chimps, and the difference between western and non-western chimps closely resembles the difference between repetitive and nonrepetitive DNA. Our results suggest that rapidly evolving mutation rate modifiers tend to act broadly in trans across the whole genome, whereas cis regulators of mutation in specific genomic compartments are highly conserved between species.

Mutation Rate Evolution
SMBE-OR-266
Mutation rate plasticity evolution in bacteria
R. Krasovec [1,*], D. Gifford [1], H. Richards [2], R. V. Belavkin [3], A. Channon [4], E. Aston [4], A. J. McBain [1], C. Knight [1]
[1]The University of Manchester, Manchester, United Kingdom, [2]University of Waikato, Hamilton, New Zealand, [3]Middlesex University, London, [4]Keele University, Keele, United Kingdom

Abstract: Spontaneous mutation rate depends upon both an organism's genotype and its environment. We have previously demonstrated, using fluctuation assays, that mutation rate decreases in dense populations, relative to sparse populations, by as much as 23-fold across bacteria and yeast. We find that this 'density-associated mutation rate plasticity' (DAMP) critically depends on the activity of particular NUDIX hydrolase proteins responsible for removing damaged nucleotides. We ask how evolutionarily robust the association is between population density and mutation rates. DAMP varies among organisms, for instance, *Escherichia coli* wild-type K-12 and B strains have DAMP, but another gamma proteobacterium, *Pseudomonas aeruginosa* PAO1, does not. When we estimated mutation rates across a panel of 62 *E. coli* strains, isolated from a wide range of natural environments (the ECOR collection), we observed variation in both average mutation rates and in the degree of DAMP: 21 strains lacked DAMP, whereas in 41 strains mutation rates decreased in populations grown to higher density using increased nutrients. We then experimentally evolved (for ~600 generations) seven *E. coli* K-12 strains with different degrees of DAMP, including three constitutive mutator strains. We observed that, in evolved genotypes, the degree of DAMP did not change – DAMP is strikingly robust even in constitutive mutators, where evolved genomes contain many point mutations. We discuss the relationship of DAMP to the fitness gains and mutations identified across 185 experimentally evolved lines, considering the wider biological role of mutation rate plasticity.

Mutation Rate Evolution

SMBE-OR-268

Shifts in the polymorphism spectrum reflect changes in generation times in recent human evolution

P. Moorjani [1 2 3,*], Z. Gao [3 4], G. Amster [5], M. Przeworski [5 6]

[1]Department of Molecular and Cell Biology, [2]Center for Computational Biology, University of California Berkeley, Berkeley, United States, [3]These authors contributed equally, -, [4]Howard Hughes Medical Institute & Department of Genetics, Stanford University, Palo Alto, [5]Department of Biological Sciences, [6]Department of Systems Biology, Columbia University, New York, United States

Abstract: Recent sequencing studies of human pedigrees have shown that the numbers of de novo mutations increases with ages of reproduction of both parents in a sex-specific and context dependent manner (depending on flanking sequence), reflecting characteristics of the underlying mutational mechanisms. An important implication is that changes in life history traits, notably the male and female generation times (i.e. mean age of reproduction), are expected to affect the mutation spectra at polymorphic sites. Conversely, these considerations imply that the mutation spectra of polymorphism data carries information about life history traits over millions of years. To exploit this idea, we focus on the relative proportions of mutation types seen in pedigree studies, for different combinations of male and female generation times. We show how these quantities can be related to the observed proportions of mutation types in polymorphism data in order to obtain estimates of the mean generation times. By considering rare variants, common polymorphisms and divergent sites in turn, we estimate mean generation times over different timescales of human evolution. Our method thus provides an estimator of generation time over evolution, applicable to any species with polymorphism and de novo mutation data. We illustrate the method by estimating the recent and historical generation times for modern human populations, using data from the deCODE Genetics, 1000 Genomes Project, gnomAD and Simons Genome Diversity Project. These analyses uncover a marked increase in generation time over the course of human evolution, supporting a slow down in yearly mutation rates on the modern human lineage. Application of this method to archaic hominins further reveals that while Neanderthals had a very similar generation time to modern humans, Denisovans had a significantly shorter generation time.

Origins, evolution and function of novel genes

SMBE-OR-270

Variation and novelty in evolution: novel genes continuously arise from non-coding regions, enable protein structural innovation and function in the brain

V. Luria [1,*], A. Karger [2], W. Cain [3], A. O'Donnell [4], T. Gujral [5], M. Kirschner [1]

[1]Systems Biology, [2]IT - Research Computing, Harvard Medical School, Boston, [3]Mathematics, Harvard University, [4]Broad Institute of MIT and Harvard, Cambridge, [5]Human Biology, Fred Hutchinson Cancer Research Center, Seattle, United States

Abstract: The origin of new protein-coding genes is a central, unsolved evolutionary question. Most genes were thought to arise by duplicating or transferring existing genes. Yet novel genes arising *de novo* from genomic "junk" DNA or from long non-coding RNA were recently found. Novel genes present in one or few species are taxon-restricted and may encode new structurally new proteins. Strikingly, novel genes are invariably expressed in the brain and germline. We first found a taxon-restricted gene, *APCDD1*, and showed it functions in vertebrate neurons and skin. To then understand how novel genes appear and what their general properties are, we combined mathematical, computational and experimental approaches. To evaluate how frequently novel genes arise, we built a mathematical model based on gene and genome parameters and dynamic factors including mutation. We found genomes should continuously make many novel genes, yet keep few. We computationally identified candidate novel genes in 25 eukaryotic genomes by integrating Ribo-Seq and proteomics data with phylostratigraphy, and evaluated their predicted biophysical properties. Compared to ancient proteins, novel proteins are shorter, basic, fragile, disordered, promiscuous but less toxic. We experimentally compared novel and ancient proteins, tested gene function *in vivo* in zebrafish brains using CRISPR, and showed many novel genes are expressed in human brains throughout life. Thus genomic DNA turnover governs the flux of genes that are continuously created and destroyed. The few new genes kept encode proteins with distinct structural features and brain expression, showing genome dynamics generates variation that enables new structures and functions.

Origins, evolution and function of novel genes
SMBE-OR-271
New Gene Formation in Hybrid Drosophila
R. Rogers [1,*]
[1]UNC Charlotte, Charlotte, United States

Abstract: New gene formation in Hybrid *Drosophila*

Gene expression changes in hybrids can be a source of phenotypic variation. Transgressive gene expression occurs when hybrid offspring have expression changes beyond variation within parent species. Qualitative changes in expression have been well studied in a number of hybrid systems. However, the ability of hybrids to create new genes or activate existing genes has been largely overlooked. We have assayed expression changes in *D. santomea* x *D. yakuba* hybrids. We observe as many as 56 cases of new gene formation in a single cross. These new genes appear when trans acting factors from one species bind to a cis regulatory module that was present in the other species. We find that the number of new transcripts created depends on the direction of the cross. In *D. yakuba* ♀ *x D. santomea* ♂ crosses, a greater number of transcripts are formed in female offspring (21 vs 9). In *D. santomea* ♀ *x D. yakuba* ♂ crosses, a greater number of transcripts are formed in male offspring (56 vs 5) (χ^2=34.671, P=3.905 x 10^{-9}). These new transcripts formed in hybrids might potentially spread through populations via introgression. This source of genetic novelty may offer a source of new variation that could influence natural populations.

Genome-wide methods for detecting selection

SMBE-OR-272

Deep learning reveals population-specific changes to the landscape of recombination in Drosophila melanogaster

J. Adrion [1,*], J. Galloway [1], A. Kern [1]

[1]Biology, University of Oregon, Eugene, United States

Abstract: Accurately inferring the genome-wide landscape of recombination rates in natural populations is a central aim in genomics, as patterns of linkage influence everything from genetic mapping to understanding evolutionary history. Recent advances in machine learning, including the development of deep neural networks suitable for DNA sequence alignment data, have shown great promise in their initial application to population genetics questions. Here we describe ReLERNN, a deep learning method for accurately predicting the landscape of recombination rates across the genome using as few as four chromosomes. Rather than use summaries of linkage disequilibrium as its input, ReLERNN uses raw alignment columns, which are then modeled as a sequence across the genome using a recurrent neural network. We show that ReLERNN can predict complex landscapes of recombination with a high degree of accuracy. Moreover, we demonstrate that ReLERNN improves accuracy and reduces bias relative to existing methods and maintains high accuracy in the face of demographic model misspecification. We apply ReLERNN to natural populations of African *Drosophila melanogaster* and show that genome-wide recombination landscapes, while correlated among populations, show important population-specific differences. Finally, we explore how a combination of structural variation and natural selection contribute to the evolution of the landscape of recombination in *Drosophila*.

The molecular basis of major transitions in evolution

SMBE-OR-273

Phylogenomics Resolves Early Events in Bacterial Evolution

G. A. Coleman [1,*], G. J. Szöllősi [2], T. A. Williams [1]

[1]Biological Sciences, University of Bristol, Bristol, United Kingdom, [2]Biological Physics, Eötvös Loránd University, Budapest, Hungary

Abstract: A rooted tree of Bacteria is essential to reconstruct the early evolutionary history of life and the emergence of key geobiological interactions which affect us to this day. Many current ideas pertaining to the nature of bacterial evolution are informed by hypotheses of prokaryotic phylogeny. However, rooting the tree of Bacteria has proven difficult. Recent discoveries of a huge diversity of new uncultured phyla provide new data, but are often difficult to resolve within the bacterial tree, with the relationships between the major bacterial lineages still showing little resolution. We attempt to construct a rooted tree of Bacteria using probabilistic gene tree-species tree reconciliation methods. These hierarchical models integrate horizontal gene transfers (HGTs), gene duplications and gene losses into an overall model of genome evolution using amalgamated likelihood estimation, where patterns of gene family evolution contain information about the root of the tree. This rooting method also allows us infer ancestral gene content and reconstruct ancestral metabolisms for the internal nodes, including the last bacterial common ancestor, and to explore character evolution and rates of HGT over time. This provides insights into early bacterial evolution and key major evolutionary transitions, such as the evolution of photosynthesis, the double-membrane and terrestrialisation. Preliminary results support a root position between two large and diverse clades; one encompassing the 'Terrabacteria' and 'Candidate Radiation Phyla' superphyla, along with several other lineages, and another which comprises the Proteobacteria, PVC and FBC superphyla, and related lineages. These results tentatively support the early origin of both double-membranes and terrestrialisation.

The molecular basis of major transitions in evolution

SMBE-OR-274

Reconstructing LECA's 'duplome': The phylogenetic origins and order of gene duplications during eukaryogenesis

J. Vosseberg [1,*], J. J. E. van Hooff [1 2], L. M. van Wijk [1], M. Marcet-Houben [3 4], T. Gabaldón [3 4 5], B. Snel [1]

[1]Theoretical Biology & Bioinformatics, Department of Biology, Utrecht University, [2]Hubrecht Institute - KNAW (Royal Netherlands Academy of Arts and Sciences), Utrecht, Netherlands, [3]Centre for Genomic Regulation (CRG), The Barcelona Institute of Science and Technology, [4]Universitat Pompeu Fabra, [5]Institució Catalana de Recerca i Estudis Avançats (ICREA), Barcelona, Spain

Abstract: Eukaryogenesis - the emergence of the eukaryotic cell from its prokaryotic ancestors - is one of the most enigmatic evolutionary events and has sparked many controversies. It has become clear that the last eukaryotic common ancestor (LECA) already had a complex nature, reflected by its full eukaryotic cellular compartmentalisation and a relatively large chimaeric genome consisting of many paralogues. The numerous gene duplications that occurred during eukaryogenesis (the 'duplome' of LECA) were likely a driving force behind the rise in cellular complexity. It has remained largely unknown how and in which order the gene repertoire of LECA was shaped by gene duplications, horizontal and endosymbiotic gene transfers and gene inventions. We here present a novel phylogenomics-based estimate of the number of genes derived from prokaryotes and of newly invented genes, and their expansion via duplication. LECA's genome was estimated to contain over 10,000 genes, of which nearly half were the result of duplications. Although genes donated from bacteria - not alphaproteobacteria - and inventions comprised the largest part of LECA's gene repertoire, genes derived from the archaeal host were overrepresented in the duplome. In addition, we timed gene duplications relative to other gene acquisitions using branch lengths. Genes inherited from the archaeal host duplicated throughout eukaryogenesis - both before and after the bacterial influx - and duplications in alphaproteobacteria-derived and invented genes were late events. These findings support a 'mito-intermediate' scenario in which a relatively complex archaeal host took up a bacterial endosymbiont and many additional bacterial genes, which resulted in a further increase in cellular and genomic complexity.

Novel insights into evolutionary genetics from emerging technologies

SMBE-OR-275

Evolution of the common cold virus in healthy and asthmatic individuals

S. A. Megremis [1,*], P. West [1], S. Kandathil [2], B. Constantinides [1], G. Guibas [1], D. Robertson [3], S. Lovell [1], N. Papadopoulos [1]
[1]University of Manchester, Manchester, [2]University College London, London, [3]MRC-University of Glasgow Centre for Virus Research, Glasgow, United Kingdom

Abstract: Rhinoviruses (RVs), responsible for the common cold, are also causative agents in exacerbations of chronic airway diseases, which can lead to hospitalisation and death, such as in young children and atopic asthmatic patients. RVs are short-genome RNA viruses, with more than 160 characterised genomes to date that cluster in phylogenetic trees in 3 subgroups: A, B and C. They have evolved to subvert host control in several ways: using multiple receptors (ICAM-1, LDLRs, CDHR3), suppressing immunogenic dinucleotides in their genome sequence, and by diverting antibody responses. The asthmatic respiratory epithelium has been the target of investigation for many years and a deficient anti-viral innate immune response against rhinoviruses has been implicated in asthma pathology. We hypothesize that rhinoviruses infecting the asthmatic epithelium follow a different evolutionary trajectory compared to infection of the healthy non-asthmatic epithelium. To investigate this we have isolated, propagated and infected primary human nasal epithelial cells (PNECs) obtained from healthy (n=3) and asthmatic (n=3) adults with a characterised rhinovirus species A strain 1B (propagated on HeLa-H1: 4 passages at 18 hours) along with three biological replicates of Calu3 (lung cancer cell line) to evaluate the effect of different cell types on virus mutation. During the course of the infection we sequenced (Illumina NextSeq) the virus at three time points, 4, 24 and 72 hours post infection. Sequencing reads were aligned on our characterised 'stock' A1B genome. Most nucleotide changes were observed as early as 4 hours post infection in both healthy and asthmatic PNECs infections. Analysis of the A1B polyprotein revealed that the 3D polymerase was rich in destabilising amino acid changes. Overall, neutral and destabilising mutations were observed with a frequency higher than 2% in PNECs. A significantly higher number of A1B genome locations with nucleotide changes was found in asthma patients than in healthy non-asthmatic controls. Taking into account nucleotide changes detected in all donors per clinical group and at all time points, we observed a differential distribution within the A1B genome between the two clinical groups. This included both nucleotide changes which were retained from the virus stock (acquired during virus propagation) and de novo changes observed during the infection but not found in the virus stock. More changes were observed in the polymerase, protease, VP2, VP4 and IRES in asthma. Our results indicate that the accumulation of a distinct set of nucleotide changes at different locations of the rhinovirus genome and/or polyprotein is dependent on the host susceptibility and increased in the asthmatic phenotype at least *in vitro*. If this is true, rhinoviruses infecting asthmatics could be evolving under different selection pressure acquiring new advantageous traits.

Open Symposium
SMBE-OR-276
Accurate inference of natural selection and ancestral side-chain conformation from amino acid sequences
C. C. Weber [1,*], U. Perron [1], N. Goldman [1]
[1]EMBL-EBI, Wellcome Genome Campus, United Kingdom

Abstract: How can we best learn the history of a protein's evolution? Ideally, a model of sequence evolution should capture both the process that generates genetic variation and the functional constraints determining which changes are fixed. However, in practical terms the most suitable approach may simply be the one that provides a particular output. For example, we might be interested in a measure of the strength of selection (typically obtained using a codon model) or an ancestral structure (obtained using an empirical model incorporating rotamer states).
But what if the data in the relevant state-space (i.e. codons or rotamer states) are not readily available? We provide a proof of concept and show that it is possible to obtain asymptotically accurate estimates of the outputs of interest using a well-established method for handling missing data. Encoding observed amino acids as ambiguous characters allows the application of models with the desired features, operating in a larger state-space. This strategy is viable because the path taken through amino acid space contains information about states that were likely visited in the "unseen" state-space.
To illustrate this, we consider two examples. We show that *dN*/*dS*, a parameter describing the relative strength of selection on nonsynonymous and synonymous changes, can be estimated in an unbiased manner using an adapted version of a standard codon model. Additionally, we find that ancestral amino acid side-chain configuration can be inferred by applying a 55-state empirical model to 20-state data (amino acid sequences). Adding structural information to as few as 10% of the tips of a phylogeny results in remarkable performance compared to a benchmark that considers the full rotamer state information. Hence, ambiguity encoding allows inference using data and models of varying structures, whether they consider features of the generative process or a more detailed view of selective constraints at the level of protein structure.

Novel insights into evolutionary genetics from emerging technologies
SMBE-OR-279A
Evolutionary centromere repositioning fuels karyotype evolution in Drosophila
D. Bachtrog [1,*]
[1]University of California Berkeley, Berkeley, United States

Abstract: Centromeres are the basic unit for chromosome inheritance, but their underlying sequence and associated centromeric histone proteins vary substantially among species. Yet, given their embedment in highly repetitive pericentromeric DNA, their evolutionary dynamics is poorly understood. Here, we study centromere evolution in Drosophila by generating high-quality reference genomes using single-molecule sequencing and chromatin conformation capture for multiple species of the *obscura* group. In this species group, both acro- and metacentric chromosomes exist. We show that karyotype evolution in the *obscura* group of Drosophila was driven by the emergence and loss of centromeres, and in most cases, not accompanied by any chromosomal rearrangements that are usually thought to trigger these events. Flies in this group are ancestrally acrocentric, and the emergence of new centromeres in the middle of two autosomes in species of the *obscura* subgroup resulted in the creation of metacentric chromosomes. In both cases, new centromeres were seeded *de novo* at ancestrally gene-rich regions and the emergence of a centromere function resulted in a drastic size increase due to an accumulation of repetitive DNA. This was accompanied by dozens of genes previously located in euchromatin now being embedded in pericentromeric heterochromatin. These metacentric chromosomes secondarily became acrocentric in the *pseudoobscura* subgroup, and centromere repositioning and a pericentric inversion, respectively, appear to have shifted the location of the centromere towards the chromosome end. The former centromeres and pericentric sequences left behind in the chromosome arms shrunk dramatically in size after their inactivation, yet these regions contain remnants of their evolutionary past, including increased repeat content and heterochromatic environment. Centromere movements are accompanied by rapid turnover of the major satellite DNA detected in (peri) centromeric regions.

Open Symposium

SMBE-OR-280

Evidence for genetic exchange in the bdelloid rotifer Adineta vaga inferred from multiple genome sequences

O. A. Vakhrusheva [1,*], E. A. Mnatsakanova [2], Y. R. Galimov [3], T. V. Neretina [4], E. S. Gerasimov [4], S. G. Ozerova [3], A. O. Zalevsky [4], I. A. Yushenova [5], I. R. Arkhipova [5], A. A. Penin [4], M. D. Logacheva [6], G. A. Bazykin [1], A. S. Kondrashov [7]

[1]Center of Life Sciences, Skolkovo Institute of Science and Technology, [2]Department of Hydrobiology, Faculty of Biology, M.V. Lomonosov Moscow State University, [3]N.K. Koltzov Institute of Developmental Biology of the Russian Academy of Sciences, [4]M.V. Lomonosov Moscow State University, Moscow, Russian Federation, [5]Marine Biological Laboratory, Woods Hole, United States, [6]Genomics Core Facility, Skolkovo Institute of Science and Technology, Moscow, Russian Federation, [7]University of Michigan, Ann Arbor, United States

Abstract: Sexual reproduction is ubiquitous among eukaryotes. While transitions to asexuality occur frequently, most asexual lineages are short-lived and abandoning sexual reproduction is commonly regarded as an evolutionary dead end. However, existence of ancient asexuals challenges this point of view. Bdelloid rotifers, one of the most striking examples of putative ancient asexuals, presumably lost conventional meiosis tens of millions of years ago. The main reason to assume that bdelloids lack meiotic sex is the failure to detect males in numerous bdelloid species. However, this does not exclude the possibility of other modes of genetic exchange in their populations. Indeed, recent analyses based on several genomic regions suggested genetic exchanges in this group, but whole-genome evidence was missing. Here, we sequenced whole genomes of 11 wild-caught individuals of the bdelloid rotifer *Adineta vaga*. The observed patterns of population structure in *A. vaga* are incompatible with strict clonality. First, we found that linkage disequilibrium in *A. vaga* decays with physical distance and showed that this decay cannot be explained by gene conversion. Next, we detected a substantial number of triallelic sites harboring all three possible heterozygous genotypes, with the observed number significantly exceeding the one expected due to recurrent or back mutations. Finally, we reconstructed haplotype-based phylogenies at many genomic loci and identified multiple cases when the two haplotypes of a single individual clustered with haplotypes from different individuals. Collectively, our findings provide compelling evidence for genetic exchange in *A. vaga*, a species previously thought to be anciently asexual.

Contemporary Evolution

SMBE-OR-281

Fine-scale genetic relationships and potentially adaptive admixture among the rock-dwelling cichlids of Lake Malawi

T. Linderoth [1,*], M. Samborskaya [1], H. Svardal [2], I. Bista [1 3], G. Turner [4], R. Durbin [1 3]

[1]Department of Genetics, University of Cambridge, Cambridge, United Kingdom, [2]Department of Biology, University of Antwerp, Antwerp, Belgium, [3]Wellcome Trust Sanger Institute, Cambridge, [4]School of Natural Sciences, Bangor University, Bangor, United Kingdom

Abstract: For over half a century the cichlid fishes of the African Great Lakes have evoked awe among researchers as champion speciators. As many as 600 species of haplochromine cichlids currently inhabit Lake Malawi, with an estimated common ancestor within the last million years. Recent genomic investigations have pointed to complex evolutionary histories, often involving admixture between species, as one feature likely to underlie this rapid rate of speciation, as well as highlighting specific, pertinent genomic regions. In order to further elucidate the potential role that hybridization plays in rapid, adaptive radiations we used whole-genome sequences from 301 individuals to classify the fine-scale genetic and evolutionary relationships between nine species of the rock-dwelling cichlids from Lake Malawi. Focusing first on *Maylandia* species from the Chilumba region in the north of the lake, we see apparent admixture between *M. fainzilberi* and *M. emmiltos* from one sampling location. All of these sampled individuals of *M. emmiltos* have a larger proportion of their genome that is genetically more similar to sympatric *M. fainzilberi* than to *M. emmiltos* from another nearby location, suggesting genetic swamping via hybridization or, alternatively, that *M. emmiltos* has introgressed into *M. fainzilberi* and the latter have been misclassified based on morphology. We are conducting demographic analyses to identify which of these scenarios is more plausible. Another case of admixture is in Lake Malawi National Park in the south of the lake, where *Cynotilapia afra* was introduced via escape from fish exporters within the last 60 years. We identify at least three regions of the *C. afra* genome where material from the native *Maylandia zebra* has risen to high frequency within the last tens of generations, suggesting adaptive introgression post introduction. We are performing high-resolution characterization of these introgressed regions to determine whether they could have played a role in *C. afra*'s successful establishment and spread throughout the National Park region. Our findings support that interspecific hybridization is likely a relevant mechanism for generating the phenotypic diversity represented by the Lake Malawi cichlid radiation.

Evolution of phenotypes: understanding diversity and the role of plasticity in adaptation to new environments
SMBE-OR-282

Laboratory natural selection in Drosophila suggests widespread adaptive phenotypic plasticity

C. Schlötterer*, F. Mallard

Abstract: Phenotypic plasticity is the ability of a single genotype to produce different phenotypes in response to environmental variation. The importance of adaptive phenotypic plasticity in natural populations and its contribution to phenotypic evolution during rapid environmental change is widely debated. Here, we show that plasticity of gene expression in natural populations is adaptive: evolution to more extreme environments increases ancestral plasticity rather than mean genetic expression. We determined the adaptive value of plasticity in gene expression by conducting laboratory natural selection on a *Drosophila simulans* population in hot and cold environments. After almost 60 generations in the hot environment, 598 genes evolved a change in plasticity relative to the natural ancestral population. Plasticity increased in 75% of these genes, which were strongly enriched for several well-defined functional categories (e.g. chitin metabolism, glycolysis, and oxidative phosphorylation). Furthermore, we show that plasticity in gene expression of populations exposed to different temperatures is rather similar across species. We conclude that most of the ancestral plasticity is adaptive and can further evolve in more extreme environments.

Open Symposium

SMBE-OR-284

Is there a goose on the loose?: investigating introgression into the Swedish population of Lesser White-fronted Goose

D. Díez-Del-Molino [1,*]

[1]Bioinformatics and Genetics, Swedish Museum Of Natural History, Stockholm, Sweden

Abstract: Interspecific introgression is a potential threat to endangered taxa. One particular case where this had a major impact on conservation policy is the Lesser White-fronted Goose (LWfG) in Sweden. After declining dramatically to a only a few breeding pairs in the early 80s, a restocking program was set to release captive breeding birds. The program successfully released 341 LWfG birds to the wild until 1999 when some genetic studies detected mtDNA haplotypes characteristic from the Greater White-fronted Goose (GWfG) in captive LWfG birds, including some 16% from the breeding program. The program was then shut in 2000 as precautionary measure, but none of the released birds was tested. Thus, even though the presence of GWfG introgression in the wild LWfG Swedish population remains unknown, this has been used to motivate contentious conservation legislation and policies.

In this study we sequenced complete genomes from 21 LWfG individuals and used D-statistics to investigate GWfG introgression into LWfG birds. Our results are consistent with no introgression into the Swedish LWfG population respect the Russian or Norwegian ones, suggesting that even if some birds in the captive breeding program had such genes, these were never passed on to the wild population. This study demonstrates that genomic approaches can provide the robust scientific evidence necessary for the design of better conservation legislation and policies, and should be used to help minimize the risk of introducing hybrids into wild endangered populations through careful genetic monitoring of wild and captive individuals in breeding and restocking programs.

Using Ancient DNA to Study Natural Selction: New Models and New Data
SMBE-OR-286
27,566 Icelandic genomes reveal the nature of the introgressing Neanderthal
L. Skov*, M. M. Coll, G. Sveinbjörnsson, E. Lucotte, H. Jonsson, B. Halldorsson, D. Gudbjartsson, A. Helgason, M. H. Schierup, K. Stefánsson

Abstract: Human evolutionary history is rich with interbreeding between divergent populations. The impact of one such event can still be seen in Eurasians, who trace about 2% of their ancestry to introgression from Neanderthals 50-60 thousand years ago. The characteristics of these archaic hominins and their phenotypic impact on modern humans is of considerable interest. Here, we shed light on these questions using 14.4 million archaic chromosome fragments detected in fully phased whole genome sequences (WGS) from 27,566 Icelanders. They correspond to 111,616 unique archaic fragments that cover 50.3 % of the callable genome.

Based on similarity with known archaic genomes, we assign 84.5% of the fragments a Neanderthal origin, 3.3 % a Denisova origin and 12.4% an unknown origin. We propose that the introgressing archaic population was Neanderthal with an ancient contribution from Denisovans after splitting from the sequenced Vindija Neanderthal.
A paired comparison of archaic fragments with syntenic non-archaic fragments suggests that the mutation rate per year in Neanderthals was at most slightly higher than for humans for the roughly 500 thousand years during which these fragments existed separately in humans and Neanderthals. However, we detect significant differences to the spectrum of mutations and a significant excess of derived variants with predicted damaging effect in archaic fragments. Finally, we report five cases of phenotypic associations driven by archaic variants in 271 phenotypes, but show that the majority of previously reported associations can be better explained by non-archaic variants.

Using Ancient DNA to Study Natural Selction: New Models and New Data
SMBE-OR-287

Ancient DNA reveals that few disease-associated loci have been strongly selected during recent human history

C. N. Simonti [1,*], J. Lachance [1]
[1]Biological Sciences, Georgia Institute of Technology, Atlanta, United States

Abstract: Neutral and selective processes can alter the allele frequencies of disease-causing loci, and ancient samples allow us an unprecedented look at how our species has changed over time. Here, we analyzed a curated set of 2,709 GWAS loci in 143 ancient and 503 modern European genomes (including 24 loci that have introgressed alleles from Neanderthals). This time series dataset enabled us to identify disease-associated loci that have been targets of recent natural selection. First, we determined the maximum potential contribution of ancient hunter-gatherer, pastoral, and agricultural populations to modern European genomes. We then used a Bayesian method to infer allele frequency trajectories and selection coefficients. We find the majority of GWAS variants that impact health have negligible effects on fitness. Specifically, 895 of the 2,709 GWAS loci analyzed in our study show signatures of modest selection ($|s| > 0.001$), and a much smaller subset of loci (9 of 2,709) appear to be under strong selection ($|s| > 0.01$). Genetic variants that are located in the MHC region on chromosome 6 are enriched for signatures of selection. As expected, we found that protective alleles are more likely to be positively selected than alleles that increase the risk of complex diseases. This pattern was more pronounced for diseases that have an age of onset that occurs prior to reproduction. However, risk alleles at many disease-associated loci have increased in frequency over the last 10,000 years. This may be due to pleiotropy, recent relaxation of selection, and/or genetic hitchhiking of disease variants. Focusing on individual diseases, we find that alleles that protect against asthma are enriched for signatures of positive selection. Overall, these results demonstrate the potential of ancient DNA to improve our understanding of recent human evolution. Our results represent an important early step in using this new source of data to better understand how disease risk has changed over time.

Barriers and drivers of evolutionary innovation by horizontal gene transfer
SMBE-OR-288
Selection-based model of prokaryote pangenomes
M. R. Domingo-Sananes [1,*], J. O. McInerney [1]
[1]School of Life Sciences, University of Nottingham, Nottingham, United Kingdom

Abstract: Prokaryote genomes show great flexibility in the gene content within a species and bear the hallmarks of extensive horizontal gene transfer between species. This results in the existence of *pangenomes*: the complete set of genes present in a species, which includes *core* genes, present in all genomes, and *accessory* genes, whose presence varies among individuals. There is currently a lot of debate about which factors and forces shape the diversity and size of pangenomes, how is this diversity in gene content maintained, and whether or not accessory genes are on average advantageous or neutral in different prokaryote populations and species. In order to explore these issues, we developed a mathematical model to simulate the gene content of prokaryote genomes and pangenomes. In contrast to previous models, we focus on testing the role of selection and highlights the importance of gene gain and loss rates in shaping pangenomes. We find that accessory genomes can be explained through the existence of nearly-neutral genes within a constant environment, niche-dependent advantageous genes when multiple niches are available, or genes with any fitness effect if gene gain and loss rates are high. We therefore show that there are multiple possible mechanisms that can be responsible for maintaining diversity in gene content. We argue that information about population structure, gene-by-environment interactions and the distributions of the parameters presented in this model is crucial for understanding the contribution of different mechanisms to pangenome diversity in different prokaryote species.

Poster Abstracts

Barriers and drivers of evolutionary innovation by horizontal gene transfer

SMBE-PO-022

Early stage of P element invasion and hybrid dysgenesis in natural populations of Drosophila simulans in Japan.

R. Kondo [1 2,*], M. Oda [1], Y. Yoshitake [3], M. Itoh [3]

[1]Grad. Sch. of Humanities and Sciences, [2]Division of Natural Science, Ochanomizu University , Tokyo, [3]Dept. of Applied Biology, Kyoto Institute of Technology, Kyoto, Japan

Abstract: P-element is one of the transposable elements (TEs) in *Drosophila melanogaster* that encode transposase that cause the elements to move, and repressor that prevents expression of transposase. They cause P-M hybrid dysgenesis that is marked by germ cell degeneration at high temperature (29℃). Experimental data suggest that horizontal transmission of P-element occurred twice, first from *D. willistoni*to *D. melanogaster* about 60 years ago and recently from *D. melanogaster* to *D. simulans* in Europe around 2006. First invasion to *D. simulans* in Japan occurred rapidly by 2008. To investigate the spreading of P-element in Japanese *D. simulans* populations and to study the dynamics of early stage of P element invasion, we collected flies in Japan from 2015 to 2016 and examined the activity of P element through gonadal dysgenesis (GD) test. Our results indicate that active P element invaded all of *D. simulans* in Japan including those in the remote islands. Many can be classified as P strain (males have strong GD induction ability and females have strong GD repression potential). However, the remote island populations showed different dynamics which may be characteristic to early stage of invasion.

Barriers and drivers of evolutionary innovation by horizontal gene transfer

SMBE-PO-018

The Construction of Gene Sharing Networks of the Prokaryotic Immune System

Y. Ou [1,*], J. McInerney [2]

[1]School of Biological Sciences, The University Of Manchester, Manchester, [2]The University of Nottingham, Nottingham, United Kingdom

Abstract: Prokaryotic species use an adaptive defense system called clustered regularly interspaced short palindromic repeats (CRISPR). The CRISPR system is composed of direct repeats (DRs) and spacers acquired from invading bacteriophages. CRISPR and its associated cas genes are formed as defense against future viral attack. The evolutionary history of CRISPR involves rapid changes and is important in understanding of prokaryotes. Genetic recombination is common among prokaryote, but whether it plays roles in CRISPR-cas system is unknown. Here, we construct a dataset containing all available archaea and bacterial genomes, and using four different programs (CRISPRCasFinder, CRISPRDetect, PILER-CR, and MinCED) – we identify CRISPR arrays and cas genes. Across these four programs, we obtain varying results as to which species have CRISPR systems. Furthermore, we amalgamated our results into an intersection dataset by selecting spacers that were identified by all four programs (as well as their neighbouring DR). In the intersection dataset, 76.53% of archaea and 30.24% of bacteria were found to have CRISPR systems. Also, using UCLUST, similar DRs and spacers were clustered into gene families. Sequences within each cluster were then interrogated for recombination analysis. Out of 14,158 clusters, 32 clusters contained spacers belonging to species from different genera; only 1 cluster contained spacers from species identified from different families. Interestingly, species across different families or genera usually possessed the same cas gene subtype (Type I-E). These results indicate that, potentially, horizontal gene transfer is unexpectedly rare in CRISPR systems, and these events may have high correlation with cas gene subtypes.

Barriers and drivers of evolutionary innovation by horizontal gene transfer

SMBE-PO-003

Constraints to Horizontal Gene Transfer

R. K. Azad [1 2,*], S. Sengupta [1]

[1]Department of Biological Sciences, [2]Department of Mathematics, University of North Texas, Denton, United States

Abstract: Horizontal gene transfer (HGT) is the exchange of genetic material between organisms from different lineages with a profound impact on prokaryotic evolution. Disparate evolutionary histories of genomic fragments complicate the "tree of life" metaphor of organismal evolution and bring new challenges in understanding evolutionary relatedness or speciation, specifically for microorganisms. HGT is a major source of phenotypic innovation and a mechanism of niche adaptation in prokaryotes. Antibiotic resistance and pathogenicity are often the consequences of HGT, however, the scope of HGT goes far beyond this. Quantification of HGT is critical to deciphering its myriad roles in microbial evolution and adaptation. Beyond quantification and association of HGT with novel phenotypes, currently the focus of most studies in the field, understanding the factors or mechanisms that constrain or facilitate HGT is central to understanding the differential gene flow within and between different lineages and the relationships that are based on shared similarity. Recent studies in this direction have focused on the impacts of ecology, phylogeny and geography on horizontal gene flow. On the other hand, some studies have attempted to associate the mobility of genes or gene clusters to their biological functions. Here, we focus on understanding the constraints on HGT by analyzing over 2,000 completely sequenced genomes of prokaryotes that also have meta-information available particularly on their ecological niches, geographical locations, lifestyle, and phylogeny. We assess various factors for their impact on horizontal gene flow by first constructing a gene sharing network. The network is decomposed into gene sharing modules using Markov clustering algorithm. These modules provide information on potential gene donors and recipients, as well as the information on the preferential donors and recipients for different lineages. The meta-information is overlaid on the gene sharing network and then the differential constraints on transfers of single genes or clusters of genes are assessed.

Barriers and drivers of evolutionary innovation by horizontal gene transfer

SMBE-PO-021

Detecting unknown diversity with horizontal gene transfers

T. Tricou [1,*]

[1]LBBE, Villeurbanne, France

Abstract: Extant species represent only a fraction of all of the species that have roamed the earth. Detecting and studying those unknown species is essential to better understand extant biodiversity and better predict its future. Horizontal transfers, the transmission of genetic material between individuals in a non-genealogical manner, is a major driver of evolution in prokaryotes and also in eukaryotes. This implies that some genes that are present in today species may have originated or evolved for some time in species that are now extinct or unsampled. These transferred genes thus carry information on the nature and existence of extinct and unsampled diversity.

The comparison of a species tree and a gene tree, with so-called reconciliation methods, has the potential to detect genome-level events like duplications, losses and horizontal transfers. If many genes are analyzed, we hypothesize that transfers originating from the same hidden clade will be found to originate from the same branch after reconciliation, thus revealing the presence of such groups in the species tree.

In order to test this hypothesis, and explore the potential of reconciliation methods to detect hidden clades (extinct or still unknown), I performed simulations using Zombi (Davin, Tricou, Tannier, de Vienne, & Szollosi, 2018), a new kind of species-gene-genome simulator that we recently developed and that takes extinct species into account.

I show that the reconciliation of gene trees and species trees, with the ALE model (Szöllosi, Tannier, Lartillot, & Daubin, 2013), allows retrieving in a phylogeny the species that are extinct or unknown, and placing them correctly in the tree. I also explore the limits of this detection according to various parameters of the model (shape and size of the species tree, rates of duplications, losses and transfers, number of genes), and discuss its applicability to biological datasets and its potential to detect macroevolutionary patterns like mass extinctions or radiations.

Davin, A. A., Tricou, T., Tannier, E., de Vienne, D. M., & Szollosi, G. J. (2018). *Zombi: A simulator of species, genes and genomes that accounts for extinct lineages*. *bioRxiv*. https://doi.org/10.1101/339473

Szöllosi, G. J., Tannier, E., Lartillot, N., & Daubin, V. (2013). Lateral gene transfer from the dead. *Systematic Biology*, *62*(3), 386–397.

Barriers and drivers of evolutionary innovation by horizontal gene transfer
SMBE-PO-020

The extent of DNA transfer between plasmids and chromosomes in prokaryotes

A. S. Kadib Alban [1,*], G. Landan [1], T. Dagan [1]
[1]CAU, Kiel, Germany

Abstract: Plasmids are extra-chromosomal genetic elements that replicate autonomously in their host cell. Mobile and transmissible plasmids are considered as important agents of lateral gene transfer (LGT) in bacterial evolution. An important step in plasmid-mediated LGT is the transfer of DNA between plasmids and chromosomes, which occurs in the donor and recipient cells. Nonetheless, the extent of DNA transfer between plasmids and chromosomes has been poorly studied. Here we quantify the extent of homologous regions between plasmids and chromosomes using a novel comparative genomics approach. Our method combines patterns of local sequence similarity into a unified framework of syntenic regions (SRs) between genome pairs. This enables us to detect syntenic regions of heterogeneous sequence similarity levels in the presence of evolutionary events such as genomic rearrangements (with or without inversions), duplications, as well as insertions and deletions. We applied our approach to 1,400 prokaryote strains including 3,270 plasmid-chromosome pairs that co-inhabit the same cell. Our analysis uncovered 423,705 SRs in 3,215 plasmid-chromosome pairs with a median of 29 SRs per plasmid-chromosome pair. The majority of SRs (70%) are continuous in both replicons, while the remaining SRs are fragmented due to genome rearrangements and insertions/deletions. Moreover, we found a significant correlation between the plasmid size and the frequency of DNA transfer with the chromosome as indicated by the size of SRs. The copy number of SRs in the chromosome is often higher than that of the plasmid, which can be explained by repeated transfer events or duplications of the acquired DNA in chromosomes. Our approach uncovers frequent DNA transfer between plasmids and chromosomes that is highly variable among different prokaryotic lineages.

Barriers and drivers of evolutionary innovation by horizontal gene transfer

SMBE-PO-025

Frequent Horizontal Transmission, Recombination, and Horizontal Gene Transfer Maintain Ever-Young Endosymbiont Genomes

S. L. Russell [1,*], R. Corbett-Detig [2]

[1]Molecular, Cell, and Developmental Biology, [2]Biomolecular Engineering, University of California Santa Cruz, Santa Cruz, United States

Abstract: Symbiotic relationships between bacteria and eukaryotes are ubiquitous and powerful drivers of genome evolution. In particular, the way in which symbionts are transmitted between host generations is one of the main factors influencing bacterial symbiont genome evolution. Theory predicts drastic endosymbiont genome size reduction and structural stasis over time; however, this is based on the strictly vertically transmitted primary symbionts of insects, and many symbioses do not fit this pattern. Here, we use Illumina and Nanopore sequencing to assemble complete mitochondrial and bacterial symbiont genomes from several populations of chemosynthetic deep sea clams, mussels, and shallow-water bivalves that span from nearly strict vertical transmission, to mixed mode transmission, to strict horizontal transmission. We develop a theoretical framework and fit our data to a coalescent model of symbiont evolution to infer the horizontal transmission rates, recombination rates and tract lengths, and horizontal gene transfer (HGT) frequencies of the different symbiont taxa. Our findings reveal that even low to moderate rates of gene flow are sufficient to maintain symbiont genomes in a state similar to their free-living bacterial ancestors. However, symbiont genomes decay rapidly at extremely low horizontal transmission rates. Given the diversity of bacterial symbiont genomes and the prevalence of mixed modes of transmission in nature, these results are relevant to a wide diversity of symbiotic associations, from insect facultative symbionts to the human microbiome.

Barriers and drivers of evolutionary innovation by horizontal gene transfer
SMBE-PO-010
Swapping segments - global patterns of reassortment in influenza A viruses

D. Venkatesh [1,*], N. Lewis [1]
[1]Pathobiology and Population Sciences, Royal Veterinary College, London, United Kingdom

Abstract: Just as we are beginning to grasp the extensive scale and impact horizontal gene transfer (HGT) in bacteria, we are finding that reassortment, it's viral equivalent, likely has a central role in the biology of influenza A viruses (IAV) and other segmented viruses. Reassortment of IAV gene segments between different lineages and hosts have been associated with all flu pandemic and several epizootic events recorded thus far. Experimental studies have shown the impact of factors like packing signals, polymerase compatibility, and segment mismatch on reassortment in IAV. However, how the effect of such factors manifests at the population level is unclear.
In this study we reconstruct the population dynamics of global IAV over the past ten years (2008-18), and use a Bayesian phylogeographic approach to quantify the spread of viral segments within and across host, HA type, lineage, and geographical barriers to identify their impact on reassortment. We look in swine, human and avian hosts, and focus on the H1 and H3 types of HA (and their lineages) which are the main types of IAV found in mammalian hosts. We hope to begin to answer questions like: How common is reassortment, really? What are the ecological factors that limit or promote reassortment? Is reassortment directional? E.g. are segments associated with certain lineages more likely to 'spread' in the overall population of IAV? Do we see evidence at a population level for molecular factors that affect reassortment? E.g. are some combinations of segments less likely to reassort than others?

Barriers and drivers of evolutionary innovation by horizontal gene transfer
SMBE-PO-009
Towards a natural classification of bacterial plasmids based on sequence similarity
M. Acman [1,*], L. van Dorp [1], F. Balloux [1]
[1]Genetics, Evolution and Environment, University College London, London, United Kingdom

Abstract: Bacterial plasmids are promiscuous and changeable mobile genetic elements helping the spread of antimicrobial resistance (AMR) and virulence via horizontal gene transfer (HGT). Current classification systems, such as those based on replicon (PlasmidFinder) or conjugation (MOB-suite) machinery, are failing to capture the full scope of the plasmid diversity with more than 70% of publicly available complete plasmid sequences remaining unclassified. Moreover, replicon-based classification scheme is sometimes difficult to interpret and it is relying on detection of a single relatively conserved gene with larger plasmids often falling into multiple groups. This project aims to devise an improved classification scheme based on genomic distances between plasmids and their core genome composition. We examined plasmid distance matrix by using hierarchical clustering and community detection algorithms. This uncovered a distinct population structure with layered and overlapping plasmid groups. We found good accordance between emerging plasmid groups and current plasmid classifiers. Many groups were found to have unique private backbones comprised of multiple genes, while other sporadically shared their core genes with other plasmids thus highlighting the plasticity of plasmid genomes. Extensive HGT between plasmids is certainly a limitation of sequence-based classification approaches and assessing its effect on the formation of plasmid groups remains a challenge. Our future work will consider how this large-scale population structure reflects plasmid evolution and some key phenotypic features like HGT potency, virulence, AMR, heavy metal resistance, transposon carriage, or host range.

Barriers and drivers of evolutionary innovation by horizontal gene transfer
SMBE-PO-005
A Plasmid Epidemics spreading AMR genes
A. Ledda [1,*]
[1]Imperial College London, London, United Kingdom

Abstract: Plasmids are known to be selfish genetic elements. Their ability to be acquired from bacteria at the time they are most needed makes them an invaluable source of highly adaptive genes, such as antimicrobial or metal resistances. To avoid being lost when not strictly needed (and hence a burden for the host bacteria), plasmids developed a realm of different strategies. These strategies range from the most selfish, like addiction modules that kill the bacterial if it gets rid of the plasmid, to the most altruistic, like diversifying their gene content as much as possible in order to maximize their opportunities to provide a selective advantage for the host bacteria.

Here we present recent data on a plasmids epidemics in an hospital ward. The outbreak involves different Enterobacteriaceae hosts and a carbapenem resistance. The plasmid responsible for the outbreak is remarkably well conserved. More than half of the bacterial hosts belong to E.coli ST399, which was not previously associated to carbapenemase resistance in literature. Further analysis shows unmistakable marks of coevolution between the plasmid and the E.coli host. Our analysis show that the backbone of the epidemics is represented by the combination of strongly linked plasmid and E.coli ST399 host. This clonally expanding epidemics supports a plasmid epidemics transmitted via HGT to different Enterobacteriacee hosts. This dynamics allows us to make an estimation of the plasmid conjugation rate.

Barriers and drivers of evolutionary innovation by horizontal gene transfer
SMBE-PO-007
Prophages and other mobile genetic elements as regulatory switches
F. Wegner*, E. P. C. Rocha

Abstract: Active lysogeny is a form of phage-host interaction in which temperate phages are integrated into functional open reading frames. Controlled excision under specific conditions leads to the complete reconstitution of the gene, whilst re-integration of these elements leads to the halt of its expression. This provides a complex regulatory mechanism of bacterial gene expression. Such regulatory switches have been observed in genes relating to competence or sporulation, and can occur with other mobilizable elements beyond phages.
Here, we present the screen of the complete genomes of 117 bacterial species for the presence of interrupted genes and the potential of intact or cryptic prophages as well as other integrative elements to act as regulatory switches. We identify more than 1,000 candidate events. Our findings indicate that interrupted genes are particularly enriched for functions relating to replication, recombination and repair, but also defence mechanisms and certain transcription factors. We further identify candidates for experimental validation.
Active lysogeny highlights the potential of microbes to co-adapt their mobile genetic elements towards genetic regulation, which is key for pathogenicity or adaptation to new environments.
Active lysogeny is a form of phage-host interaction in which temperate phages are integrated into functional open reading frames. Controlled excision under specific conditions leads to the complete reconstitution of the gene, whilst re-integration of these elements leads to the halt of its expression. This provides a complex regulatory mechanism of bacterial gene expression. Such regulatory switches have been observed in genes relating to competence or sporulation, and can occur with other mobilizable elements beyond phages.

Here, we present the screen of the complete genomes of 117 bacterial species for the presence of interrupted genes and the potential of intact or cryptic prophages as well as other integrative elements to act as regulatory switches. We identify more than 1,000 candidate events. Our findings indicate that interrupted genes are particularly enriched for functions relating to replication, recombination and repair, but also defence mechanisms and certain transcription factors. We further identify candidates for experimental validation.
Active lysogeny highlights the potential of microbes to co-adapt their mobile genetic elements towards genetic regulation, which is key for pathogenicity or adaptation to new environments.

Barriers and drivers of evolutionary innovation by horizontal gene transfer

SMBE-PO-004

Horizontal transfer of bacterial cytolethal distending toxin genes to insects

K. I. Verster [1,*], J. H. Wisecaver [2], R. Duncan [3], M. Karageorgi [3], A. D. Gloss [4], E. E. Armstrong [5], D. K. Price [6], N. K. Whiteman [1]

[1]University of California - Berkeley, Berkeley, CA, [2]Purdue University, West Lafayette, IN, [3]University of California - Berkeley, Berkeley, [4]University of Chicago, Chicago, IL, [5]Stanford University, Stanford, CA, [6]University of Nevada - Las Vegas, Las Vegas, NV, United States

Abstract: Cytolethal distending toxins (CDTs) are tripartite eukaryotic genotoxins encoded in diverse bacterial and phage genomes. The cdtB subunit is a DNAse that causes eukaryotic cell cycle arrest and apoptosis, and in one context, is associated with resistance against parasitoid wasp infections. Here we report the discovery of functional cdtB copies in the nuclear genomes of insect species from two distantly related insect orders, including fruit flies (Diptera: Drosophilidae) and aphids (Hemiptera: Aphididae). Insect-encoded cdtB copies are most closely related to bacteriophage copies, were horizontally transferred to insect genomes > 40 million years ago and encode a protein that retains ancestral DNase activity. This phage-derived toxin has been domesticated by diverse insects and we hypothesize that it is used as a defensive weapon against parasitoid wasps

Barriers and drivers of evolutionary innovation by horizontal gene transfer
SMBE-PO-008
The bacterial capsule: a sweet barrier to DNA exchanges?
M. Haudiquet [1 2,*], O. Rendueles [1], E. P. C. Rocha [1]
[1]Microbial Evolutionary Genomics, Institut Pasteur, CNRS UMR 3525, [2]Ecole Doctorale FIRE - Programme Bettencourt, CRI, Paris, France

Abstract: Capsules are the outermost layer of bacteria and are present across all major phyla. They are also major virulence factors and increase tolerance to antibiotics. Furthermore, capsules allow bacteria to colonize novel environments and to withstand harsh conditions. These conditions favour horizontal gene transfer (HGT), suggesting that capsulated bacteria are associated with high rates of HGT[1], contradicting the current paradigm that capsules act as a physical barrier to DNA exchanges. Thus, we aim to decipher the interplay between the capsule and genetic exchanges. To do so, we combine genomics, computational modelling and experimental evolution. As a model organism, we use Klebsiella *pneumoniae (Kpn),* an enterobacteria that is becoming an increasing health threat due to the gain of antibiotics resistance genes through HGT. *Kpn* encodes a single polysaccharidic capsule that can be composed of different oligosaccharide residues which result in a variety of serotypes. We functionally annotated the genomes of thousands of public genomes of *Kpn* strains to quantify several mobile genetic elements (MGE) like prophages, plasmids, integrons and horizontally transferred genes like antibiotic resistance genes. In parallel, we predicted the polysaccharide structure of the capsule of these strains with *Kaptive*[2], a recent software able to infer the serotype from draft genomes. We show that serotypes of *Kpn* are associated with different levels of HGT, reflected by wide variations in term of MGE content. These results support our hypothesis that the capsule plays a role in mediating DNA exchanges and cannot be resumed as a physical barrier to HGT.

1. Rendueles et al. PLoS Genetics, 14(12).
2. Wyres et al. Microbial Genomics, 2(12).

Barriers and drivers of evolutionary innovation by horizontal gene transfer

SMBE-PO-012

What will happen if a foreign gene meets its native counterpart? A comparison of protein products between a pair of PgiC loci in Festuca ovina

Y. Li [1,*]

[1]Computational Biology and Biological Physics, Department of Astronomy and Theoretical Physics, Lund University, Sweden, Lund, Sweden

Abstract: Through horizontal gene transfer, genes can move between species. Among the introduced foreign genes, some may have native counterpart genes in the recipient species (i.e. duplicative horizontal gene transfer). Then how could the two originally separated genes get along with each other at their abrupt encounter after the duplicative horizontal gene transfer? The unique duplicative horizontal gene transfer case that has been identified in the grass *Festuca ovina* may deepen our understanding of this fascinating question. Among the two genes (*PgiC1*, *PgiC2*) encoding the metabolic enzyme PGIC (comprising two subunits) in *F. ovina*, one (*PgiC2*) was suggested to be acquired from another grass genus. The foreign *PgiC2* gene and its native counterpart *PgiC1* in *F. ovina* can together produce a hybrid protein (PgiC1-PgiC2), which means that this foreign gene will interact with its native counterpart by influencing not only the amount of, but also the biochemical/physicochemical properties of, the protein products of the native gene. Here in this work, we have been studying how the PgiC1-PgiC2 hybrid protein differs from its two 'parent' proteins (the native PgiC1-PgiC1 and the foreign PgiC2-PgiC2) in properties like the binding affinity between the two subunits that compose a dimeric PgiC protein. To reach our goal, we first predicted the 3D protein structures for each protein using homology modeling and/or protein-protein docking. We then closely inspected these structures for their basic characteristics at both whole-protein (e.g. solvent accessible surface area) and single amino acid (e.g. charge and hydrophobicity) levels using a range of bioinformatics tools. We are currently carrying out biophysical pulling simulations to search for possible differences in inter-subunit binding affinity between the three studied proteins using the Monte Carlo based software package PROFASI. A literature survey will also be performed to gather further support for our prediction from relevant experimental evidences.

Barriers and drivers of evolutionary innovation by horizontal gene transfer

SMBE-PO-013

The distribution of fitness effects of an antibiotic resistance plasmid

A. Alonso-Del Valle [1,*], J. DelaFuente [1], J. Rodríguez-Beltrán [1], R. León-Sampedro [1], C. delaVega-Losada [1], M. Hernández-García [1], P. Ruiz-Garbajosa [1], R. Cantón [1], Á. San Millán [1]

[1]Servicio de Microbiología, Hospital Universitario Ramón y Cajal (IRYCIS), Madrid, Spain

Abstract: Plasmids play a key role in the evolution antibiotic resistance (AR) because they are responsible for the horizontal spread of AR mechanisms among bacterial pathogens. Notably, certain associations between AR plasmids and bacteria become particularly successful, creating "superbugs" that disseminate uncontrollably in clinical settings. Despite the advantages conferred by plasmids they usually impose a fitness cost in the host bacterium, which limits the spread of the plasmid-carrying clone. We argue that the success of bacterium-plasmid associations in clinical settings could be determined by the fitness costs produced by the plasmid. To test this hypothesis, we study one of the most concerning AR plasmid associated with enterobacteria in our hospital, pOXA-48 (carrying the carbapenemase gene $bla_{\text{OXA-48}}$). To determine the distribution of fitness effects of pOXA-48 in its bacterial hosts we follow two approaches. First, using a CRISPR-Cas9 system, we cured the plasmid from a selection of pOXA-48-carrying enterobacteria isolated from hospitalized patients (*Klebsiella pneumoniae*and *Escherichia coli*). Second, we introduced the plasmid in a set of enterobacteria ecologically compatible with this plasmid (pOXA-48-free clones isolated from the same group of patients as the first ones). Finally, we measured the fitness effects produced by pOXA-48 and sequenced the genomes of the clones to look at the distribution of fitness effects across the phylogeny. Our results indicated that pOXA-48 produced lower costs in *K. pneumoniae* than in*E. coli.*Interestingly, we observed that, at the clonal level, plasmid permissiveness seems to correlate with the epidemiological success of the bacterium-plasmid associations in the hospital.

Barriers and drivers of evolutionary innovation by horizontal gene transfer
SMBE-PO-016

The origin, evolution and population structure of 4,071 global Escherichia coli ST131 genomes

A. G. Decano [1,*], T. Downing [1]
[1]School of Biotechnology, Dublin City University, Dublin, Ireland

Abstract: Background: Advances in the genomic resolution of large bacterial collections facilitates the study of specific evolutionary events across the core and accessory genomes. In pathogenic bacteria, these evolutionary changes are driven by the widespread use of beta-lactam antibiotics countered by bacterial extended-spectrum beta-lactamase (ESBL) genes. Within patients, these ESBL genes often transferred to new chromosomal or plasmid locations by mobile genetic elements (MGEs), during which the ESBL genes are often amplified, truncated or mutated. Between patients, genomic disease surveillance has identified plasmid transfer between bacteria and subsequent recombination of ESBL genes and MGEs as another adaptive pattern. Here, we use *Escherichia coli* ST131 as a paradigm for genomic bacterial epidemiology because its accessory genome, encompassing ESBL genes, MGEs and plasmids, is highly dynamic. ST131 is a major cause of infection worldwide that has three major clades A, B and C, where C has the widest level of drug-resistance and is the most prevalent globally.
Methods: By extracting all available high-quality global ST131 Illumina HiSeq read libraries, we automated quality-control, genome *de novo* assembly, plasmid reconstruction, ESBL gene screening and DNA read mapping in the largest ST131 sample collection examined this far. We reconstructed the genealogical histories of 4,071 genomes to infer their population structure and recombination patterns. We used long Nanopore and PacBio reads, published reference genomes and efficient k-mer-based methods to contextualise the core and accessory genome diversity observed to pinpoint the key emerging ST131 subclades.
Results: Our results provided a deep resolution of the origins of each ST131 subclade and the genetic changes corresponding to five genetically distinct C subclades. We refined the evolutionary history of the latter in particular to highlight plasmid, MGE and ESBL gene changes associated with their recent origin and spread. We compared the genealogical history based on mutations at the homogeneous core genome to the highly variable accessory genome, including plasmids and ESBL genes transposed by MGEs. This indicated instances where accessory genome changes corresponded to new radiations of descendant lineages and subclades.
Conclusion: Our findings underpin the tremendous power of improving infection treatments using high-throughput pan-genomics of large pathogen collections to highlight specific genetic switches driving new outbreaks.

Barriers and drivers of evolutionary innovation by horizontal gene transfer
SMBE-PO-011
Sexual selection in bacteria?
M. Vos*, A. Buckling, B. Kuijper

Abstract: A main mechanism of lateral gene transfer in bacteria is transformation, where cells take up free DNA from the environment which subsequently can be recombined into the genome. Bacteria are also known to actively release DNA into the environment through secretion or lysis, which could aid uptake via transformation. Various evolutionary benefits of DNA uptake and DNA release have been proposed but these have all been framed in the context of natural selection. Here, we interpret bacterial DNA uptake and release in the context of sexual selection theory, which has been central to our understanding of the bewildering diversity of traits associated with sexual reproduction in the eukaryote world but has never been applied to prokaryotes. Specifically, we explore potential scenarios where bacteria releasing DNA into the environment could compete for successful uptake by other cells, or where bacteria could selectively take up DNA to enhance their fitness. We conclude that there is potential for sexual selection to act in bacteria, and that this might in part explain the considerable diversity in transformation-related behaviours.

Barriers and drivers of evolutionary innovation by horizontal gene transfer

SMBE-PO-017

E. coli ST73: not good plasmid hosts

J. A. Gama [1,*], J. Kloos [1], P. J. Johnsen [1], Ø. Samuelsen [1 2]

[1]Department of Pharmacy, UiT The Arctic University of Norway, [2]Department of Microbiology and Infection Control, University Hospital of North Norway, Tromsø, Norway

Abstract: Plasmids are well known to promote the transfer of antibiotic-resistance genes among bacteria. Acquisition of plasmids may reduce the relative fitness of their hosts in the absence of selection. This has been shown mainly with model organisms and drug-resistance determinants. Here, we studied the effects of two clinical plasmids encoding resistance against carbapenems (last resort antibiotics) on 21 genetically diverse clinical *E. coli* strains representing 14 different sequence types (ST).

One plasmid (IncA/C2 bla_{NDM-1}) was stably maintained in most strains after ≈300 generations. In contrast, a IncL bla_{OXA-48}-encoding plasmid was not so stably maintained. Neither plasmid imposed significant fitness costs in the majority of strains when newly acquired with the exception of the IncA/C2 plasmid that reduced host fitness mainly in ST73 strains.

Conjugative transfer from the original isolates was higher for the IncA/C2 plasmid (≥ $1x10^{-6}$ transconjugants/donor) than for the IncL plasmid (< $1x10^{-6}$ transconjugants/donor). Moreover, both plasmids and particularly the IncL plasmid showed reduced conjugative transfer into ST73 strains. Analysis of *E. coli* genomes belonging to the studied STs revealed lower frequency of plasmids in ST73 genomes than in other STs. Interestingly, ST73 genomes encode type III restriction-modification systems more frequently than other STs. We also observed, among genomes belonging to ST73, a negative correlation between plasmid carriage and presence of type III restriction-modification systems.

In conclusion, we report two clinical plasmids harboring resistance to last resort antibiotics that do not impose detectable fitness cost in a considerable amount of strains. Additionally, ST73 strains exhibit a weaker ability to acquire plasmids than other STs. We suggest that type III restriction-modification systems may pose a barrier to plasmid acquisition in ST73. Overall, these data may explain, at least in part, why molecular epidemiology studies suggest that ST73 is less associated with antibiotic resistance determinants.

Barriers and drivers of evolutionary innovation by horizontal gene transfer

SMBE-PO-024

Delivery of CRISPR-Cas9 targeting antimicrobial resistance genes using a broad host-range conjugative plasmid

E. Pursey [1 2,*], D. Sünderhauf [1 2], F. Paganelli [3], W. Gaze [2], E. Westra [1], S. van Houte [1]

[1]College of Life and Environmental Sciences, University of Exeter, Penryn, [2]University of Exeter Medical School, University of Exeter, Truro, United Kingdom, [3]Department of Medical Microbiology, University Medical Center , Utrecht, Netherlands

Abstract: Of the numerous approaches to counteracting the spread of antimicrobial resistance (AMR), one that is receiving increased attention is the use of CRISPR-Cas-based gene editing technology to remove resistance-conferring sequences from bacteria. Proof-of-principle studies have shown that CRISPR-Cas is able to successfully remove AMR genes from monocultures or very simple bacterial communities in the laboratory, but these constructs and their delivery must be adapted to be used in real-world medical settings. Broad host-range conjugative plasmids are a delivery method that is highly adaptable to diverse microbial communities. We recently demonstrated that delivery of an AMR-targeting CRISPR-Cas construct with the conjugative plasmid pKJK5 is able to reduce the proportion of *E. coli* K12 recipients carrying plasmid-based gentamicin resistance by 33% compared to a nontargeting control. However, one issue with delivery to natural isolates using this system will be host defence factors affecting plasmid spread. When using an *E. coli* MG1655 donor with a recipient set of multidrug-resistant (MDR) clinical isolates of *E. coli*, all MDR isolates can receive pKJK5, and there is significant variation in conjugation frequency between *E. coli* recipients. MDR isolates of *E. coli* and *Klebsiella pneumoniae* will therefore be compared with drug-susceptible isolates to determine whether uptake of the delivery plasmid varies between the two groups. Data on conjugation frequencies alongside whole genome sequences will be used to interrogate the factors affecting plasmid transfer, such as restriction-modification systems, endogenous CRISPR-Cas systems, and plasmid incompatibility.

Barriers and drivers of evolutionary innovation by horizontal gene transfer

SMBE-PO-015

Systematic survey of non-retroviral virus-like elements in eukaryotic genomes

S. Nakagawa [1,*], K. Kirill [1], M. T. Ueda [1], T. Imanishi [1]

[1]Department of Molecular Life Science, TOKAI UNIVERSITY SCHOOL OF MEDICINE, Isehara, Japan

Abstract: Endogenous viral elements (EVEs) are viral sequences that are endogenized in the host cell. Recently, several eukaryotic genomes have been shown to contain EVEs. To improve the understanding of EVEs in eukaryotes, we have developed a system for detecting EVE-like sequences in eukaryotes and conducted a large-scale nucleotide sequence similarity search using all available eukaryotic and viral genome assembly sequences (excluding those from retroviruses) stored in the National Center for Biotechnology Information genome database (as of August 14, 2017). We found that 3,856 of 7,007 viral genomes were similar to 4,098 of 4,102 eukaryotic genomes. For those EVE-like sequences, we constructed a database, Predicted Endogenous Viral Elements (pEVE, http://peve.med.u-tokai.ac.jp) which provides comprehensive search results summarized from an evolutionary viewpoint. A comparison of EVE-like sequences among closely related species may be useful to avoid false-positive hits. We believe that our search system and database will facilitate studies on EVEs.

Barriers and drivers of evolutionary innovation by horizontal gene transfer
SMBE-PO-019

Horizontal gene transfer throughout the grass family

L. Dunning [1,*], S. Hibdige [1], P.-A. Christin [1]
[1]Animal and Plant Sciences, University of Sheffield, Sheffield, United Kingdom

Abstract: There is increasing evidence that horizontal gene transfer (HGT) is a prevalent evolutionary force in certain groups of eukaryotes. Among plants, we have shown that HGT spreads functional genes across the grass family. Initially, we sequenced the genome of the grass *Alloteropsis semialata* and identified 59 genes that were acquired as part of 23 horizontally transferred fragments of DNA from at least 9 different donors. The majority of the 59 HGTs in *A. semialata* are expressed, and we show that they have added to the genetic toolkit of the recipient genome. We have now expanded our HGT scans to other model grasses including many wild species in addition to widely grown crops (e.g. maize, sorghum, barley, rice and wheat). Our results show that HGT is widespread in the family, and can spread genes with diverse functions including photosynthesis, soil adaptation and disease resistance. We show that the rate of HGT differs among species, and these differences may be associated with certain phenotypic traits. Overall, our results demonstrate that HGT among certain eukaryotes can be significant, representing a recurrent source of material for adaptive diversification.

Biochemistry, epistasis and the evolutionary process

SMBE-PO-029

Emergence and propagation of epistasis in metabolic networks

S. Kryazhimskiy [1,*]

[1]Division of Biological Sciences, University of California, San Diego, La Jolla, United States

Abstract: The effect of a mutation on a phenotype of interest often depends on the presence of other mutations in the genome. Such dependencies are known as epistasis or genetic interactions. The evolutionary process fundamentally depends on the structure and type of these interactions. Certain types of epistasis are involved in explaining the evolution of sexual reproduction, historical contingency, robustness to deleterious mutations, etc. Epistasis is also extensively used in genetics to identify genes involved in various biological processes. Despite its prominent role in biology, our current understanding of the mechanistic origins of epistasis is poor, especially for mutations affecting different genes. In particular, we lack a null expectation for what types of epistasis (if any) should be common to many or even all biological systems and what types of epistasis may be signatures of potentially interesting idiosyncratic interactions between specific gene products.

Here, I develop a mathematical theory for understanding what types of epistasis we might expect to observe between mutations affecting microbial metabolism. I consider a hierarchy of increasingly coarse-grained descriptions of a metabolic network, such that more coarse-grained ("higher-level") descriptions typically have fewer effective parameters than more detailed ("lower-level") descriptions, with the growth rate being the single top-level parameter. I find that mutations that exhibit no epistasis for lower-level parameters (e.g., mutations affecting different enzymes) almost certainly exhibit epistasis for higher-level parameters, and that epistasis for lower-level parameters generically implies epistasis for higher-level parameters. This suggests that any metabolic mutations that have effects on growth rate are generically expected to exhibit epistasis. Moreover, I show that, for networks with first-order reaction kinetics, negative epistasis at a lower level remains negative at all higher levels, and strong positive epistasis at a lower level remains strongly positive at all higher levels. Finally, I show that certain topological relationships between reactions within the network impose constraints on the sign of epistasis for growth rate that mutations affecting these reactions can exhibit.

This theory provides a foundation for interpreting epistasis observed in experiments and for constructing more realistic models of genome-wide fitness landscapes.

Biochemistry, epistasis and the evolutionary process

SMBE-PO-028

To see or not to see: molecular evolution of the rhodopsin visual pigment in neotropical electric fishes

A. Van Nynatten*, F. H. Janzen, K. Brochu, J. A. Maldonado-Ocampo, W. G. Crampton, B. S. Chang, N. R. Lovejoy

Abstract: Functional variation in the dim-light-specialized visual pigment, rhodopsin, frequently occurs in species inhabiting light-limited environments. Variation in visual function can arise through two processes: relaxation of selection or adaptive evolution improving photon detection in a given environment. Here, we investigate the molecular evolution of rhodopsin in Gymnotiformes, an order of South American fishes with sophisticated electrosensory capabilities. These nocturnal fishes are thought to have poor vision, indicating a possible sensory trade-off between the visual system and a novel electrolocation sensory system. To test this idea, we surveyed rhodopsin from 147 gymnotiform species, spanning the order, and analyzed patterns of molecular evolution. In contrast to our expectation, we detected strong selective constraint in gymnotiform rhodopsin, with rates of non-synonymous to synonymous substitutions lower in gymnotiforms than in other vertebrate lineages. In addition, we found evidence for positive selection on the branch leading to gymnotiforms and on a branch leading to a clade of deep-channel specialized gymnotiform species. On the gymnotiform branch, positively selected sites include a substitution associated with visual disease in humans, but deleterious effects associated with this substitution are likely masked by epistatic substitutions at nearby sites. Our results suggest that rhodopsin remains an important component of the gymnotiform sensory system alongside electrolocation, and that photosensitivity of rhodopsin is well adapted for vision in dim-light environments. It is this functional flexibility of rhodopsin that has allowed adaptation to many different environments and appears to have cemented its place in vertebrate vision and in the overall integration of sensory information.

Biochemistry, epistasis and the evolutionary process

SMBE-PO-030

Mitochondrial mutational spectrum in vertebrates is shaped by generation time

A. G. Mikhaylova [1 2,*], A. A. Mikhaylova [1], K. Ushakova [1], E. Tretyakov [1], K. Popadin [3], K. Gunbin [4], I. Mazunin [1], D. Knorre [5], S. Nikolaev [6], A. Yurchenko [7], K. Khrapko [8], A. Reymond [3]

[1]Center for Mitochondrial Functional Genomics, Immanuel Kant Baltic Federal University, Kaliningrad, [2]The Vavilov Institute of General Genetics RAS, Moscow, Russian Federation, [3]Center for Integrative genomics, University of Lausanne, Lausanne, Switzerland, [4]Institute of Cytology and Genetics SB RAS, Novosibirsk, [5]The A.N. Belozersky Institute Of Physico-Chemical Biology, Moscow State University, Moscow, Russian Federation, [6]Gustave Roussy Cancer Center, Paris, France, [7]Institute of Biodiversity Animal Health and Comparative Medicine, University of Glasgow, Glasgow, United Kingdom, [8]Northeastern University, Massachusetts, United States

Abstract: To understand evolution, which is a function of mutagenesis and selection we have to split observed molecular changes of DNA sequences into ones, driven by mutation process and by selection. For example, it has been shown previously that long- versus short- lived mammals have GC rich mitochondrial genomes [doi:10.1089/rej.2008.0676.], but without additional investigation it is impossible to disentangle whether this is the result of positive or it is a neutral consequence of mutational bias. In order to estimate and compare the mutational spectra for the maximal number of vertebrate species we focused on mitochondrial genes, extensively used in numerous ecological, evolution and population genetics studies. It has been shown that mitochondrial mutational spectrum varies between species [doi:10.1101/gr.3128605]. We reconstructed mtDNA mutations from more than 2000 vertebrate species and revealed two principal mtDNA mutational processes: temperature sensitive transition G>A and generation-time/cell-division-rate sensitive T>C. Altogether they explain more than 50% of all mtDNA mutations.

For each vertebrate species we collected all available intra-species protein-coding mtDNA sequences, reconstructed within-species phylogeny using an outgroup, inferred ancestral sequences, derived a list of polarized single-nucleotide substitutions and, focusing on the nearly neutral synonymous fourfold degenerate sites, we observed strong excess of transitions over transversions.

Further focusing on mammalian species we used generation length as a metric, which is both well known and is associated with numerous ecological and physiological parameters. For 424 mammalian species we observed positive correlation between generation length [doi:10.3897/natureconservation.5.5734] and Ts/Tv. This correlation was driven mainly by transition T>C which positively correlated with the generation time while several rare transversions negatively correlated with generation time.

In order to derive mutational signatures in unsupervised way we performed principal component analysis of mutational spectra of 424 mammalian species. We observed that the first component is mainly driven by G>A substitutions while the second is driven mainly by T>C substitutions. Interestingly, the correlation of the second principal component with generation length was significantly higher as compared to the sole effect of T>C and Ts/Tv, meaning that the second principal component may reflect a complex signature of a mutagen associated with generation length. Majority of T>C mutations are results of spontaneous deamination of adenine to hypoxanthine on single-stranded heavy chain during replication and we hypothesize that increased chemical damage in long-lived species may be driven by either more aerobic conditions in long-lived animals or due to increased pH inside mitochondria of oocytes of long-lived animals, or due to increased chemical damage in oocytes of short-lived mammals.

Mutational spectrum may affect long-term changes in nucleotide content in the absence of selections. To test it we used 3464 complete mitochondrial genomes of vertebrate species and for each species estimated nucleotide content of the most neutral sites. The most common transition T>C according to our results is stronger in long-lived species. Testing all pairwise correlations between generation time and neutral nucleotide content we observed two the strongest correlations: positive with C and negative with T.

Biochemistry, epistasis and the evolutionary process
SMBE-PO-027
A Model to Study Distribution of Fitness Effects in Genetic Networks
R. G. Brajesh [1,*], S. Saini [1]
[1]Department of Chemical Engineering, Indian Institute of Technology Bombay, Mumbai, India

Abstract: Mutations give rise to genetic variation leading to an increase in fitness of the population. The nature of these mutations can be harmful, advantageous or neutral for the organism. A number of computational and theoretical studies report distribution of beneficial and deleterious mutations in organisms/systems. However, a precise, quantitative understanding of these distributions is currently not well understood**.** In this study, we consider lactose utilization system in *E. coli* and define its fitness with the help of cost-benefit analysis, later we develop computational framework to employ mutations in the system and evaluate distributions linked to beneficial or deleterious mutations. We use this framework to answer two specific questions, how does the distribution of mutations changes with respect to current fitness of the system and how this distribution varies between multiple parameter sets associated with the same fitness? Our study shows that beneficial mutations can be accurately represented by an exponential distribution at any fitness level. However, deleterious mutations can be represented by exponential distributions only when the system is at low fitness and cannot be represented by any standard distributions when the system is at intermediate or high fitness level. With the help of our computational framework, we also explore the epistatic interactions between beneficial mutations and quantitatively show that the spread of effect of beneficial mutation is drastically different in different genetic background. Overall, we provide a computational framework to explore the distribution of beneficial and deleterious mutations and the dynamics of beneficial mutations in a controlled laboratory setup.

Biochemistry, epistasis and the evolutionary process

SMBE-PO-042

Implications of epistasis on protein evolution: evolutionary Stokes and anti-Stokes shifts.

N. Youssef [1,*], C. Jones [2], E. Susko [2 3], J. Bielawski [1 2 3]

[1]Department of Biology, [2]Department of Mathematics and Statistics, [3]Center for Comparative Genomics and Evolutionary Bioinformatics, Dalhousie University, Halifax, Canada

Abstract: The influence of epistasis on the evolution of proteins remains largely unknown and has been the subject of recent controversy. Here we implement a mechanistic thermostability informed model (TSIM) to investigate the short and long-term impact of site interactions on protein evolution. We explore trajectories in sequence space under three regimes: Mutation-Drift (MD), Mutation-Selection (MS), and Mutation-Drift-Selection (MDS). We find that the degree of epistatic interactions is heavily dependent on the evolutionary regime. Under MD, sites evolve virtually independently of each other. By contrast, the inclusion of a selective force (MS and MDS) results in an interdependent sequence where each substitution affects the amino acid preferences at all other sites. An important prediction of the TSIM is the *evolutionary Stokes shift,* which refers to an increase in the preference of the resident amino acid due to compensatory substitutions at other sites. We observe and define a reciprocal phenomenon, the *evolutionary anti-Stokes shift*, meaning a decrease in the preference of the resident amino acid due to replacements at other sites. Over short time scales, the proportion of induced Stokes (P_{Stokes}) and anti-Stokes ($P_{antiStokes}$) shifts are equal under all evolutionary regimes albeit at different values (approximately 1% under MD, and 50% under MS and MDS). Given longer adjustment times, the P_{Stokes} and $P_{antiStokes}$ differed depending on the interplay between drift and selection. We show that these observations are an epiphenomenon of the underlying distribution of substitution selection coefficients.

Biochemistry, epistasis and the evolutionary process

SMBE-PO-031

The strategy of reproduction and its role in Y chromosome evolution

D. Mackiewicz [1,*], P. Posacki [1], M. Burdukiewicz [1], P. Błażej [1]

[1]Department Bioinformatics and Genomics, University of Wroclaw, Wrocław, Poland

Abstract: Sex determination in mammals is based on sex chromosomes. Females possess two copies of X chromosome, whereas males have one X and one Y chromosome. It is well known that sex chromosomes are derived from ordinary autosomes. Despite their shared ancestry, Y chromosome displays enormous degeneration over evolutionary time, whereas its X counterpart maintains most of ancestral genes. It has been suggested that the degeneration process started with the acquisition of a male-determining gene. Next, sex-specific genes became linked to this gene, and heterozygous sex evolved because it was profitable for these genes to be inherited together. It was possible by suppression of recombination, which might have been achieved through chromosomal inversions on the proto-sex chromosomes. However, the other reasons for the Y degeneration should be taken account, e.g. the strategy of reproduction in mammals. To study all these aspects, we applied a more general and advanced computer simulation model in which the recombination rate between the sex chromosomes can freely evolve and individuals can create unfaithful or faithful pairs. We found that only under the unfaithfulness of mates the number of females increases at the expense of males in the evolving population and the accumulation of mutations on the Y chromosome occurs. Therefore, the recombination rate between the X and Y decreases very quickly and the Y degenerates. Thus, the X chromosomes are cleaned off defective alleles and the reproduction potential of population, measured by the number of females, is not reduced. The simulation showed that the suppression of recombination between X and Y chromosomes is spontaneous and does not require inversions.

Biochemistry, epistasis and the evolutionary process

SMBE-PO-033

The chemical arms race between Heliconius butterflies and their Passiflora host

E. De Castro [1,*], J. Musgrove [2], O. McMillan [2], S. Bak [3], C. Jiggins [1]

[1] Zoology Department, University of Cambridge, Cambridge, United Kingdom, [2]Smithsonian Tropical Research Institute, Gamboa, Panama, [3]Department of Plant and Environmental Sciences, University of Copenhagen, Copenhagen, Denmark

Abstract: Understanding how the interactions between species promote their diversification is one of the most fascinating questions in evolutionary biology. *Heliconius* butterflies and *Passiflora* plants are classical examples of coevolution. Curiously, both butterfly and plant contain toxic cyanogenic glucosides (cyanogens). *Heliconius* biosynthesize cyanogens, but can also sequester simple cyanogens from plants of the genus *Passiflora,* where over 30 cyanogen structures have been reported. Although all *Heliconius* species utilize only *Passiflora* plants as hosts, different degrees of specialization are observed in this relationship: some *Heliconius* utilized preferentially one or few specific *Passiflora* species whereas others are generalist. In this talk, I will discuss how biosynthesis and sequestration of cyanogens has shaped the arms race between heliconiine butterflies and *Passiflora* plants. By performing metabolomics analyses of 22 heliconiines and over 50 *Passiflora* species, we discovered that some of these plants have decorated their cyanogens structures with unusual sugars and phosphate groups to avoid sequestration and evade herbivory by heliconiines. *Heliconius* species are more specialized in sequestration than other heliconiines which are overall more generalist. Moreover, I will also explain the role of host-plant specialization in the sympatric speciation of *H. melpomene rosina* and *H. cydno chioneus* in Panama. Larval-performance experiments revealed that diet affects size and weight of these butterflies, and *H. cydno* has the worst perfomance when reared on the favourite host of *H. melpomene.*

Biochemistry, epistasis and the evolutionary process

SMBE-PO-032

The Role of Gene Duplication in the Evolution of the Glucosinolate Biosynthesis Pathway

R. S. Abrahams [1 2,*], M. Kerstens [2], K. Bouwmeester [2], J. C. Pires [1], E. Schranz [2]

[1]University of Missouri, Columbia, United States, [2]Wageningen University , Wageningen, Netherlands

Abstract: Glucosinolates (Mustard Oils) are a class of diverse defense compounds found in plants of the order Brassicales and are considered a key innovation. With over 120 known compounds, their diversity and evolutionary history has obscured by patterns of convergence and parallel evolution. Some conflicting evidence exists around the role and significance of different duplication events in the evolution of the group and the majority of glucosinolate research has been done on a intra-species context without reference to phylogeny. In this study we use a phylogenomic synteny network approach, using over 35 plant genomes to investigate patterns of gene family expansion in the biosynthetic pathway. Here we shed insight on the evolution of this complex group, revealing lineage specific patterns and novel insights.

Biochemistry, epistasis and the evolutionary process

SMBE-PO-037

Ancestral reconstruction and experimental testing of evolvability using the AP-superfamily as a model system

B. Eenink [1,*], M. Heberlein [1], T. Kaminski [2], C. Dilkaute [3], J. Jose [3], F. Hollfelder [2], E. Bornberg-Bauer [1], B. van Loo [1]

[1]Institute for Evolution and Biodiversity, University of Münster, Münster, Germany, [2]Department of Biochemistry, University of Cambridge, Cambridge, United Kingdom, [3]Institute for Pharmaceutical and Medical Chemistry, University of Münster, Münster, Germany

Abstract: Natural protein evolution has resulted in a diverse array of structures catalyzing diverse reactions with unrivaled efficiency. To emulate evolution *in vitro*, we generate mutant libraries of individual genes using error-prone PCR, and heterologously express these mutant libraries in an *E. coli*. A fundamental understanding of the natural evolution of enzymes would be essential to harness this knowledge in order to devise improved strategies for directed evolution campaigns. In particular how many and which combinations of mutations are needed for desired improvements remains a mystery. Furthermore, these improvements may not be reachable via one-by-one mutational steps due to the occurrence of epistatic ratchets, raising the question how evolutionary dead ends can be bypassed. In addition to these considerations, to identify the optimal starting point for laboratory evolution, the shape of the local fitness landscape could be even more critical than the initial activity levels of an enzyme.

We investigate the effect of mutational load on extant and reconstructed ancestral members of the Alkaline Phosphatase (AP) superfamily. We experimentally explore sequence space around extant and resurrected ancestral enzymes in parallel using micro-droplet-based high-throughput screening methods. We attempt to connect 'fitness parameters', i.e. catalytic efficiency, enzyme stability and the occurrence of epistatic effects, of both extant and ancestral enzymes to the shape of their local fitness landscapes under different mutational loads. This will provide quantitative insight into how these properties are contributing to shaping the trade-off between robustness and evolvability, and how these differ between extant and ancestral enzymes.

Biochemistry, epistasis and the evolutionary process

SMBE-PO-039

Evidence of compensatory epistasis when comparing the full distribution of fitness effects of new amino acid mutations across great apes

D. Castellano [1,*], M. Coll Macià [1], P. Tataru [1], T. Bataillon [1], K. Munch [1]

[1]Aarhus University, Aarhus, Denmark

Abstract: The distribution of fitness effects (DFE) is central to many questions in evolutionary biology. However, little is known about the differences in DFEs between closely related species. In this work, we use more than 9,000 coding genes orthologous one-to-one across great apes, gibbons, and macaques to assess the stability of the full DFE across great apes. We use the information contained in the unfolded site frequency spectrum of polymorphic mutations to estimate the full DFE. First, we find that the shape of the deleterious DFE is likely shared across great apes as expected if protein function is similar in these species. Second, our proxy for the species effective population size, synonymous nucleotide diversity, is a very strong predictor of the strength of negative selection consistent with the Nearly Neutral Theory. Third, we find that the rate of new beneficial mutations is higher in great ape populations with low levels of genetic diversity. This is in agreement with models of compensatory epistasis which predict that, in small populations, the proportion of new weakly beneficial compensatory mutations is greater than in large populations. Compensatory mutations improve fitness in genotypes that contain deleterious mutations but have no beneficial effects otherwise. All these results are replicated using only GC-conservative mutations and we find that GC-biased gene conversion is not affecting our main conclusions. We conclude that while the deleterious DFE is likely conserved between closely related species, the proportion and the DFE of beneficial variants is more dynamic and affected by compensatory epistasis.

Biochemistry, epistasis and the evolutionary process

SMBE-PO-040

Investigations of natural and disease-associated variants on rhodopsin structure and function

M. G. Woo [1,*], B. S. Chang [1 2]

[1]Cell and Systems Biology, [2]Ecology and Evolutionary Biology, University of Toronto, Toronto, Canada

Abstract: Rhodopsin is a critical class A G protein-coupled receptor (GPCR) responsible for mediating the first step in dim-light vision. Epistatic interactions in rhodopsin are particularly interesting as residues at specific sites are observed to give rise to varying degrees of altered function in certain species while in others, function appears to be unaffected. Due to its critical role in vision, rhodopsin is highly conserved in nature, however over 150 mutations in this gene have been found to be associated with hereditary retinal disease, such as Retinitis Pigmentosa (RP). Although a large number of these disease-associated mutations have been characterized, many remain poorly studied in terms of fundamental mechanistic effects on structure-function. Here, we use site-directed mutagenesis to investigate the functional impact of disease-causing sites on rhodopsin structural stability and biochemistry, as well as the role of potential compensatory mutations or intramolecular epistasis at sites in proximity to the disease sites. We utilize a comparative approach to identify natural variants at proposed compensatory sites in vertebrate rhodopsins, introduce those mutations into both bovine and human background and characterize protein fitness using spectroscopic assays and computational modeling of rhodopsin structure. These findings allow us to gain further insight into how protein function and structural stability have been achieved, and moreover, the underlying molecular mechanisms that may be contributing to inherited visual diseases such as RP.

Biochemistry, epistasis and the evolutionary process

SMBE-PO-045B

Nitrogen availability and plant genome evolution

J. Y. Choi [1,*], G. Eshel [1], R. Gutierrez [2], G. Coruzzi [1], M. Purugganan [1]

[1]New York University, New York, United States, [2]Pontificia Universidad Católica de Chile, Santiago, Chile

Abstract: Since elements such as nitrogen, found as nutrients for organisms, are often limited in nature, they are believed to impact the evolution of the molecular composition of transcriptomics and proteomic sequences. This selective pressure would especially affect plants as they rely heavily on environmentally available nitrogen. Past studies, however, have been equivocal on whether the molecular constraint arising from ecologically-limited nitrogen levels can affect the molecular sequence evolution in plants. We examined the genome-wide coding and protein sequences from 113 plant species from artificially or naturally nitrogen poor and rich environments, and tested whether plants from nitrogen poor environments had higher selective pressure to conserve nitrogen at the molecular level. In the end, we did not find any consistent evidence to suggest ecological nitrogen availability is driving genome evolution in plants. We discovered, however, that gene expression patterns were strongly associated with genome-wide nitrogen content, suggesting that sequence composition in plants is likely to be physiologically- and not ecologically-driven.

Biochemistry, epistasis and the evolutionary process

SMBE-PO-034

Experimental assessment of predicted coevolving amino acids in Escherichia coli

M. B. Haider [1,*], J. Y. Dutheil [1]

[1]Department of Evolutionary Genetics, Max Planck Institute for Evolutionary Biology, Plön, Germany

Abstract: Amino acids within a protein sequence interact to maintain its structure and function. Consequently, mutations at a given functionally significant position can be potentially compensated by a mutation at an interacting position. For instance, a “small-to-big” substitution from an amino acid with a small side chain to an amino-acid with a big side chain can be compensated by a reciprocal “big-to-small” substitution. Such coevolution scenario implies that the first mutation leads to a fitness reduction, while the compensating mutation restores it. Several methods have been developed to detect coevolving positions from sequence alignments. The validation of the resulting predictions, however, relies so far only on indirect evidence such as residue contact maps in proteins for which an experimental structure is available. In this study, we mapped substitutions for a protein sequence alignment on each branch of the phylogeny, to detect coevolving amino-acids in bacterial homologous protein families. Accounting for the biochemical properties of amino acids, we identified thousands of coevolving groups. We then selected candidate groups displaying a pattern of co-substitutions in the *Escherichia coli* lineage, and experimentally reconstructed the local adaptive landscape, resurrecting the ancestral genotype and comparing its fitness to single and double mutants (the reconstructed ancestral genotypes). We short-listed 13 candidate groups of coevolving positions in essential, non-essential and conditional genes. We introduced the corresponding mutations in the E. coli genome using homologous recombination and assessed the local fitness with growth curves and competition experiments between all genotypes (single mutants, double mutant, and wildtype). This study provides the first experimental assessment of an evolutionary method predicting coevolving positions in protein sequences.

Biochemistry, epistasis and the evolutionary process
SMBE-PO-038
Evolution and ecology in 3D culture of cancer cells
P. Mukashyaka [1], J. Noorbakhsh [1,*], P. Kumar [1], E. Courtois [1], O. Anczukow [1], P. Robson [1], E. Liu [1], J. Chuang [1]
[1]The Jackson Laboratory For Genomic Medicine, Farmington, CT, United States

Abstract: Cancer cells within the same tumor evolve and compete for shared resources and may also respond differently to drug treatments. In previous work [1], we found two genetically distinct coexisting subclones within breast cancer samples with different sensitivity to chemotherapy. To explore the dynamics of competition and cooperation between these subclones, we derived cell lines from each and tagged them with fluorescent markers. We then co-cultured these cell lines in 3D organoids and tracked them using confocal microscopy. To determine the location of each organoid and the cells within it, we developed a 3D cell segmentation procedure. Our procedure relies on a two-step process where organoids are segmented first and then individual cells are detected within each organoid using a combination of standard filtering, thresholding, and watershedding. Preliminary analysis of hundreds of organoids suggests that in the absence of treatment the two cell types preserve their relative ratio regardless of organoid size, suggesting that the system is in a state of ecological coexistence. We will present results on cellular mixing during growth and treatment and the implications for adaptive therapy ecologies in tumors.

1. *Kim, Hyunsoo, et al. "High-resolution deconstruction of evolution induced by chemotherapy treatments in breast cancer xenografts." Scientific reports 8.1 (2018): 17937*

Biochemistry, epistasis and the evolutionary process
SMBE-PO-035
An evolutionary path: from hydrolase to ligase by substitution
N. Hecht [1,*], E. Gur [1 2]
[1]Life Sciences, [2]The National Institute for Biotechnology in the Negev, Ben-Gurion University, Beer-Sheva, Israel

Abstract: The understanding of how proteins evolve to perform novel functions has long been sought by biologists. Two homologous bacterial enzymes, PafA and Dop, pose an insightful case study, as both rely on similar mechanistic properties, yet catalyze distinct and opposite reactions. PafA catalyzes a ligation reaction of a small protein tag to target protein substrates, whereas Dop removes the tag by hydrolysis of the iso-peptide bond between the tag and substrate. Given that both enzymes maintain a similar fold and high sequence homology, we wondered what are the critical differences in amino acid sequence and folding that are responsible for each distinct activity. We tackled this question using horizontal and vertical analysis of sequence-function relations, and identified a set of uniquely conserved residues in each enzyme. Reciprocal mutagenesis of the hydrolase, Dop, completely abolished the native activity, at the same time yielding a catalytically active ligase. Our findings suggest that the change of activity is a result of a conformational change resulting with a different geometry of the active site. Further analysis revealed conserved residues to be essential for stabilization of this alternative conformation, rather than effecting the catalytic mechanism directly. Overall, our analysis captures in mechanistic and evolutionary detail the changes required for the emergence of a new catalytic function from a preexisting one.

Biochemistry, epistasis and the evolutionary process

SMBE-PO-045

The non-neutral impact of synonymous mutations at the protein and cellular levels

I. G. Bravo [1,*], F. Borvetö [1], J. Daron [1], A. Demange [1], F. Leblay [1], A. Willemsen [1], M. Picard [1]

[1]Centre National de la Recherche Scientifique, Montpellier, France

Abstract: Synonymous codons encoding for the same amino acid are not used at random. Instead, codon usage preferences vary between positions along a gene, between genes within a genome, and between species. The overall balance between mutational biases and selection forces at each of these integration levels shapes codon usage preferences.

We have aimed here at quantifying the impact of codon usage preferences on protein performance and on cellular fitness.

At the protein level, we have designed and synthesised synonymous versions of the commonly used reporter gene encoding for the secreted alkaline phosphatase (seap) differing exclusively on the codon used to encode for the conserved D-S residues in the active site of the enzyme (twelve combinations). We have transfected human cells, analysed DNA and mRNA seap levels, and measured phosphatase activity in the supernatant. Our results show that protein activity levels are different depending on the actual combination usedto encode for the enzyme active center, even after accounting for DNA and mRNA levels.

At the cellular level, we have designed and synthesised synonymous versions of the *Shble* gene, conferring resistance to the antibiotic bleomycin, fused to a fluorescent reporter. We have transfected human cells, analysed DNA and mRNA *Shble* levels, quantified cellular fluorescence, analysed transcriptome and proteome, and finely measured cell growth and proliferation in the presence and in the absence of antibiotic. Our results show that both SHBLE levels and cellular fitness strongly depend on the codon usage preferences of our focal gene Shble.

We will discuss the implications and fitness costs differentially associated to synonymous codons preferences in terms of protein functioning, cellular fitness as well as the alternative evolutionary solutions to compensate for such fitness costs, with or without modification of such codon usage preferences.

Biochemistry, epistasis and the evolutionary process

SMBE-PO-026

The effect of phylogenetic branch length on protein structural properties

C. A. Weibel[1 2,*], L. Kosinski[3], J. Masel[4 5 6 7]

[1]Physics , [2]Mathematics, [3]Molecular and Cellular Biology, [4]Ecology and Evolutionary Biology, [5]Graduate Interdisciplinary Programs (GIDP), [6]BIO5 Institute, [7]Biochemistry and Cellular Molecular Biology , University of Arizona, Tucson, United States

Abstract: Prior work based on phylostratigraphy suggests that proteins become more structurally ordered with age. If longer phylogenetic branch lengths result in species that are “more evolved”, then we predict lower intrinsic structural disorder (ISD) of proteins in species residing at the end of a long branch, compared to orthologous proteins in species at the end of a short branch. We identified twelve phylogenetically independent pairs of species which have been shown in the literature to have one member of the pair evolving faster than the other. Ensembl-annotated orthology is available for five of the pairs, and we are in the process of annotating the orthologs of the remaining seven pairs using a BLASTp reciprocal best hits algorithm. We are running IUPred2 to predict ISD for the coding sequences and will build mixed linear models to test whether orthologous genes have different ISD in different species. W e are also exploring other sequence-based metrics of protein properties, such as the clustering of hydrophobic amino acids along the primary sequence, or protein aggregation propensity, that show long-term trends as a function of protein age.

Biochemistry, epistasis and the evolutionary process
SMBE-PO-043
Pairwise and higher-order epistatic interactions within gene regulatory networks
F. Baier*, Y. Schaerli

Abstract: Epistasis, the context-dependency of mutations, is pervasive in biological systems and severely influences evolutionary outcomes and their predictability. Despite its importance experimental studies on how epistasis influences the functional interactions within gene regulatory networks (GRNs) remain elusive, because of difficulties to perform and interpret such experiments in natural systems.
Here, we employ synthetic three-node GRNs, which translate a morphogen gradient into a "stripe" gene expression pattern (low-high-low) in a population of *Escherichia coli* cells. In particular, we systematically combine and characterize pairwise and three-way combinations of 10 x 10 x 10 mutants of the three nodes, which separately exhibit only slight differences in expression levels and maintain a "stripe-forming" output.
We find that much of the phenotypic variation, in terms of gene expression levels at various morphogen concentrations, is due to epistatic interactions, defined as a significant deviation from a multiplicative model. Epistatic interactions increase from pairwise to three-way combinations and also change with the morphogen concentration, meaning that the magnitude and sign of the epistatic interactions depends also on the environment. Interestingly, such high-order epistatic interactions between genotype and environment result in novel "non-stripe" patterning phenotypes, which would not be accessible in a purely multiplicative scenario without epistasis.
Using a mathematical model, we provide a mechanistic explanation for epistasis between genotype x genotype and genotype x environment interactions, which suggests that most epistatic interactions might be predictable as long as the networks topology and regulatory mechanisms are sufficiently understood.
Our work provides exciting insights into the mechanistic causes and evolutionary consequences of epistasis in GRNs.

Biochemistry, epistasis and the evolutionary process
SMBE-PO-041
Effects of an RNA chaperone on mutation tolerance
V. Soo [1,*], T. Warnecke [1]
[1]MRC London Institute of Medical Sciences, London, United Kingdom

Abstract: Due to their intrinsic thermodynamic properties, RNA can misfold easily in cells. One way to mitigate RNA misfolding is through the actions of RNA chaperones, which bind and unwind structured RNA molecules and thereby offer opportunities for these misfolded species to refold properly. Such rescue activity has implications for the fitness effects of individual mutations – at least mutations that compromise RNA folding or structure might be buffered by RNA chaperones. However, little is known about the rules governing such mutation buffering. Here, we describe how a model RNA chaperone, the DEAD-box RNA helicase CYT-19, affects the fitness effects of mutations in a model structured RNA, the *Tetrahymena* group I intron, whose self-splicing activity is dependent on its structure. We performed deep mutational scanning on the P1ex region of the intron which is critical for its self-splicing activity, and assayed differential splicing activity of all possible P1ex mutants in the presence and absence of CYT-19 to identify mutations that are buffered by RNA chaperone activity. I will discuss the properties of the chaperone-dependent and chaperone-independent mutation pools. Our results highlight that, to understand RNA robustness *in vivo*, we need to consider how mutational fitness effects are modulated by RNA chaperones and other *trans*-acting factors.

Contemporary Evolution

SMBE-PO-069

Genomic introgression in Oreochromis tilapias

A. Ciezarek [1,*], A. Ford [2], T. Mehta [1], W. Nash [1], L. Penso-Dolfin [1], G. Turner [3], M. Genner [4], F. Di Palma [1 5], W. Haerty [1]
[1]Organisms and Ecosystems, Earlham Institute, Norwich, [2]Department of Life Sciences, University of Roehampton, London, [3]School of Natural Sciences, Bangor University, Bangor, [4]School of Biological Sciences, University of Bristol, Bristol, [5]Norwich Medical School, University of East Anglia, Norwich, United Kingdom

Abstract: The *Oreochromis* tilapias are an economically important group of fish for aquaculture, whose production has expanded dramatically in the last two decades. A direct consequence of this success has been the introduction of exotic species in Tanzania, a hotspot for *Oreochromis* diversity. Introductions have had significant negative ecological effects on indigenous *Oreochromis* species, including species displacement and loss of population structure through hybridisation, having potential implications for local adaptations. This history of introgression has made untangling the population history of *Oreochromis* difficult. We address this issue using genome-wide sequencing data across seven *Oreochromis* species, identifying strong signatures of introgression between species at both the population and individual level, although each species has remained phylogenetically distinct. We quantified the impact of gene flow on genomic diversity and differentiation, and investigated whether some genomic regions are resistant to introgression. Such regions may maintain adaptive variation, and therefore be important for population viability. We anticipate that our results will have important implications when managing the translocation of species.

Contemporary Evolution

SMBE-PO-065

DNA methylation as a mechanism of transgenerational phenotypic plasticity in Daphnia.

M. Villalba [1,*], R. Hager [1], L. Weiss [2], R. Tollrian [3]

[1]Evolution and Genomic Sciences, University of Manchester, Manchester, United Kingdom, [2]Animal Ecology, Evolution and Biodiversity, Ruhr-Universitat, Bochum, [3]Animal Ecology, Evolution and Biodiversity, Ruhr-Universitat Bochum, Bochum, Germany

Abstract: Phenotypic plasticity is defined as the ability of a given genotype to differentially respond to different environmental challenges. Organisms may respond quickly to environmental changes within one generation, but transgenerational plasticity has also been observed. The freshwater crustacean Daphnia is well known for its high degree of plasticity. For example, it responds to an increased predation risk by developing defensive morphological features. Transgenerationally, the clonal offspring of defended mothers display even stronger defences when the predation risk remains high. As Daphnia reproduce via parthenogenesis, this transgenerational plasticity may be via epigenetic mechanisms. In our study we sought to elucidate the molecular mechanisms of this non-genetic inheritance of predator induced phenotypic plasticity, focussing on differential DNA methylation in two species of daphnia (D. lumholtzi and D. pulex). We compare DNA methyltransferase (DNMT) activity by measuring differential gene expression of the DNMT genes in two successive generations of predator exposed and control daphnids. This will provide first insights into whether DNA methylation is involved in enabling transgenerational plasticity in Daphnia.

Contemporary Evolution

SMBE-PO-046

Stacks v2: Novel and improved methods to leverage RAD-seq data in a wide range of study systems

N. Rochette [1,*], A. Rivera-Colon [1], J. Catchen [1]

[1]Animal Biology, University of Illinois at Urbana-Champaign, Urbana-Champaign, United States

Abstract: Restriction-associated DNA sequencing (RAD-seq) is an eminently popular approach for population genomics, phylogenetics, and linkage analyses. Although genome-wide resequencing is gradually becoming more affordable, RAD-seq approaches still benefit from considerably lower library preparation costs, and from scaled-down computational resource and expertise needs. We therefore expect that RAD-seq will remain a profoundly relevant strategy for all study systems that do not benefit from extensive genomic resources, for studies that involve hundreds or thousands of individuals, and as a means to inexpensively survey promising novel systems. In addition, RAD-seq data analysis methods have received much less attention than genome-wide ones, and considerable methological improvements remain possible.

Here, we introduce a series of major improvements to the prominent RAD-seq data analysis pipeline, Stacks. These new developments improve the results obtained with all RAD-seq protocols, and dramatically expand the potential of the widely used sdRAD and bestRAD protocols, especially for *de novo* approaches. First, we substantially improve the existing core elements of the pipeline, including the clustering of RAD loci, the alignment of reads, and the SNP-calling model. Second, we introduce a de Bruijn graph-based method that makes it possible and practical to assemble a long 500-700bp mini-contig (up from 100-150bp) for each RAD locus from paired-end sequencing RAD-seq data. Furthermore, noting that these longer loci typically cover several SNPs, we describe an accurate method to phase SNP alleles within a RAD locus to yield multi-state haplotype markers that contain more information than the individual SNPs they comprise, and that may readily be used in many phylogeography and demography analyses.

We demonstrate the great effectiveness of these methods using simulations and re-analyses of datasets previously published for fish and bird systems. Finally, we illustrate the new ways in which they allow RAD-seq data to be used. For instance, longer RAD loci built *de novo* for the system of interest can be more easily mapped to any available genomic resources within the genus or family. Additionally, we note that assembling a contig for each RAD locus implies that it is natively possible to remove PCR duplicates in the sdRAD and bestRAD protocols. Building on this property, we characterize the levels of PCR duplicates arising in these protocols, then propose and demonstrate experimentally a mechanistic explanation for the elevated levels of PCR duplicates observed, and show how to adjust the existing molecular protocols to suppress this behavior.

The work presented here provides novel, widely applicable tools to further leverage RAD-seq data, and highlights the versatility and usefulness of the RAD-seq design in all biological systems.

Contemporary Evolution

SMBE-PO-048

Effects of pesticide exposure on gene expression in the bumblebee Bombus terrestris

F. Lopez-Osorio [1,*], T. J. Colgan [1], I. K. Fletcher [1], A. N. Arce [2], R. J. Gill [2], A. Ramos Rodrigues [2], E. Stolle [3], L. Chittka [1], Y. Wurm [1]

[1]School of Biological and Chemical Sciences, Queen Mary University of London, London, [2]Department of Life Sciences, Imperial College London, Ascot, United Kingdom, [3]Institut für Biologie, Martin-Luther-Universität Halle-Wittenberg, Halle, Germany

Abstract: Insect pollinators such as bees play a major role in the maintenance of biodiversity and provide an essential service in agricultural productivity. Worldwide declines of insect pollinators, however, pose major challenges and concerns. These declines have been the outcome of several interplaying factors, such as exposure to pesticides and habitat loss. The widespread use of pesticides has received special attention due to their neurotoxic effects on bees. However, our understanding of the molecular pathways involved in the response to pesticide exposure remains limited. Here, we use a transcriptomics approach to evaluate the effects of pesticide exposure in the bumblebee *Bombus terrestris*. We chronically exposed colonies of *B. terrestris* to a total of fifteen pesticide treatments, including combinations of compounds and controlling for factors such as colony social environment and worker age. We reveal that genes involved in a variety of metabolic processes are differentially expressed in response to pesticide exposure, and that there are differences in the effects of pesticide combinations compared to single compounds. Our study shows that transcriptional profiles enable the discovery of unique and common responses to different pesticide classes in *B. terrestris*.

Contemporary Evolution

SMBE-PO-068

Genetics of the PanAf project: chimpanzee diversity from fecal samples

C. Fontsere [1,*], E. Lizano [1], M. Alvarez-Estape [1], P. Gratton [2], J. Lester [2], D. A. Hughes [3], M. Kuhlwilm [1], M. de Manuel [1], F. Calvet [1], P. Dieguez [2], T. Aebischer [4], A. Agbor [2], P. Alvarez Varona [5], S. Angedakin [2], A. K. Assumang [6], F. Aubert [7], E. A. Ayimisin [2], E. Bailey [2], D. Barubiyo [2], M. Bessone [2], M. Bonnet [8], G. Brazzola [2], R. Chancellor [9], H. Cohen [2], K. Corogenes [2], C. Coupland [2], E. Danquah [6], B. Curran [10], E. Dilambaka [11], V. E. Egbe [2], H. Eshuis [2], T. Deschner [12], D. Dowd [13], A. Dunn [14], J. Dupain [15], D. Hedwig [16], V. Hermans [17], S. Jones [2], P. Kadam [18], A. Goedmakers [19], A.-C. Granjon [20], M. Kaiser [2], J. Head [21], M. Kambi [2], M. V. Kambale [2], I. Imong [22], I. Kienast [2], J. Lapuente [2], K. J. Jeffery [23 24], B. Larson [2], A. Laudisoit [25], K. Lee [2], V. Leinert [7], M. Kambere [26], K. Langergraber [27], D. Kujirakwinja [14], V. Lapeyre [28], S. Marrocoli [2], R. Martin [2], F. Mulindahabi [11], M. Murai [2], A. Meier [2], G. Maretti [26], S. Nixon [29], E. Neil [2], C. Orbell [30], L. J. Ormsby [2], A. Piel [31], S. Regnaut [7], C. Sanz [32], L. Sciaky [2], P. Niyigabae [11], D. Morgan [33], A. Tickle [2], E. Normand [34], J. van Schijndel [2], E. Terrade [2], E. Ton [19], J. Willie [17], Y. G. Yuh [2], K. Yurkiw [2], J. Preece [26], M. Robbins [35], L. Vigilant [2], F. A. Stewart [36], C. Boesch [2 7], A. Rundus [37], H. Kühl [2 38], M. Arandjelovic [2], T. Marques-Bonet [1 39 40 41], V. Sommer [42], L. Pacheco [5], N. Tagg [43], H. Vanleeuwe [14], V. Vergnes [34], R. M. Wittig [44], K. Zuberbuehler [45], R. McElreath [46]

[1]Institut de Biologia Evolutiva (UPF-CSIC), Barcelona, Spain, [2]Max Planck Institute for Evolutionary Anthropology (MPI EVA), Leipzig, Germany, [3]MRC Integrative Epidemiology Unit, University of Bristol, Bristol, United Kingdom, [4]University of Fribourg, Fribourg, Switzerland, [5]Instituto Jane Goodall Espana, Barcelona, Spain, [6]Department of Wildlife and Range Management, Faculty of Renewable Natural Resources, Kwame Nkrumah University of Science and Technology, Kumasi, Ghana, [7]Wild Chimpanzee Foundation (WCF), Leipzig, Germany, [8]The Aspinall Foundation, Port Lympne, United Kingdom, [9]Depts of Anthropology & Sociology and Psychology, West Chester University, West Chester, United States, [10]The Aspinall Fondation, Port Lympne, Germany, [11]Wildlife Conservation Society (WCS), New York, United States, [12]Max Planck Institite for Evolutionary Anthropology, Leipzig, Germany, [13]Wild Chimpanzee Foundation, Leipzig, , [14]Wildlife Conservation Society, New York, [15]Africa Wildlife Foundation, Washington DC, United States, [16]The Aspinall Foundation, Kent, United Kingdom, [17]KMDA, Centre for Research and Conservation, Royal Zoological Society of Antwerp, Antwerp, Belgium, [18]University of Cambridge, Cambridge, United Kingdom, [19]Chimbo Foundation, Amsterdam, Netherlands, [20]Max Planck Institute for Evolutionary Anthropology, Leipzig, , [21]Max Plack Insttitute for Evolutionary Anthropology, Leipzig, [22]Max Plack Insitute for Evolutionary Anthropology, Peipzig, Germany, [23]School of Natural Sciences, Stirling, United Kingdom, [24]Agence National des Parcs Nationaux, Libreville, , [25]EcoHealth Alliance, New York, United States, [26]Max Planck Institute for Evolutionary Anthropology, Leipzig, Germany, [27]School of Human Evolution and Social Change, Tempe, United States, [28]Wildlife Chimpansee Foundation, Leipzig, Germany, [29]Chester Zoo, Chester, [30]STICS - Stirling Conservation Science, Stirling, [31]School of Natural Sciences and Psychology, Liverpool John Moores University, Liverpool, United Kingdom, [32]Department of Anthropology, Washington University in Saint Louis, St. Louise, [33]Lester E. Fisher Center for the Study and Conservation of Apes, Chicago, United States, [34]Wild Chimpanzee Foundation, [35]Max Planck Insitute for Evolutionary Anthropology, Leipzig, Germany, [36]School of Natural Sciences and Psychology, Liverpool John Moores University, Liverpool, United Kingdom, [37]West Chester University, Wset Chester, United States, [38]German Centre for Integrative Biodiversity Research (iDiv), Leipzig, Germany, [39]Catalan Institution of Research and Advanced Studies (ICREA), [40]CNAG-CRG, Centre for Genomic Regulation (CRG), Barcelona Institute of Science and Technology, Barcelona, [41]Institut Catala de Paleontologia Miquel Crusafont, Universitat Autonoma de Barcelona, Cerdanyola del Vallès, Spain, [42]University College London, London, United Kingdom, [43]Royal Zoological Society of Antwerp, Antwerp, Belgium, [44]Max Planck Instiute for Evolutionary Anthropology, Leipzig, Germany, [45]Universite de Neuchatel, Neuchatel, Switzerland, [46]Max Planck Institute for human Behaviour, Ecology & Culture, Leipzig, Germany

Abstract: The *Pan African Programme: The Cultured Chimpanzee* (PanAf) is a large-scale research project that aims to understand and collect systematic ecological, social, demographic and behavioral data across the entire geographic

range of extant chimpanzees (*Pan troglodytes*). Included in the PanAf project is the collection of geo-localized non-invasive (NI) samples to study the genetic diversity of chimpanzees. NI sampling is preferred to avoid disturbance and negative effects on the studied species. However, they come with additional difficulties, specifically lower sample quality, low proportions of endogenous content and higher rates of DNA degradation.

We report our efforts to leverage methodological advances in next-generation sequencing to acquire genomic data (chr21 and exome) from more than 800 NI samples from 51 chimpanzee sites using target capture methods. We analyze how sample variability and geographical site influences DNA quality, estimate demographic parameters and characterize the migration patterns among chimpanzee communities. As a result of our extensive sampling effort we discover novel genetic diversity that can be traced to a specific geographical location. For example, despite the low levels of genetic diversity in western chimpanzee communities (*Pan troglodytes verus*), we observe a tight correlation between genetic markers and geography. With this new fine-scale geographic and genetic variation map of chimpanzees we are building a vital resource to aid the management and protection of this endangered species by, for instance, genotyping illegally trafficked individuals to determine their country of origin. The PanAf is an extremely valuable resource that is broadening our knowledge of chimpanzees from numerous perspectives.

Contemporary Evolution

SMBE-PO-050

Within-host diversity of the global influenza population

L. Zhao [1,*], R. Daniels [2], J. McCauley [2], C. Illingworth [1]

[1]Department of Genetics, University of Cambridge, Cambridge, [2]Crick Institute, London, United Kingdom

Abstract: Many viruses form large within-host populations and evolve under the influence of high mutation rates. As a consequence, within-host viral populations may contain a large amount of sequence diversity. Viral population diversity has a close relationship with the evolution of viral populations and is therefore a useful metric with which to understand and evaluate the biology of infectious disease. We here use viral diversity to explore the relationship between viral evolution and the local climate. Firstly, we show that some metrics that have been used to evaluate population diversity in viral populations contain systematic biases; a careful choice of metric is required to achieve meaningful comparison between the diversities of different populations. Secondly, we use within-host sequence data generated from clinical cases of influenza A infection from multiple countries worldwide to explore systematic links between climate and within-host viral diversity. We identify a strong statistical link between within-host influenza A/H3N2 diversity and the mean absolute humidity of the area in which the sample was collected. We explore how links between climate and viral transmission may shape the evolution of influenza populations

Contemporary Evolution

SMBE-PO-051

miRNAs preferentially regulate whole genome duplicated genes

L. R. Tincopa Marca [1,*], S. Griffiths-Jones [1], M. Ronshaugen [1]

[1]The University of Manchester, Manchester, United Kingdom

Abstract: MiRNAs are small non-coding RNAs approximately 20-22 nucleotides long which function is to post-transcriptionally regulate the expression of most genes in animals and plants. Gene duplication is an important evolutionary process that can drive diversity and novelty. However, gene duplication can generate a gene dosage imbalance in the cell. Here, we investigate the possible roles of miRNAs in buffering gene dosage post duplication. We propose that small-scale duplications (SSD) can generate a stoichiometric imbalance in gene products, however miRNAs may play a role in buffering expression of such genes limiting the imbalance. Alternatively, genes duplicated during whole genome duplication (WGD) processes may not create such an imbalance, and therefore will show evidence of being less targeted by miRNAs. To address this scenario, we determine the properties of targeted WGD and SSD genes. We predicted miRNA target sites in 3'UTRs of SSD, WGD and single copy genes in human, mouse, rat, pig, dog and chicken genomes. Contrary to our hypothesis, we found that in human, mouse, rat, miRNAs preferentially regulate WGD genes. In addition, we found that miRNAs preferentially regulated haploinsuficient genes.

Contemporary Evolution

SMBE-PO-054

PATTERNS OF INTER-SUBTYPE RECOMBINATION IN FULL HIV GENOMES FROM UGANDA

H. Grant [1,*], E. Hodcroft [2], D. Ssemwanga [3], N. Bbosa [3], R. Nsubuga [3], U. Bahemuka [3], Y. Mayanja [3], M. Kuteesa [3], A. Kamali [3], F. Kibengo [3], R. Newton [3], J. Seeley [3], D. L. Robertson [4], P. Kaleebu [3], A. Leigh Brown [1]

[1]Institute of Evolutionary Biology, University of Edinburgh, Edinburgh, United Kingdom, [2]University of Basel, Basel, Switzerland, [3]Uganda Virus Research Institute, MRC/UVRI & LSHTM Uganda Research Unit, Entebbe, Uganda, [4]Centre for Virus Research, University of Glasgow, Glasgow, United Kingdom

Abstract: Human immunodeficiency virus (HIV) is a highly diverse retrovirus at both the within-individual and population level. This diversity is generated by high mutation rates and pervasive recombination. Previous work has shown that the Ugandan HIV epidemic is predominantly composed of subtype A and subtype D, and that these subtypes have co-circulated in the country for nearly 50 years. As part of the PANGEA HIV consortium project, 531 near full length genome sequences have been obtained from samples collected by the MRC/UVRI & LSHTM unit in Uganda. This dataset provides a unique opportunity to test for the favorable distribution of recombinogenic breakpoints along the HIV genome in a natural population. Subtyping was performed with the program SCUEAL. We find 288 (54.2%) 'pure' subtype genomes (153 A; 127 D; 8 C), and 243 (45.8%) inter-subtype recombinants, where almost all recombinants have a unique recombination pattern. We find evidence for reduced recombination within envelope (gp120) and the favorable inheritance of gp120 as an intact gene region.

Contemporary Evolution

SMBE-PO-053

Exploring the hypothetical domain of the eukaryote common ancestors through ancestral state reconstuction

D. Newman*, J. McInerney[1]

[1]University of Nottingham, Nottingham, United Kingdom

Abstract: The emergence of eukaryotes changed the face of life on the planet; from multicellularity to the eukaryotic cell itself being a significant increase in size and complexity compared to prokaryotes, as well as the vast diversity of forms amongst eukaryotes themselves. Recently, a profusion of studies have made great advances in suggesting what biology and genetic material can be inferred as being present in the Last Eukaryote Common Ancestor (LECA) and the idea that LECA already possessed most features that characterise modern eukaryotes has become well established. In an effort to go beyond our phylogenetic inferences we have employed MrBayes and gene families we identified as being in LECA that possess prokaryote homologs as the basis for Bayesian Ancestral Sequence Reconstructions to construct hypothetical LECA amino acid sequences. These hypothetical LECA and their corresponding modern eukaryote sequences were then analysed for their domain complement and architecture using InterProScan. With the domain complement and architecture of these hypothetical LECA proteins compared against their modern forebears, we will examine how comparable the molecular functions of conserved gene families are across the kingdoms of eukaryotic life and against their hypothetical ancestral form.

Contemporary Evolution

SMBE-PO-058

PGG.SNV: Understanding the evolutionary and medical implications of human single nucleotide variations in diverse populations

S. Xu*

Abstract: Our current biological understanding of the human genome is still very much in its infancy. The functions and phenotypic effects of the majority of our genome remain unknown. We developed a database, *PGG*.SNV, which gives much higher weight to previously under-investigated indigenous populations in Asia, whose genomes harbor an enormous number of variants that have not been observed in the extensively studied populations of European ancestry. *PGG*.SNV archives genetic variations, in particular, single nucleotide variants (SNVs), across 302,567 genomes and estimates their frequency in 980 diverse populations, including 1,009 newly sequenced genomes representing 16 indigenous populations living in unusual environments (e.g., tropical forests and highlands) in East Asia and Southeast Asia that were not covered by any previous projects or studies. Moreover, *PGG*.SNV archives 1,018 ancient genomes. *PGG*.SNV documents more than 10 billion individual SNVs. *PGG*.SNV provides various approaches to query SNV information and nine types of annotations for each variant. Notably, estimation of population genetic diversity and evolutionary parameters is available in *PGG*.SNV, which is a unique feature compared with other databases. Our efforts are expected to: i) build reference genomic resources for diverse human populations including indigenous groups in East Asia and Southeast Asia; ii) enable studies of variants that are rare or not existing in well-studied populations; iii) improve the interpretations of phenotype-genotype associations and adaptations to local environments for both Asian and non-Asian populations; and iv) help advance our understanding of the biological meaning of the human genome sequence.

Contemporary Evolution

SMBE-PO-064

A genomic approach to contemporary local adaptation in a marine mammal, the harbour porpoise (Phocoena phocoena) in the Baltic Sea

M. Autenrieth [1,*], K. Havenstein [1], L. Lah [1 2], J. Saeurich [3], J. Ucar [4], A. Savas [5], S. Hartmann [6], H. Benke [7], I. Pawliczka [8], A. Roos [9], U. Siebert [10], T. Gunnlaugsson [11], G. Vikingsson [11], A. B. Dennis [1], R. Tiedemann [1]

[1]Evolutionary Biology & Systematic Zoology, Potsdam University, Potsdam, Germany, [2]Novartis BTDM Mengeš, Mengeš, Slovenia, [3]Hannover University, Hannover, [4]Freie Universität Berlin, [5]Humboldt University, Berlin, [6]Evolution and Adaptive Genomics, Potsdam University, Potsdam, [7]Deutsches Meeresmuseum, Stralsund, Germany, [8]Hel Marine Station, University of Gdansk, Hel, Poland, [9]Swedish Museum of Natural History, Stockholm, Sweden, [10]Institute for Terrestrial and Aquatic Wildlife Research, University of Veterinary Medicine Hannover, Büsum, Germany, [11]Marine and Freshwater Research Institute, Reykjavík, Iceland

Abstract: The Baltic sea, beeing aprox. 7000 years old, and with high salinity differences to North Sea provides potentially strong selective pressure for contemporary evolution. The harbour porpoise (Phocoena phocoena), a highly mobile cetacean found in the Northern hemisphere, inhabits basins that vary broadly in salinity, temperature, and food availability, which drives population differentiation. ddRAD analyses screening several thousand SNPs show little differentiation in the Eastern North Atlantic, e.g. genetically similarities of North Sea to Iceland, despite their geographical distance. In contrast, porpoises within the Baltic Sea have significantly diverged from the adjacent North Sea population, with a further, inner Baltic split. The habitats of these genetically identified populations differ substantially in salinity, ranging from oceanic to brackish. We produced a full genome of a Belt Sea porpoise with a draft annotation yielding 22,154 predicted genes. Of these we analysed candidate genes adaptive to the aquatic lifestyle, e.g. ELOVL genes associated with blubber building and the TMC1 gene associated with cetaceans evolution of hearing. We further identified genes related to osmoregulation, potentially associated with local adaptation. Combining these data, we aim at identifying non-synonymous SNPs potentially related to local adaption. Given the young age of the Baltic Sea, we hypothesize standing genetic variation as a potential source for rapid local adaptation. With the example of an illustrative marine mammal species, we aim to also contribute to the discussion on the mode of contemporary evolution, i.e., hard selective sweeps at a few candidate loci vs. soft sweeps throughout the genome.

Contemporary Evolution

SMBE-PO-063

Evolution of the extracytoplasmic function sigma factor protein family

D. Pinto [1,*], R. R. da Fonseca [2]

[1]Institut für Mikrobiologie, Technische Universität Dresden, Dresden, Germany, [2]Center for Macroecology, Evolution and Climate, Natural History Museum of Denmark, University of Copenhagen, Copenhagen, Denmark

Abstract: Understanding the process of transcription has been a central goal of the scientific community for decades. However, much is still unknown, especially concerning how this process is regulated. In bacteria, a single DNA-directed RNA-polymerase performs the whole of transcription. This enzyme contains multiple subunits, among which the σ factor that confers promoter specificity. Besides the housekeeping σ factor, which is responsible for transcription of all genes required for growth, bacteria encode several alternative σ factors that initiate transcription of specific subsets of genes crucial for processes that range from motility to stress responses.

The most abundant and diverse family of alternative σ factors, the extracytoplasmic function (ECF) family, regulates transcription of genes associated with stressful scenarios, making them key elements of adaptation to specific environmental changes. Despite this, the evolutionary history of ECF σ factor family has never been investigated.

Here, we report on our analysis of thousands of members of this family. We show that single events are in the origin of alternative modes of regulation of ECF σ factor activity that require partner proteins, but that multiple events resulted in the acquisition of regulatory extensions. We show that in the Bacteroidetes, there is a duplication of an ecologically relevant gene cluster, whereas in Planctomycetes, the duplication generates distinct C-terminal extensions after fortuitous insertion of the duplicated σ factor. Finally, we also demonstrate horizontal transfer of ECF σ factors between soil bacteria.

Contemporary Evolution

SMBE-PO-066

Genomic responses to extreme selective pressure over the last 100 years: insecticide resistance in a pest moth

A. Mcgaughran*

Abstract: The cotton bollworm (*Helicoverpa armigera*) is a major pest of cotton and other agricultural crops, costing billions of dollars in management and yield losses globally. A key aspect of its success as a 'megapest' is the moths' ability to rapidly evolve insecticide resistance.

My research uses DNA from Australian moth samples that were collected over the last 100 years. Included in my sampling design are populations that were never exposed to insecticides and populations that were exposed to different classes of insecticide following the rapid occurrence of field resistance. This allows me to compare genomic sequences from historical, insecticide-free, pest samples to samples that were collected when resistance to various insecticides was being recorded in the field, as well as to contemporary samples that may no longer carry the historical signal of selective pressure from these chemicals.

Examining the genetic make-up of a rapid adaptive response to an environmental change (i.e., insecticide resistance), provides a template for testing questions about the presence or absence of pre-adapted resistance genes and for understanding the mechanisms behind successful pest status. In my talk, I will focus on these questions, presenting preliminary results that look at genomic shifts (e.g., Fst, allele frequency) between pre-, mid-, and post-insecticide sampling points.

Contemporary Evolution

SMBE-PO-061

Evolutionary processes involved in the emergence of Treponema pallidum and the current global re-emergence of syphilis

M. Pla-Diaz [1], V. J. Schuenemann [2], D. Smajs [3], L. Sanchez-Buso [4], P. Bosshard [5], H. C. Bagheri [6], K. Nieselt [7], L. Giacani [8], N. Arora [9], F. Gonzalez-Candelas [10,*]

[1]Unidad Mixta Infeccion y Salud, FISABIO-University of Valencia, Valencia, Spain, [2]Institute of Evolutionary Medicine, University of Zurich, Zurich, Switzerland, [3]Department of Biology, Masaryk University, Brno, Czech Republic, [4]Welcome Trust Sanger Institute, Hinxton, United Kingdom, [5]Department of Dermatology, University Hospital of Zurich, Zurich, Switzerland, [6]Repsol Technology Center, Madrid, Spain, [7]Center for Bioinformatics, University of Tübingen, Tübingen, Germany, [8]Univesity of Washington, Seattle, United States, [9]Institute for Evolutionary Biology and Environmental Studies and Zurich Institute of Forensic Medicine, University of Zurich, Zurich, Switzerland, [10]Unidad Mixta Infeccion y Salud, University of Valencia-FISABIO, Valencia, Spain

Abstract: Introduction: *Treponema pallidum* subsp. *pallidum* (TPA) is the etiological agent of syphilis, a widespread potentially devastating sexually transmitted disease that has been resurgent in the last decades in developed countries and is still endemic in underprivileged settings. TPA genome is highly conserved and very similar to that of the closely related *T. pallidum* subsp. *pertenue* (TPE) and *T. pallidum* subsp. *endemicum* (TEN), which cause the human treponematoses yaws and bejel, respectively. The time and location of origin of TPA are still under debate. Here we have enquired about the evolutionary processes driving TPA evolution since the common ancestor of the agents of human treponematoses and its recent epidemiological spread. To this end, we have analysed 75 complete genome sequences from the three different *T. pallidum* subspecies looking for potential recombination events and to establish the role of natural selection in their evolution.

Material and methods: We generated maximum likelihood phylogenetic trees for the concatenated alignment and each protein-coding gene in the dataset with IQTREE, and used a novel procedure, named Phylogenetic Incongruence Method (PIM), for the analysis of recombination in these genomes. PIM evaluates phylogenetic signals and performs congruence tests with IQTREE to determine potentially recombinant genes. The PIM results were compared with those obtained with Gubbins and ClonalFrame. The analyses of selection were performed with SNPeff, PAML (Codeml) and Hyphy.

Results: We found 12 genes with 21 different recombination events, which constitutes very strong evidence for recombination among *T. pallidum* subspecies. Only one recombination event per gene was detected for each of the identified genes, with the exception of *tp0136* and *tp0326* genes, which showed 7 and 4 events, respectively. All but one events (in *tp0136*) corresponded to inter-subspecies transfers (TPE/TEN to TPA).

Clear evidence for natural selection acting on the recombinant genes was provided by significantly higher values of the non-synonymous/ synonymous substitution rate ($w = dN/dS > 1$) in the recombinant regions than in the non-recombinant zones of those genes. This signal was confirmed by Codeml and SNPeff analysis. Additionally, 14 non-recombinant genes with evidence of positive selection, and 23 non-recombinant genes with significantly more SNPs than expected were identified. These signals of natural selection were also confirmed by Hyphy (Busted test) in ten of the recombinant genes, two of the non-recombinant genes with abundant SNPs and in one of the other non-recombinant genes with positive selection signal.

Discussion: Our results indicate that recombination and selection were forces strongly implicated in the emergence of TPA.as a subspecies. The phylogenetic location of some of these events and the functional role of the genes involved along with the detection of positive selection acting on them suggest that both recombination and selection may have had an important role in the evolution and current re-emergence of the syphilis pathogen.

Evolution Ecology and Host-Virus dynamics in the Microbiome
SMBE-PO-081

Wolbachia incidence and host shift in scale insects

E. Sanaei [1,*], Y.-P. Lin [2], L. G. Cook [1], J. Engelstädter [1]
[1]School of Biological Science, University of Queensland, Brisbane, Australia, [2]Department of Plant Medicine, College of Agriculture, National Chiayi University, National Chiayi University, Taiwan

Abstract: *Wolbachia* is a group of endosymbiotic alpha-proteobacteria that mostly live as reproductive parasites in arthropods. It is a master manipulator, modifying its hosts' biology in many ways to increase its vertical (maternal) transmission. *Wolbachia* can also undergo host shifts that can be mediated by ecological vectors such as shared host plants or parasitoids. Many surveys of *Wolbachia* have been conducted in different arthropods that have uncovered a huge diversity of strains across host species. However, the standard methodology based on PCR and Sanger sequencing is unsuitable to determine strain diversity within individual hosts. Here, we present the first study investigating the prevalence, genetic diversity and phylogeny of *Wolbachia* in scale insects (Coccomorpha). We screened 780 individual scale insects from 120 species and nine families. Through a new approach of Illumina amplicon sequencing of seven *Wolbachia* and three host genes we were able to obtain a fine-scaled picture of *Wolbachia* diversity, both across host species and within infected individuals. High strain diversity within individual hosts were observed. Although we could not detect *Wolbachia* in some families, others had high infection rates up to 29% (mealybugs, Pseudococcidae). After strain identification, we investigated potential routes of host-shift among scale insects by also screening intimately associated species such as ants and parasitoids. Among all screened associate species, it appears that close ecological interactions between scale insects and ants may be a major route of *Wolbachia* transfer.

Evolution Ecology and Host-Virus dynamics in the Microbiome

SMBE-PO-082

Dynamics, evolution and stability of the respiratory microbiome

S. Megremis [1], A. Tapinos [1,*], J. Busby [2], L. S. Kong [3], C. F. Yap [4], T. Gilman [5], M. Muldoon [4], D. Robertson [2], N. Papadopoulos [6]

[1]Division of Evolution & Genomic Sciences, The University of Manchester, Manchester, [2]MRC-University of Glasgow Centre for Virus Research, Glasgow, [3]Faculty of Biology, Medicine and Health, [4]School of Mathematics, [5] School of Earth & Environmental Sciences, [6]Division of Infection, Immunity & Respiratory Medicine, The University of Manchester, Manchester, United Kingdom

Abstract: The respiratory tract is a region of microbe influx in the human body and is under constant pressure from the environment. A main role of the respiratory microbiome is to prevent colonisation or overgrowth of pathogens. The complex interactions occurring in respiratory microbial populations and the dynamic properties of this system are relatively under-studied. It is thus important to better understand healthy and disease specific respiratory microbiome steady states from ecological and evolutionary perspectives. In this study we present analysis of longitudinal nasopharyngeal samples collected from asthmatic donors in Poland, Greece, and Finland. The cohort consists of pre-school children who had no asthma exacerbation or respiratory infection over the two-year sampling period. The inter-patient microbiome comparison, identified a 'core' respiratory microbiome shared by all donors. We also focus our investigation on identifying taxonomic similarities and differences that appear in donors from different age groups. In addition, we investigated longitudinal changes in the composition of each individual donor's microbiome, providing information on concurrent changes, recurring fluctuations, and/or evolutionary characteristics within species. The functional and taxonomic stability of the respiratory microbiome are described using co-abundance networks and topological analysis of their structure. Evolutionary properties such as taxonomic similarity and species diversity of individual donors' microbiomes will establish the stability and resilience of both the bacteriome and the virome, over time.

Evolution Ecology and Host-Virus dynamics in the Microbiome

SMBE-PO-073

Diversity of gut microbiome in healthy individuals revealed by RNA sequencing

V. N. Pimenoff [1,*], F. Moratalla-Navarro [1], J. Mas-Lloret [1], M. Obón-Santacana [1], R. Carreras-Torres [1], G. Ibanez-Sanz [1], E. Guinó [1], V. Moreno [1]

[1]Oncology Data Analytics Program, Bellvitge Biomedical Research Institute, L'Hospitalet de Llobregat, Spain

Abstract: Background: Gut microbiome has a dynamic role in human health and disease. Current metagenomic and phylogenetic methods have enable an extensive exploration of the culture-independent gut microbial communities even at strain resolution. Indeed, pathogenic potential is strain-specific for most gut bacteria and, in consequence, phylogenetic variants may differentially disrupt the gene expression patterns at the host colonic epithelia.

Methods: Here, we present metagenomic and phylogenetic analysis of the gut microbiome from mucosal biopsies of 308 healthy elderly individuals enrolled from a population-based colorectal cancer screening program. Colon mucosal tissue samples were sequenced at high coverage using RNA-sequencing, and the microbial fraction of the data was classified using an exact alignment of K-mers algorithm against a large repository of bacterial reference genomes. Further strain-level analyses were performed using phylogenetic inference on specific species taxa and differential host gene expression patterns were explored for signs of host-microbiome interactions.

Results: Our microbiome taxonomic profiling from the metatranscriptome data identified between 30,000 and 1,200,000 reads per sample associated to a particular gut bacterial species. Variation of the microbiome diversity and distribution was observed among healthy individuals at phylum to species-level. Moreover, a differential strain-level bacterial distribution for particular taxa were observed.

Conclusions: Strain-level phylogenetic diversity and distribution of the gut microbial species is crucial in determining their functional capabilities, including interaction with the gut epithelia and immune homeostasis. Moreover, gut microbiome diversity distribution at strain-level may contribute to inconsistencies in genus and species-level associations. Taken together, strain-level bacterial diversity should be contemplated when examining host-microbiome interactions.

Evolution Ecology and Host-Virus dynamics in the Microbiome

SMBE-PO-076

Identification of pathogenic coevolution marks between Fusarium oxysporum f. sp. vanillae and Vanilla planifolia by fungi comparative genomics.

I. A. Carbajal [1,*], A. cibrián [1]

[1]LANGEBIO MX, CINVESTAV, irapuato, Mexico

Abstract: *Vanilla planifolia* is one of the most important orchids, its extract is sold as the second most valuable vegetal product worldwide. The main biological challenge with its production is the root and stem rot disease, caused by *Fusarium oxysporum*, a fungus that develops close host-pathogen interactions, holding high genetic variability. Transposable elements play an important role in this ongoing arms race, specially a miniature inverted-repeat transposable elements family (mites), related with genes capable of confer pathogenicity by horizontal gene transfer in different strains. Our goal in this work is to identify specific *Fusarium oxysporum f. sp. vanilleae* pathogenic sequences (homologous and novel) and to infer about their phylogenetical origin.

We isolated 76 morphologically different axenic fungal endophytes from healthy and infected vanillas, from this, we identified 9 different genera genetically including *Fusarium*. Five *Fusarium* strains were selected for whole genome sequencing.

A good-quality assembly of *Fusarium oxysporum f. sp. vanillae* (Fov) has been generated for the first time along with genomes of additional pathogenic and non-pathogenic *Fusarium* endophytes of vanilla. As evidence suggested, *Fov* holds a two-evolutionary rates genome. Pathogenicity genes were identified by genomes alignment for mites sequences search and we suggest an effector profile for *Fov,* useful for early diagnosis and pathogen identification tools development on the field.

Evolution Ecology and Host-Virus dynamics in the Microbiome
SMBE-PO-072
Host phylogeny constrains bacteriophage host range
B. Stamp [1,*], D. L. Robertson [1]
[1]MRC-University of Glasgow Centre for Virus Research, Glasgow, United Kingdom

Abstract: The role of the human microbiome in health and disease has become increasingly recognised and studied. Microbiome communities are defined by a set of complex interactions with ecological, spatiotemporal and evolutionary dimensions, between species from all domains of life. Bacteriophage-host interactions form a key part of this system and also offer a unique potential to be able to influence and manipulate this system via bacteriophage therapy. Understanding these interactions relies on a characterisation of the host range of these viruses and, given many recent co-evolutionary studies have demonstrated bacteriophage host switching is common, likely changes in host range must also be considered. It's intuitive to expect host relatedness to be one of the most significant determinants of host range and host switching. This pattern of preferential host switching has been well characterised for eukaryotic viruses, assisting novel approaches to computationally predict viral host range and likely zoonoses. However, this phylogenetic/evolutionary relatedness dimension is often neglected when virus-host interactions are studied in prokaryotes. Broad and narrow host range are also still poorly defined, further hindering an adequate characterisation of viral host range. Here, we present a taxonomic approach to characterise breadth of host range, analysing a dataset of over 14,000 interactions between around 10,000 viruses and 3,000 hosts. While it's common for both prokaryotic and eukaryotic viruses to infect multiple host species, it becomes increasingly uncommon for viruses to infect multiple host groups as taxonomic resolution is reduced. By phylum level, viruses infecting multiple host taxonomic groups is extremely uncommon. Arboviruses standout as an interesting exception to this for eukaryotes, able to infect both Chordates and Arthropods. We demonstrate a similar pattern of host phylogeny restricting host range for both eukaryotes and prokaryotes which defines the taxonomic limits to host range. This host range restriction will have consequences for the probability of successful host switching and therefore offers some power to predict the likelihood of host switching events as well as characterise bacteriophage host range in the microbiome.

Evolution Ecology and Host-Virus dynamics in the Microbiome

SMBE-PO-071

Microbial diversity within lower and higher wood-feeding termites

A. Chakraborty [1,*], L. Michalčíková [1], K. Votýpková [1], J. Sobtnik [1]

[1]Faculty of Forestry and Wood Sciences, Czech University of Agricultural Sciences, Suchdol, Prague, Czech Republic

Abstract: The co-existence of insects and their microbial allies play an important role in shaping their ecology and evolution. Termites are eusocial insects that thrive in great abundance in terrestrial ecosystems. The symbiotic gut microbiota produces digestive enzymes that enable termites to feed on nutritionally poor substrates such as wood and soil. The presence of diverse microbial consortium consisting ecto- and endo-symbionts within termites' functions in the decomposition of complex organic matter and recalcitrant plant material, carbon mineralization, N waste recycling or N-fixation. We investigated the microbial diversity both bacterial (via 16S) and fungal (via ITS) through MiSeq in two different species of wood-feeding termites namely *Reticulitermes* (lower termite) and *Microcerotermes* (higher termite) feeding on a variety of wood with different lignin content. Our data pinpoint towards the evolution of diverse symbiotic associations within lower and higher termites based on lignin content in their diet.

Evolution Ecology and Host-Virus dynamics in the Microbiome
SMBE-PO-083A

Functional role of endophytic fungi within coralloid roots of cycads.

F. López Restrepo [1,*]
[1]Langebio, Cinvestav, Irapuato, Mexico

Abstract: Cycads are an endangered gymnosperm order that has existed since the Permian period. These plants are considered a model for symbiosis in eukaryotes, as they constitute the only order that fixates nitrogen through a symbiotic relationship with cyanobacteria within a specialized organ, the coralloid root. These structures also contain a great diversity of organisms, among which are many species of fungi. While similar systems have been extensively studied, showing that fungi help increase the rates of nitrogen fixation of cyanobacteria, the role fungi fill within the coralloid root, as well as the way they organise is currently unknown. We have observed the presence of endophytic fungi within the cells of the coralloid root through the use of microscopy, and in order to determine the effect fungi are having within the coralloid root, we have collected samples from cycads living both in research greenhouses and in its natural habitat to investigate the activity of fungi present in the coralloid root through metatranscriptomics. We hope to determine which genes are being expressed differentially in the coralloid root when compared to the closest structure, the secondary root. We expect to find that the endophytic fungi of the coralloid root present a similar role to those in other systems, likely as a result of convergence.

Evolution Ecology and Host-Virus dynamics in the Microbiome

SMBE-PO-070

Preliminary study on bacterial symbionts in free-living protists using single-cell 'omics

Y. Yan [1], M. A. Cerón-Romero [1 2,*], A. K. Weiner [1], L. A. Katz [1 2]

[1]Biological Sciences, Smith College, Northampton, [2]Program in Organismic and Evolutionary Biology, University of Massachusetts Amherst, Amherst, United States

Abstract: Symbiotic interactions, especially between bacteria and eukaryotes, exist in nearly all ecosystem. While extensive work has been done on the diversity of bacterial symbionts within animal and plant hosts, limited cases have been studied of symbiotic relationship between bacteria and microbial eukaryotes (i.e. protists). Endosymbiotic bacteria have been reported to be able to increase the fitness of both the host organisms and themselves by altering the gene expression of their hosts or manipulating the host cell cycles. However, such studies have focused on a few cultivable lineages. Given that protists represent the majority of diversity in eukaryotes, understanding their associations with bacteria will expand our knowledge of adaptation and genome evolution. Collecting single-cell transcriptomes from uncultivable free-living protists *Loxodes* sp. (Ciliophora), *Hyalosphenia papilio* and *Hyalosphenia elegans* (Amoebozoa), we identified bacterial sequences that represent candidate symbionts. We are now using fluorescent *in situ* hybridization (FISH) on samples with species-specific probes to confirm the presence of symbionts and to identify their location within the protist host. These data provide a first step in understanding the microbiomes of free-living protists.

Evolution Ecology and Host-Virus dynamics in the Microbiome

SMBE-PO-075

Molecular epidemiology and evolution of viral respiratory infections in Vietnam

L. Lu [1,*], G. Robertson [1], J. Ashworth [1], M. Woolhouse [1]

[1]Centre for Immunity, Infection and Evolution, University of Edinburgh, Edinburgh, United Kingdom

Abstract: Acute respiratory infections (ARIs) represent a major public health problem in Vietnam. The genetic diversity and transmission patterns of respiratory viral pathogens that are co-circulating in Vietnam in the same time period are not well characterized. We enrolled 4326 patients with ARI from 5 hospital sites in Vietnam within 4 years. Real-time PCR (RT-PCR) testing was performed for 14 respiratory viruses in nasopharyngeal or throat (NT) swabs samples. We found that 63.8% of the samples were positive with viral pathogens. The most frequently detected viruses were *Respiratory syncytial virus* (RSV) (22.7%), *Human Rhinovirus* (HRV) (12.7%) and *Influenza A virus* (10.6%). Statistical analyses showed that patients with different viral pathogen infections differed significantly in the age, locations and time of hospital admission as well as disease severity. Disease severity also differed significantly between patients with viral infections and those virus-negative patients; while the differences were not significant between patients with single virus infection and those with multiple infections.

Further, respiratory viruses sequences belonging to 13 families and 24 genera were retrieved in metagenomic sequencing, including 113 complete/nearly complete genomes. 42.6% of the whole genomes (n=46) are the first time reported in Vietnam, including *Human Parainfluenza virus* (HPIV), *Human adenovirus* (HAdV), *Human metapneumovirus* (HMPV), Human coronavirus (CoV), HRV, *Enterovirus B virus*, *Influenza C virus, Measles morbillivirus* (MeV*)*, *Cardiovirus* and *Human polyomavirus* (HPyV), the latter four are viruses associated with respiratory infections but are not commonly tested. The molecular evolution of major respiratory viral pathogens was further compared by using phylogenetic methods. We found multiple subtypes/clades of viruses were co-circulating within the same locations in Vietnam, where mainly receiving major respiratory viruses transmitted from neighboring countries.

Our results suggest that linking epidemiological data and molecular diagnostic technologies can provide a better understanding of the evolutionary and epidemiological dynamics of respiratory viral infection. We also highlight the importance of extensive samplings to further examine the viral pathogens that are associated with respiratory diseases in humans in Vietnam.

Evolution Ecology and Host-Virus dynamics in the Microbiome

SMBE-PO-078

Diverse phage communities enhance plant health by specifically targeting the pathogen and boosting the rhizosphere microbiome

X. Wang [1], Z. Wei [1], A. Jousset [2], V.-P. Friman[*]

[1]Nanjing Agricultural University, Nanjing, China, [2]Utrecht University, Utrecht, Netherlands

Abstract: Pathogen-specific parasitic viruses (phages) has been proposed as an alternative strategy for controlling agricultural pathogens. However, phage-biocontrol outcomes are still highly variable and poorly understood in natural rhizosphere microbiomes at timescales considering both ecological and evolutionary processes. Here we studied how phage community diversity affects the biocontrol efficacy of devastating *Ralstonia solanacearum* plant pathogenic bacterium in tomato plant rhizosphere. We find that increasing phage community richness decreases disease incidence up to 80% both in greenhouse and field experiments. While this could be explained by pathogen density reduction and selection for highly resistant but slow-growing pathogens, phages also indirectly changed the composition and the diversity of rhizosphere microbiome and enriched bacteria that were highly antagonistic towards the pathogen. Together these results demonstrate that diverse phage communities can be used as 'precision' tools to enhance plant health by specifically targeting the pathogen and by indirectly boosting the functioning of surrounding rhizosphere microbiome.

Evolution Ecology and Host-Virus dynamics in the Microbiome

SMBE-PO-079

Coexistence of related bacteria within the honey bee gut microbiota influenced by host diet

S. Brochet [1,*], K. M. Ellegaard [1], R. Mars [2], U. Sauer [3], P. Engel [1]

[1]Department of Fundamental Microbiology, University of Lausanne, Lausanne, Switzerland, [2]Kashyap laboratory, Mayo Clinic, Rochester, United States, [3]Institute of Molecular Systems Biology, ETH Zürich, Zürich, Switzerland

Abstract: In this study, we use the honey bee gut microbiota to investigate how closely related bacteria coexist in natural microbial communities. We focus on the abundant phylotype *Lactobacillus* Firm5. This phylotype has diversified into four different species, each of which harbors extensive strain-level diversity. Interestingly, all four species consistently coexist in individual bees. We hypothesize that this coexistence is facilitated by nutritional niche partitioning. To test this hypothesis, we first investigated if strains of the four Firm5 species coexist *in vitro*, depending on the availability of nutrients derived from the honey bee diet. We then explored the metabolic profiles of 14 strains in the presence of these nutrients using untargeted metabolomics. Finally, we used RNA sequencing to investigate the genetic basis of species-specific adaptations to the bee diet. Our preliminary results show that the *in vitro*coexistence of closely related species tends to be more stable in the presence of honey bee diet-derived nutrients compared to simple sugars. Moreover, our metabolomics results reveal substantial differences in metabolic profiles between the four species, but also between certain strains of the same species. Therefore, we are currently carrying out a systematic screen for comparing patterns of co-existence for strains from the same or a different species. Finally, our RNA sequencing results show large differences in the expression of genes involved in carbohydrate transport and metabolism in response to diet. Together these results show that coexistence of related bacteria in the honey bee gut is likely to be influenced by diet.

Evolution Ecology and Host-Virus dynamics in the Microbiome
SMBE-PO-080
Adding Biology to a Widespread, Metagenomically-defined Bacteriophage

P. Kirchberger [1,*], H. Ochman [1]
[1]Department of Integrative Biology, University of Texas at Austin, Austin, United States

Abstract: Gokushoviruses are a group of single stranded circular DNA bacteriophages found in every ecosystem on the planet. Their small genomes have been recovered and assembled from metagenomes ranging from the open ocean to salt flats, hydrothermal vents and the human gut, where they are often the dominant and most quickly evolving viral group. Despite their presence in a variety of environments, little is known about their biology or bacterial hosts: Gokushovirus isolates are exceedingly rare, having only been cultivated from a few obligate intracellular bacteria. We have recovered numerous Gokushoviruses that reside as prophages, integrated into the genomes of diverse genera within the *Enterobacteria*. The enteric Gokushoviruses form a monophyletic group within a larger clade consisting of phage identified from metagenomic samples, and they likely represent more typical Gokushoviruses than those previously isolated. Through synthesis and transformation of circularized phage genomes into *E. coli*, we produced viable Gokushovirus phage particles, allowing us to resolve the properties of this group. Gokushoviruses were previously regarded to be exclusively lytic due to their lack of integrases, but we show that revived phages readily integrate into host genomes. This activity is accomplished by hijacking a phylogenetically conserved chromosome dimer resolution system, similar to that employed by cholera phage CTX as well as some non-phage mobile elements. We show that much of what has been previously assumed about the biology of Gokushoviruses is incorrect, and we provide an experimental system that facilitates further investigation of this group in a genetically tractable model organism.

Evolution of phenotypes: understanding diversity and the role of plasticity in adaptation to new environments
SMBE-PO-094

Exploring protein phenotypes by translational errors in yeast

A. S. Kleppe [1,*], E. Bornberg-Bauer [1]
[1]Institute of Evolution and Biodiveristy, University of Münster, Münster, Germany

Abstract: Mutations are the source of genetic novelty and diversity. Thus, mutations are needed for a trait to be readily adaptable in an alternating environment. Simultaneously, most mutations are deleterious and are therefore selected against. These conflicting roles of mutations are poorly understood with respect to protein evolvability. We have analysed translational errors (TrE) in yeast by ribosome profiling data, to elucidate the bio-physical context that enables or hinders a protein to evolve. We find that expression and codon usage correlate negatively and strongly with TrE rate, suggesting that codon usage diminishes the rate but fail to eliminate TrE entirely. We also find that error-prone proteins have more disordered C-termini, which we believe may offer structural robustness to initial protein fold and function. Furthermore, by inter-species comparisons we have investigated the evolutionary rate (by McDonald-Kreitman test) of error-prone proteins and their orthologs. We find that there is no selective pressure to reduce - or enable - TrE. Effectively, we conclude that the investigated TrE are neutral, or near-neutral, as they seemingly go undetected by natural selection. I will present and discuss these findings, specifically the role or TrE in protein plasticity and phenotypic landscape exploration.

Evolution of phenotypes: understanding diversity and the role of plasticity in adaptation to new environments
SMBE-PO-091

Trade-offs driven by density-dependent selection in free-living mammalian cell lines

T. Li, J. Liu, J. Feng*, X. Lu [1]
[1]Kunming Institute of Zoology, NO 32. Jiaochang East Road, Kunming,Yunnan, China

Abstract: Evolutionary life history theory proposes that the trade-offs, such as those between reproduction and survival, apply to the organisms that are subject to natural selection to help to determine the evolution of phenotypes. It leads to a puzzle how rapid cell proliferation, cell survival strategies and other cancer hallmarks could be entirely achieved within an individual tumor in complex and challenging microenvironments. The hypotheses can be that trade-offs have been developed among the heterogeneous neoplastic cell populations during cancer evolution. In addition, the evolutionary selection pressures and adaptive strategies that govern the trade-off between increasing proliferation and survival for cancer phenotypes are currently unknown. In this study, we apply the theory of density-dependent (also named r- and K-) selection, which was one of the first and most refined models for life history, to puzzle out the evolutionary process. The central idea of r/k selection theory is that populations in stable environments with limited resources evolve slow life histories; environments with rapid and stochastic fluctuations in resource availability select for fast life histories. We perform the experimental scheme of r- and K-selection using a clonal cancer cell line for cell survival at high density and fast cell proliferation at low density. After 200 generations r/k selection, adaptation has evolved. The r and K cells show better fitness than their counterparts at high and low density, respectively. RNA-seq analysis indicates that the trade-off phenotypes in r and K cells are associated with their distinguished pattern of gene expression in cell cycle, adhesion, apoptosis, contact inhibition and so on. Intriguingly, both empirical observations and estimations of mathematical modelling demonstrate that r and k cells are competitively interacting when co-existing, and the interaction conduces to the total population growth. This experimental evolutionary study in Hela cell line suggests that cancer cells may be subject to trade-offs between cell proliferation and survival which have emerged during somatic evolution driven by r- and K-selection, resulting in the improvement in whole fitness of a tumor with heterogeneous cell populations.

Evolution of phenotypes: understanding diversity and the role of plasticity in adaptation to new environments
SMBE-PO-093

Adaptation to new environment: the pH challenge

R. Di Martino [1,*], S. Mitri [1]
[1]DMF - Department of Fundamental Microbiology, University of Lausanne, Lausanne, Switzerland

Abstract: Perturbations of environmental parameters such as pH or carbon sources induce microbes to develop various strategies to maintain high fitness in novel conditions. In this study, we focus on an *Agrobacterium tumefaciens* strain isolated from a four-species community growing in a metal-working fluid (MWF), a coolant and lubricating agent. To untangle the metabolic pathways of *A. tumefaciens* in this environment, we designed a minimal medium in which different concentrations of carbon sources commonly present in the MWF were added separately. Among these tested compounds, citric acid got our attention: after several days of incubation, *A. tumefaciens* switched from no growth to a sudden 20-fold increase compared to the starting inoculum. Growth occurred only at the highest-tested citric acid concentration simultaneously among all the replicates. Moreover, this specific concentration has a very acidic pH, and we verified that cells collected from the stationary phase could grow instantly in the same fresh medium but not in a neutralized citric acid medium. Our hypothesis is that the exposure to the acidic condition triggered an adaptive response. We are evaluating what has changed both at the genomic and transcriptomic level. Our aim is to understand the impact of pH on metabolic functions of *A. tumefaciens* and its ability to adapt to extreme environments. The result of this experiment will give us more insight on how bacteria react to external variations, highlighting which are the pathways responsible for fitness preservation in a new environment.

Evolution of phenotypes: understanding diversity and the role of plasticity in adaptation to new environments
SMBE-PO-087

The shapes of macronuclei in ciliates

W.-J. Chang [1,*], K. Miller [1], N. Stover [2], N. Nannas [1]
[1]Biology, Hamilton College, Clinton, [2]Biology, Bradley University, Peoria, United States

Abstract: One of the identifying features of ciliates that is shared among all ciliate classes is nuclear dualism, which consists of the presence of at least one macronucleus and at least one micronucleus. The specialized somatic macronucleus and germline micronucleus work in tandem to fulfill the nuclear requirements of the cells. One unusual aspect of the macronucleus that distinguishes it from almost all other eukaryotic nuclei is that it can take on a wide variety of distinct shapes in different species of ciliates. Most eukaryotic cell types have spherical nuclei, however the macronuclei of ciliates often have shapes ranging from ovoid or spherical to tubular, nodular, and bean-shaped. Until now, no evolutionary benefit or explanation has been identified that would justify an advantage to these alternative macronuclear shapes. In most eukaryotes, the nuclear envelope and cytoskeleton, made up of microtubules, lamin intermediate filaments, and actin microfilaments, support and maintain a spherical nucleus within the cell. The cytoskeleton and macronuclear envelope of ciliates share many similarities with most other eukaryotes, despite the unusual shape of the ciliate macronucleus. Here we construct a database classifying shapes of macronuclei from more than 200 ciliates to help analyze trends, if any, of the evolution of diverse shapes of macronuclei.

Evolution of phenotypes: understanding diversity and the role of plasticity in adaptation to new environments
SMBE-PO-090

Gene amplification as a primitive mechanism for gene expression regulation

I. Tomanek [1,*], R. Grah [1], M. Lagator [1 2], A. Andersson [1], J. Bollback [3], G. Tkačik [1], C. Guet [1]
[1]IST Austria, Klosterneuburg, Austria, [2]The University of Manchester, Manchester, [3]University of Liverpool , Liverpool, United Kingdom

Abstract: Transcriptional regulation enables fast adjustment of gene expression levels to environmental changes. But what if no dedicated regulatory circuit is available? We ask whether the intrinsic processes of duplication and amplification, which modify expression by changing gene copy number, can act as a form of gene regulation. Using an experimental system designed to monitor gene copy number in real time, we show that under fluctuating selection amplifications lead to rapid adaptive alterations to gene expression levels of the population. This 'amplification-mediated gene expression tuning' depends on the continual introduction of expression polymorphisms that are rapidly acted upon by selection and ultimately also rapidly removed. Mathematical modeling shows that amplifications successfully tune gene expression in environments where transcription factor-based schemes are hard to evolve or maintain. The fleeting nature of gene amplifications gives rise to a universal population-level mechanism for rapidly tuning expression of any gene, without leaving any genomic signature.

Evolution of phenotypes: understanding diversity and the role of plasticity in adaptation to new environments
SMBE-PO-124

The Komodo dragon genome reveals adaptations in the cardiovascular, muscular, and chemosensory systems of monitor lizards

A. Lind [1,*], Y. Lai [2], Y. Mostovoy [2], A. Holloway [1], A. Iannuci [3], A. Mak [2], M. Fondi [2 3], V. Orlandini [2 3], W. Eckalbar [4], M. Milan [5], M. Rovatsos [6 7], I. Kichigin [8], A. Makunin [8], M. Pokorná [6], M. Altmanová [6], V. Trifonov [8], E. Schijlen [9], L. Kratochvíl [6], R. Fani [3], P. Velenský [10], I. Rehák [10], T. Patarnello [5], T. Jessop [11], J. Hicks [12], O. Ryder [13], J. Mendelson [14 15], C. Ciofi [3], P.-Y. Kwok [4 16 17], K. Pollard [1 4 18 19 20], B. Bruneau [1 16 21]

[1]Gladstone Institutes, [2]UCSF, San Francisco, United States, [3]Department of Biology, University of Florence, Sesto Fiorentino, Italy, [4]Institute for Human Genetics, UCSF, San Francisco, United States, [5]Department of Comparative Biomedicine and Food Science, University of Padova, Legnaro, Italy, [6]Department of Ecology, Charles Univeristy, Prague, [7]Institute of Animal Physiology and Genetics, The Czech Academy of Sciences, Liběchov, Czech Republic, [8]Institute of Molecular and Cellular Biology, SB RAS, Novosibirsk, Russian Federation, [9]BU Bioscience, Wageningen University, Wageningen, Netherlands, [10]Prague Zoological Garden, Prague, Czech Republic, [11]Centre for Integrative Ecology, Deakin University, Waurn Ponds, Australia, [12]Department of Ecology and Evolutionary Biology, University of California at Irvine, Irvine, [13]Institute for Conservation Research, San Diego Zoo, Escondido, [14]Zoo Atlanta, [15]School of Biological Sciences, Georgia Institute of Technology, Atlanta, [16]Cardiovascular Research Institute, [17]Department of Dermatology, [18]Department of Epidemiology and Biostatistics, [19]Institute for Computational Health Sciences, UCSF, [20]Chan-Zuckerberg BioHub, [21]Department of Pediatrics, UCSF, San Francisco, United States

Abstract: Monitor lizards are unique among ectothermic reptiles in that they have a high aerobic capacity and distinctive cardiovascular physiology which resembles that of endothermic mammals. We have sequenced the genome of the Komodo dragon (Varanus komodoensis), the largest extant monitor lizard, and present a high resolution de novo chromosome-assigned genome assembly for V. komodoensis, generated with a hybrid approach of long-range sequencing and single molecule physical mapping. Comparing the genome of V. komodoensis with those of related species showed evidence of positive selection in pathways related to muscle energy metabolism, cardiovascular homeostasis, and thrombosis. We also found species-specific expansions of a chemoreceptor gene family related to pheromone and kairomone sensing in V. komodoensis and several other lizard lineages. Together, these evolutionary signatures of adaptation reveal genetic underpinnings of the unique Komodo sensory, cardiovascular, and muscular systems, and suggest that selective pressure altered thrombosis genes to help Komodo dragons evade the anticoagulant effects of their own saliva. As the only sequenced monitor lizard genome, the Komodo dragon genome is an important resource for understanding the biology of this lineage and of reptiles worldwide.

Evolution of phenotypes: understanding diversity and the role of plasticity in adaptation to new environments
SMBE-PO-089

Transcriptional Rewiring, Adaptation, and the Role of Gene Duplication in the Metabolism of Ethanol of Saccharomyces cerevisiae

C. Toft [1 2,*], F. Mattenberger [1 2], M. A. Fares [1 2 3], B. Sabater-Muñoz [2 3]
[1]Institute for Integrative Systems Biology (I2SysBio, CSIC-UV), [2]Instituto de Biología Molecular y Celular de Plantas, Consejo Superior de Investigaciones Científicas-Universidad Politécnica (CSIC-UPV), Valencia, Spain, [3]Smurfit Institute of Genetics, Department of Genetics, University of Dublin, Trinity College, Dublin, Ireland

Abstract: The Baker's yeast *Saccharomyces cerevisiae*is one of the most important biotechnologically microbes because of its ability to produce and tolerate high levels of ethanol. Due to these properties, much interest has been put in understanding the genetic underpinnings of alcohol tolerance to improve the quality and productions of the products. A reason study by Voordeckers et al (2015) has shown that adaptation to high ethanol tolerance involves point mutation, copy number variation, ploidy changes, and clonal interference. Nonetheless, the transcriptional re-wiring occurring during the response or adaptation to ethanol and its importance in comparison with the contribution to mutations in coding changes has not been explored. With both copy number variation and change of ploidy involved in the adaptation to ethanol, indicates that duplication plays an important role in the adaptation. We have already observed in our previous studies that ancient duplicates (small-scale and whole genome duplicates) play an important role in acute ethanol response (Mattenbergeret al. 2017). Here we use experimental evolution of *S. cerevisiae*to uncover the transcriptional rewiring during ethanol adaptation and its link with gene duplication. First the populations of *S. cerevisiae*were evolved for hundreds of generations for the purpose of diversification and later we evolved the populations with ethanol being the only carbon source. Little overlap was observed between enriched functional categories of up-regulated genes in response and adaptation to ethanol. In particular, we see a significantly larger proportion of duplicated genes responding to ethanol stress than singletons. Moreover, the fold change in duplicated genes was higher compared to singletons. All of which indicate that duplicated genes play a central role in the acute and chronic response to ethanol.

Voordeckers et al. 2005 PLoS Genet.
Mattenbergeret al. 2017 DNA Res

Evolution of phenotypes: understanding diversity and the role of plasticity in adaptation to new environments
SMBE-PO-084

Allele-specific single-cell RNA sequencing reveals different architectures of intrinsic and extrinsic gene expression noises

M. Sun [1,*], J. Zhang [1]
[1]Ecology and Evolutionary Biology, University of Michigan, Ann Arbor, United States

Abstract: Gene expression noise refers to the variation of the expression level of a gene among isogenic cells in the same environment, and has two sources: extrinsic noise arising from the disparity of the cell state and intrinsic noise arising from the stochastic process of gene expression in the same cell state. Due to the low throughput of the existing method for measuring the two noise components, the genome-wide architectures of intrinsic and extrinsic expression noises remain elusive. Using allele-specific single-cell RNA sequencing, we here estimate the two noise components of 3975 genes in mouse fibroblast cells. The noise estimates obtained allowed us to evaluate the predicted effects of various factors. For instance, we found that the presence of the TATA-box in the promoter of a gene increases both the intrinsic and extrinsic expression noise of the gene, whereas miRNAs lower the intrinsic noise but increase the extrinsic noise of their target genes. Considering gene functions, we predict and demonstrate that nuclear genes functioning in the mitochondrion have reduced extrinsic noise, genes encoding protein complex members have decreased intrinsic noise, and cell cycle genes have lowered intrinsic noise but elevated extrinsic noise. Hypothesis-independent exploration of our noise data further revealed functional groups of genes with exceptionally large or small intrinsic and/or extrinsic expression noise. These findings unravel differential regulations, optimizations, and biological consequences of intrinsic and extrinsic noises and can aid the construction of desired synthetic circuits.

Evolution of phenotypes: understanding diversity and the role of plasticity in adaptation to new environments
SMBE-PO-116

The role of the epigenome in plastic responses to rapidly changing social environments

J. Rouse [1,*], E. Duncan [1], T. Leech [2], A. Bretman [1]
[1]University of Leeds, Leeds, United Kingdom, [2]Max Planck Institute for biology of ageing, Cologne, Germany

Abstract: Social environments can be very dynamic and have important consequences for fitness. For example, males use social information to predict the amount of mating competition they will face and adjust investment in particular mating opportunities accordingly. To accurately match their reproductive strategy to this fluctuating environment requires plasticity that is fast acting and reversible. Whilst the role of the epigenome in responses to environmental change is becoming more understood, it has been suggested that it is not invoked in such flexible plasticity. We sought to test this using a *Drosophila melanogaster* fruit fly model. Male flies adjust their mating duration and ejaculate composition depending on the level of sperm competition signalled by exposure to rival males before mating. The behavioural component of this change occurs quickly (under 24h) and is entirely reversible. A previous transcriptome study showed that the expression of some epigenetic modifiers is sensitive to exposure to rival males. Here, we used chemical inhibitors and RNAi to test whether epigenetic remodelling is required to achieve this plastic response. We found histone deacetylation is crucial to a male's ability to plastically respond to increased sperm competition, but in a tissue specific manner. We are now investigating how this plasticity alters the epigenetic landscape using ChIP-seq. Overall, this suggests epigenetic remodelling is an important mechanism in short-term, reversible plastic phenotypes, and must be considered when exploring adaptations to fluctuating environments.

Evolution of phenotypes: understanding diversity and the role of plasticity in adaptation to new environments
SMBE-PO-099

Phenotypic heterogeneity promotes adaptive evolution

C. Pal [1,*]
[1]Biological Research Centre, Szeged, Hungary

Abstract: Phenotypic heterogeneity of genetically identical cells can generate nonheritable variation in a population. Is this heterogeneity favorable for microbes? In a changing environment, the answer is a definite yes. While scholars have argued that stochastically generated variation precedes genetic changes and thereby facilitate the evolution of complex traits, this idea has remained disputed, not least because of the shortage of experimental studies. We address this long-standing and controversial issue by integrating synthetic biology, laboratory experimental evolution, and genomic analyses. We explicitly tested the mechanisms whereby phenotypic heterogeneity may promote evolvability. Our work demonstrates that phenotypic heterogeneity facilitates evolutionary rescue from deteriorating environmental stress by generating individuals with exceptionally high fitness. Remarkably, elevated phenotypic heterogeneity evolves as a direct response to stress and thereby it promotes evolution of rare combinations of mutations. These results indicate that phenotypic heterogeneity might have an important role in the evolution of key innovations.

Evolution of phenotypes: understanding diversity and the role of plasticity in adaptation to new environments
SMBE-PO-097

The role of the enhancers in the intraspecific variability

E. Minkina [1,*], N. Akberova [2], E. Shagimardanova [3], O. Gusev [1 4], I. Adameyko [5 6]
[1]Institute of Fundamental Medicine and Biology, [2]Institute of Fundermental Medicine and Biology, Kazan (Volga Region) Federal University, [3]Institute of Fundermental Medicine and Biology, Kazan (Volga Region) Federal University, Kazan, Russian Federation, [4]RIKEN, Yokohama, Japan, [5]Department of Physiology and Pharmacology, Karolinska Institutet, Stockholm, Sweden, [6]Department of Molecular Neurosciences, Medical University Vienna, Vienna, Austria

Abstract: Sequence changes within the protein-coding genes can lead to substantial disruptions in their work and, consequently, to diseases. Such inflexible affect cannot to be responsible for the normal traits variability within a species, such as the craniofacial shape in humans. It's assumed that the morphological divergence between related scpecies conditioned by a fine tuning of gene activity derived by enhancers. That implies also hyper-variable enhancers within the species must be involved in intraspecies variability.

In the current study we combined polymorphisms data with the predicted human enhancers from two fundamental projects (FANTOM5, n = 65,423; RoadMap, n = 1,598,323), as well as a enhancers set identified in a study of human embryonic craniofacial tissues (n = 75,928), to investigate the role of hypervariable enhancers in the traits variability. We calculated the enhancer variability scores, defined as the ratio of the variants number to the enhancer length and normalized to the entire genome variants ratio.

We observed no correlation the intraspecific variability of enhancer regions with the interspecific conservativity of these regions between vertebrates, that can imply the presence of some positive selection for variable enhancers in human population. It also supports the idea that hypervariable enhancers can participate in fine tuning of traits without disrupting the functionality of essential regulatory networks. We also tested enhancers of 40 most important craniofacial genes and find a hight intraspecific variability for each of them, however interspecific conservation still different within this group.

Our followed analysis of hyper-variable enhancers in different human populations will provide additional understunding of their role in intraspecific variability.

This study was supported by Russian Science Foundation grant No 19-14-00260.

Sequence changes within the protein-coding genes can lead to substantial disruptions in their work and, consequently, to diseases. Such inflexible affect cannot to be responsible for the normal traits variability within a species, such as the craniofacial shape in humans. It's assumed that the morphological divergence between related scpecies conditioned by a fine tuning of gene activity derived by enhancers. That implies also hyper-variable enhancers within the species must be involved in intraspecies variability.In the current study we combined polymorphisms data with the predicted human enhancers from two fundamental projects (FANTOM5, n = 65,423; RoadMap, n = 1,598,323), as well as a enhancers set identified in a study of human embryonic craniofacial tissues (n = 75,928), to investigate the role of hypervariable enhancers in the traits variability. We calculated the enhancer variability scores, defined as the ratio of the variants number to the enhancer length and normalized to the entire genome variants ratio.We observed no correlation the intraspecific variability of enhancer regions with the interspecific conservativity of these regions between vertebrates, that can imply the presence of some positive selection for variable enhancers in human population. It also supports the idea that hypervariable enhancers can participate in fine tuning of traits without disrupting the functionality of essential regulatory networks. We also tested enhancers of 40 most important craniofacial genes and find a hight intraspecific variability for each of them, however interspecific conservation still different within this group.Our followed analysis of hyper-variable enhancers in different human populations will provide additional understunding of their role in intraspecific variability.This study was supported by Russian Science Foundation grant No 19-14-00260.

Evolution of phenotypes: understanding diversity and the role of plasticity in adaptation to new environments
SMBE-PO-095

Dynamic evolution of bitter taste receptor gene repertoire in Old World monkeys

M. Hou [1,*], M. Hayashi [1], A. Melin [2], S. Kawamura [1]
[1]Department of Integrated Bio-science, University of Tokyo, Kashiwa, Japan, [2]Department of Anthropology and Archaeology & Department of Medical Genetics, University of Calgary, Calgary, Canada

Abstract: Taste perception is fundamental in food selection for many animals. Bitter taste perception is important not only in dietary selection but also in preventing animals from ingesting potentially toxic compounds. Previous studies have revealed evolutionary divergence of bitter taste receptor gene (*TAS2R*) repertoire in mammals including primates using publicly available whole genome sequence (WGS) data. Old World monkeys are an excellent subject for studying adaptive evolution of bitter sense because they are dietarily diverged into folivorous colobines and omnivorous cercopithecines within which many genera are further diverged. However, only a few genera with WGS data available have been subjected for evolutionary study of *TAS2R*s. Dependence on WGS data is also a matter of concern due to its inherent incompleteness especially for multigene families such as *TAS2R*s. In this study, we employed the target capture method specifically probing *TAS2R*s followed by massive-parallel sequencing for eight species of Old World monkeys (six cercopithecines: two *Papio*, one each of *Macaca*, *Cercopithecus*, *Chlorocebus* and *Erythrocebus* species; two colobines: one each of *Semnopithecus* and *Colobus* species). Though still in progress, our preliminary phylogenetic analysis showed various patterns of evolutionary gain and loss of *TAS2R*s in these species, prompting further analyses for possible connection to dietary adaptation.

Evolution of phenotypes: understanding diversity and the role of plasticity in adaptation to new environments
SMBE-PO-098

A bayesian model of phenotypic evolution unveils random forces as drivers of different evolutionary rates of alternative splicing

C. Martí-Gómez*, C. Vogl [1], F. Sánchez-Cabo [2], E. Lara-Pezzi [3]
[1]Department of biomedical sciences, University of Veterinary Medicine, Vienna, Austria, [2]Bioinformatics Unit, [3]Molecular regulation of heart failure, Fundación Centro Nacional de Investigaciones Cardiovasculares Carlos III (CNIC), Madrid, Spain

Abstract: Alternative splicing (AS) is thought to underlie major phenotypic changes along species evolution and divergence. However, these results are based on descriptive statistics that are compatible also with neutral evolution of AS patterns. To gain further insight into the evolution of AS, one can use models of phenotypic evolution over a phylogenetic tree, such as the Browniam Motion (BM) model. However, application of such models to transcriptomic data remains challenging due to the small number of species for which data are available and the noisy nature of such data.

In this study, we use a generalization of the BM model, an Ornstein-Uhlenbeck (OU) process, to model AS evolution. The OU model combines directional evolution towards a phenotypic optima with random changes accumulating over evolutionary time as in the BM model. We extended the basic OU model to include intraespecific variability and measurement error and to take into account the binomial nature of AS data from RNA-seq experiments. We then performed bayesian inference of the underlying parameters by sampling from the joint posterior distribution using a Markov Chain Monte Carlo (MCMC) algorithm.

We showed that the limitation of using small phylogenies can be circumvented by pooling information from a relatively high number of exons and reach accurate parameter inferences using simulated data. Fitting the model to a real dataset, we find that most putatively alternatively spliced exons are actually nearly constitutively spliced and that the set of high confidence alternatively spliced exons is small. Inclusion levels of alternatively spliced exons evolved relatively slowly mainly due to a reduced stochastic rather than to stronger deterministic effects.

Evolution of phenotypes: understanding diversity and the role of plasticity in adaptation to new environments
SMBE-PO-105

Evolution of growth rate and its plasticity in response to light in natural Arabidopsis thaliana populations

B. Wieters [1,*], J. de Meaux [1]
[1]University of Cologne, Cologne, Germany

Abstract: Genome-wide association studies are popular approaches to search for signatures of local adaptation in genomes. *Arabidopsis thaliana* covers a range of highly contrasting environments, which led to broad patterns of local adaptation and to an array of plastic responses. We investigated the growth response to light, which is crucial for plant performance and strongly variable across the distribution range, in an attempt to understand how a trait of crucial ecological relevance was remodeled by local selective pressures.
Our study entailed genotypes from three broad regions, Iberic Peninsula, Scandinavia and South China. The Iberian genotypes reached higher final rosette diameter, while Northern European genotypes exhibited faster growth. Moreover, Spanish and Chinese genotypes displayed plasticity differences between treatments. Fst/Qst comparisons were indicative of a signature of local adaptation. GWAS of European accessions revealed numerous SNPs associated with different aspects of the growth response including candidate genes related to light, auxin or transcription. Polygenic scores indicated that GWAS hits captured approximately 25% of the phenotypic variance we quantified, yet we also observed that they were strongly biased towards outliers of population structure. Plant final size in high light was negatively correlated with the number of rare loss-of-function alleles, which suggests that it is sensitive to the mutation load. Taken together, our results suggest that the plant growth response in *A. thaliana* is locally adapted. Yet, our work also illustrates the pitfalls of searching for genomic signatures of local adaptation in highly structured populations.

Evolution of phenotypes: understanding diversity and the role of plasticity in adaptation to new environments
SMBE-PO-125

Adaptive changes in anthrax toxin receptor expression in humans

L. A. Choate [1 2,*], P. McMahon [1], I. Said [2], S. Lin [1], J. J. Lewis [1], C. G. Danko [1]
[1]Baker Institute for Animal Health, [2]Molecular Biology and Genetics, Cornell University, Ithaca, United States

Abstract: Modern humans were exposed to new environmental pathogens during the development of agriculture. Anthrax disease is caused by a soil borne bacterium whose natural host is grazing herbivores, but can be passed to a secondary host, including humans, through contact with infected animals. Multiple lines of historical evidence suggest that anthrax disease was an important source of mortality during human history, especially in Europe and the Fertile Crescent. Here we show that anthrax toxin receptor 2 (*ANTXR2*), the gene encoding a transmembrane protein required for anthrax toxins to access host cells, is 8-fold down-regulated in human CD4+ T-cells compared with other primates. We used CRISPR activation to restore human *ANTXR2* expression to non-human primate levels, revealing a 2-fold difference in sensitivity to recombinant anthrax toxins. Using genome-wide maps of enhancer activity in primary CD4+ T-cells from multiple human and non-human primates, we identified the precise coordinates of 9 putative enhancers in a 100 kb gene desert upstream of *ANTXR2*. Enhancers in this region were enriched for higher Hi-C interaction frequency with the *ANTXR2* promoter, and often overlapped *ANTXR2* expression quantitative trait loci (eQTLs). Finally, several enhancers were outliers for population differentiation (Fst) in Europeans, consistent with recent selection. Using a combination of machine learning, luciferase assays, and CRISPR genome editing, we identified DNA sequence changes responsible for differences in *ANTXR2* expression. Our results suggest that exposure to anthrax disease may have driven adaptive changes in *ANTXR2* during the advent of agriculture.

Evolution of phenotypes: understanding diversity and the role of plasticity in adaptation to new environments
SMBE-PO-109

Assessing relationships between genetic and epigenetic variation in the context of phenotypic plasticity in Daphnia magna

I. W. Wilson [1,*], F. Brunner [1], A. Reynolds [1], S. Price [1], D. Atkinson [1], S. Paterson [1], S. Plaistow [1]
[1]Department of Evolution, Ecology and Behaviour, University of Liverpool, Liverpool, United Kingdom

Abstract: Unlike 'traditional' genetic mechanisms, epigenetic inheritance is thought to allow transmission of environment-induced changes in phenotypes through generations. Despite an increasing body of work on epigenetic signals and phenotypic plasticity, however, we still know very little about the relationships between genetic and epigenetic changes. We present here an overview of an ongoing novel comparison of genomic and epigenomic data from 12 *Daphnia magna*clones isolated from a single pond as well as qualitative phenotypic plasticity data from a previous study. We will use these datasets to assess the correlation between diversity in DNA methylation patterns and genetic variation and investigate whether the most heavily methylated loci in a genome are also the most variable between genotypes, and whether particular methylation patterns can be associated with relative plasticity.

We have employed a combination of PacBio and Illumina technologies to sequence and *de novo*assemble the genomes of our 12 clones. This hybrid technique combines the greater contiguity and lesser GC-bias of PacBio assemblies with the low error rate of Illumina sequencing, thus improving upon existing *D. magna*resources. These reliable assemblies have allowed us to assess genetic variation between the clones studied here and to identify those loci and genes that are under the greatest positive selection pressures and so most susceptible to variation, even within a single environment.

We have also performed Whole Genome Bisulphite Sequencing for these clones to identify variations in DNA methylation patterns both within and between clonal populations. These data sets have been combined with our group's previously generated plasticity data – based upon plastic responses to temperature and food availability – and will allow us to compare and assess both genetic and epigenetic variation within and between clones, in the context of phenotypic plasticity. Findings from this unique combination of genetic, epigenetic and phenotypic datawill inform the discussion concerning the relationship between these mechanisms of inheritance and offer further insight into their role, and the role of plasticity, in adaptation and evolution.

Findings of this experiment will also be integrated into our ongoing study into retention and decay of methylation patterns in response to environmental stimuli, as well as our wider project investigating the roles of genetic diversity, phenotypic plasticity and epigenetics in adaptation to climate change in mesocosms, which includes tracking changes in epiallele dynamics in semi-wild populations.

Evolution of phenotypes: understanding diversity and the role of plasticity in adaptation to new environments
SMBE-PO-100

Cross-altitude analysis suggests a turning point at the elevation of 4,500 m for polycythemia prevalence in Tibetans

X. Qi [1,*], H. Zhang [1], C. Cui [2], O. NA [2], B. Su [1]
[1]State Key Laboratory of Genetic Resources and Evolution, Kunming Institute of Zoology, Chinese Academy of Sciences, Kunming, [2]School of Medicine, Tibetan University, Lhasa, China

Abstract: Tibetans are well adapted to high altitude environments, and with a great majority of them living under the elevation of 4,500m. It is understood that the relatively low hemoglobin (Hb) concentration in Tibetans protects them from polycythemia (red cell overproduction). In general, Tibetans showed lower Hb concentration compared with Han immigrants residing in Tibet, and there was a linear correlation of Hb concentration with altitude. To test whether there is a turning point of elevation for polycythemia incidence in Tibetans, we collected data of Hb concentration from nearly 8,000 Tibetan individuals residing in 20 geographic regions throughout the Qinghai-Tibetan Plateau, with elevations ranging from 2,227m to 5,018m. For populations living below 4,500m, the average Hb levels are low and incidences of polycythemia are rare, an indication of adaptation for Tibetans at relatively low altitude. In contrast, the average Hb levels and incidences of polycythemia deviated from the linear expectation for the high elevations (>4,500m). In particular, at 5,000m, the majority of the population (67% in males and 65% in females) are subject to polycythemia. The nonlinear regression curves indicate that the Hb levels increase slowly with elevations from 2,200m to 4,500m, and then sharply after 4,500m. The same pattern was observed for the curve of incidence of polycythemia. We showed that the elevation of 4,500m marks a sharp increase of Hb concentration and incidence of polycythemia in Tibetan populations living above the elevation of 4,500m, which may reflect the altitude limit for Tibetans' adaptation to high altitude.

Evolution of phenotypes: understanding diversity and the role of plasticity in adaptation to new environments
SMBE-PO-102

Genomic regions showing evidence for adaptive introgression in an admixed population have undergone natural selection in the ancestral population.

M. Isshiki [1,*], I. Naka [1], R. Kimura [2], N. Nishida [3], M. Nakazawa [4], T. Inaoka [5], Y. Matsumura [6], R. Ohtsuka [7], J. Ohashi [1]
[1]Graduate School of Science, The University of Tokyo, Tokyo, [2]Graduate School of Medicine, University of the Ryukyus, Nishihara, [3]Research Center for Hepatitis and Immunology, National Center for Global Health and Medicine, Chiba, [4]Graduate School of Health Sciences, Kobe University, Kobe, [5]Faculty of Agriculture, Saga University, Saga, [6]Faculty of Health and Nutrition, Bunkyo University, Chigasaki, [7]Japan Wildlife Research Center, Tokyo, Japan

Abstract: According to archaeological evidence, there are two major waves of human settlement into Oceania. The first settlement occurred in Near Oceania (e.g. New Guinea, Australia and the Solomon Islands) about 5-40,000 years ago by the ancestry of modern Papuans and Aboriginal Australians. The second one occurred about 5-3,000 years ago by Austronesian-speaking Populations originated in Asia. They reached Near Oceania and admixed with the descendants of first settlers before the occupation of Remote Oceania (e.g. Micronesia and Polynesia). Thus, modern Polynesians have these two ancestries, Papuan-related and Asian-related ancestries and mainly derived from the latter. Under the assumption that Papuan-related ancestry should have adapted to Oceanian environment much more than Asian-related ancestry, we hypothesized “adaptive introgression” from Papuan-related ancestry to Asian-related ancestry occurred in Polynesian genomes. To examine our hypothesis, the proportion of local ancestry was estimated across Polynesian genome using genome-wide SNP data of Polynesians (Tonga, n=24), modern Papuans (n=24) and Asians (n=45). The mean proportion of Papuan-related ancestry (25.5%) across genome was equivalent to values obtained in previous studies. Interestingly, we detected three genomic regions on chromosomes 6, 10, and 16 exhibiting significantly higher proportion (> 66.8% [mean + 4SD]) of Papuan-related ancestry. The regions on chromosomes 6 and 10, which contain genes involved with immune systems and metabolism, were further found to have undergone recent positive selection in modern Papuans (n=14). The present results suggest that the same driving force acted in ancestry and admixed populations living in similar geographic regions.

Evolution of phenotypes: understanding diversity and the role of plasticity in adaptation to new environments
SMBE-PO-101

The role of miRNAs in the hibernation of the central bearded dragon.

A. Capraro [1,*], D. O'Meally [2], A. Georges [3], P. Waters [1]
[1]School of Biotechnology and Biomolecular Sciences, UNSW Australia, Sydney, Australia, [2]Center for Gene Therapy, Beckman Research Institute of the City of Hope, Duarte, United States, [3]Institute for Applied Ecology, University of Canberra, Canberra, Australia

Abstract: Hibernation is a broad physiological state utilized by many animals to cope with the adverse environmental conditions associated with winter. Hibernation involves a large restructuring of cellular metabolism and function resulting from tissue-specific adaptive responses, in addition to an organism-wide response. Epigenetic mechanisms such as microRNA-mediated gene silencing, chromatin modifications, and DNA methylation are known to govern these changes in mammals, however, the role of epigenetics has yet to be uncovered in reptilian hibernation. Small RNA transcriptomic profiles of the central bearded dragon (*Pogona vitticeps*) were generated for brain, heart and skeletal muscle during late hibernation and at two months post-arousal to uncover changes in miRNA expression during the hibernation season. Across all tissue types and conditions, we discovered 1090 miRNAs; many of which have yet to be characterised. We detected tissue-specific differential expression of miRNAs, as well as differential expression of miRNAs that was common to all tissues. Subsequent gene ontology enrichment analysis of the miRNA targets revealed key pathways necessary for survival during hibernation. Furthermore, correlation between miRNA expression and target RNA expression during hibernation provides evidence for active gene silencing of hibernation-responsive genes via miRNA. This is the first study to explore miRNAs in reptiles in the context of hibernation and provides exciting insights into the role of miRNAs in regulating the expression of genes critical to hibernation in the central bearded dragon.

Evolution of phenotypes: understanding diversity and the role of plasticity in adaptation to new environments
SMBE-PO-122

Characterisation of structural variants in the adaptive radiation of African Lakes Cichlids

A. Man [1,*], L. Penso-Dolfin [1 2], W. Haerty [1], F. Di-Palma [1]
[1]Earlham Institute, Norwich, United Kingdom, [2]Deutsches Krebsforschungszentrum, Heidelberg, Germany

Abstract: African Lakes Cichlids are an impressive example of adaptive radiation with several hundreds of species arising independently in Lake Victoria, Tanganyika, and Malawi, within the last 10 million to 100,000 years. Genetic basis for this explosive radiation is well documented, but no studies have focussed on the contribution of structural variants (SVs) to speciation (via reduction of gene flow) and adaptation to different ecological niches. Here, we annotate and characterise the repertoires and evolutionary potential of different SV classes (deletion, duplication, inversion, insertions and translocations) in five Cichlid species (*Haplochromis burtoni, Metriaclima zebra, Neolamprologus brichardi, Pundamilia nyererei* and *Oreochromis niloticus*). Altogether, we provide the first overview of rearrangement evolution in East African Cichlids, and some initial insights into their possible contribution to adaptation. Interestingly, we find enrichment of genes regulating behaviour, of those involved in skeletal and visual system development, and genes associated with antigen processing and presentation and other immune related categories. The pipeline and biological results were further tested by PCR validation of selected deletions and inversions, which confirmed 7 out 10 and 6 out of 9 events, respectively.

Evolution of phenotypes: understanding diversity and the role of plasticity in adaptation to new environments
SMBE-PO-119

Phenotypic plasticity and crop domestication

J. Saban [1,*], L. Sage [1], M. Chapman [1]
[1]University of Southampton, Southampton, United Kingdom

Abstract: The extent to which phenotypic plasticity facilitates or buffers against adaptive divergence remains contentious. In our work we are using domestication in *Brassica* species as a model and transcriptomics as the tools to study the role of phenotypic plasticity in adaptive evolution. This preliminary work indicates that morphological and transcriptome plasticity differs between wild and domesticated *Brassica rapa*. Expression plasticity has been lost (canalised) in some genes, which we hypothesise is due to human selection, and we suggest that plasticity could have served as the 'raw material' for crop domestication.

Evolution of phenotypes: understanding diversity and the role of plasticity in adaptation to new environments
SMBE-PO-121

Investigating genome variation in a South African wild dog (Lycaon pictus) population: Towards understanding their susceptibility to Mycobacterium bovis infection

C. Meiring [1,*], M. Moller [1], M. A. Miller [1], L. Kleynhans [1], C. Kinnear [1]
[1]Biomedical Sciences, Stellenbosch University, Cape Town, South Africa

Abstract: The African wild dog population is declining rapidly and is at high-risk for future extinction with only 5000–6000 individuals left in the world. The population decline is due to habitat fragmentation, human-carnivore conflict and infectious disease. Management of wild dogs relies on the translocation of individuals between different game parks to maintain genetic diversity and to carry out captive breeding projects in zoos to increase the population size; however, little is known about the population genomics, such as population structure, selection and susceptibility to disease. Numerous cases of Mycobacterium bovis (M. bovis) infection and mortalities associated with bovine tuberculosis (BTB) have been reported in wild dogs within the Kruger National Park (KNP). To investigate genetic diversity, including susceptibility to M. bovis infection, whole-genome sequencing (WGS) of M bovis infected and uninfected wild dogs is required. The sequence data will be used to firstly estimate the relatedness between individuals, and as part of a larger study, to identify loci under selection and to conduct an association study, by identifying variants with large differences in allele frequencies between infected and uninfected wild dogs. The information obtained from WGS will provide new tools for conservation managers for conservation strategies and can be used in the future to develop diagnostic tools to investigate disease susceptibility. Changes in genetic composition due to environmental changes, fragmented populations and disease outbreaks can leave lasting community-wide effects on host populations, highlighting the significance of the role of host genetic diversity in species conservation.
Aim: The aim of this study is to investigate the genomic diversity of a population of free-ranging wild dogs in South Africa which will advance our understanding of adaption as well as susceptibility or resistance to diseases and other threats in this species.
Methodology: Wild dog samples from KNP will be subjected to whole-genome sequencing. The sequence data will be used to conduct a series of statistical analyses to determine the genetic diversity, population structure, relatedness and identify signatures of selection. These data will then be used to conduct an association analysis to identify genetic variants associated with M. bovis infection in exposed wild dogs in a longer-term study.
Expected Results: Whole-genome sequencing will allow us to determine genomic diversity in the current wild dog population of KNP as well as provide a foundation for metapopulation management of this species in South Africa and in zoos across the world. Determining the degree of genetic diversity in isolated African wild dog populations will promote our understanding of adaptive selection to environmental changes and introduction of novel pathogens in populations. Identifying specific regions of the genome that are associated with diversity and adaptation could lead to the development of novel tools which will improve current conservation decisions and inform genetic management of the population of wild dogs in game reserves and zoos. As part of a larger study, we will identify specific genetic regions in African wild dogs associated with adaption as well as susceptibility/resistance to M. bovis infection.

Evolution of phenotypes: understanding diversity and the role of plasticity in adaptation to new environments
SMBE-PO-120
Canalization: An alternative form of homeostasis?
A. A. Sato*

Abstract: Organismal development is robust in the face of genetic variation and environmental fluctuations. Since Conrad Hal Waddington famously coined the term 'canalization' to distinguish homeostasis in developmental from cellular, physiological homeostasis over half a century ago, developmental robustness has been a central interest of developmental biologists as well as evolutionary biologists. Since the breakthrough of Rutherford and Lindquist (1998) proposing a role of Hsp90 in developmental buffering, chaperones have gained much attention in the study of canalization. However, recent studies have revealed that a number of other molecules are also potentially involved in canalization. My group has been working on identifying molecular basis involved in developmental robustness under thermal stress in the sea squirt Ciona. Sea squirts are a simple model of chordates (which includes vertebrates), yet their embryonic stages last only for a day therefore an excellent model to study developmental robustness in chordates. We previously showed that level of robustness is maternally inherited and DnaJ chaperones plays a role in developmental robustness (Sato et al. 2015 Sci Rep 5: 16717), which was also tested in the worm *C. elegans* (Hughes et al. 2019 J. Exp. Zool). By comparing transcriptomes of hybrids of *Ciona* species, which have the same genotypes but are different in the level of robustness, we recently found that genes involved in developmental robustness are involved in those closely linked to cellular homeostasis, such as translation, protein quality control and metabolism. We conclude that developmental robustness is based on cellular homeostasis, not being able to distinguish from cellular homeostasis.

Evolution of phenotypes: understanding diversity and the role of plasticity in adaptation to new environments
SMBE-PO-085

Behavior-related gene polymorphisms in invasive African sacred ibis populations in Taiwan

Y.-C. Ku [1], Y.-R. Wang [1], C. S. Ng [1,*]
[1]National Tsing Hua University, Hsinchu, Taiwan

Abstract: It has been suggested that individual behavioral traits affect the potential to successfully encounter novel environments, such as invasion. Identifying the genetic basis of behavioral variation in invasive species thus serves an important step towards understanding the evolutionary potential of animal invaders. Using African sacred ibis *Threskiornis aethiopicus* sampled from invasive Taiwanese populations, we investigated sequence data of the dopamine receptor D4 gene (*DRD4*) and the serotonin transporter gene (*SERT*), loci suspected to be under selection for novelty-seeking behavior in a range of taxa including mammals and birds. We hypothesized that such behavior may be adaptive during invasion. The sequence data obtained in this study will allow evaluation of the involvement of *DRD4* and/or *SERT* in the invasion of African sacred ibis populations in Taiwan through genetic differentiation and selection signature. In addition to analyses to detect the presence of selection, we also estimated population differentiation and gene diversity using neutral marker data in the same individuals. The identified genetic variants also have the potential to understand the evolution of behavior in these invasive bird populations.

Evolution of phenotypes: understanding diversity and the role of plasticity in adaptation to new environments
SMBE-PO-086

Found in (mis)translation: The evolutionary benefits of translation errors

D. Agashe [1,*], L. Samhita [1], P. Raval [1], S. Tamhankar [1 2]
[1]National Centre For Biological Sciences, Bangalore, [2]IISER, Pune, India

Abstract: Mistranslation is typically deleterious, and cells have evolved various mechanisms to minimise the translation error rate. However, bacterial strains with increased mistranslation also seem to have a fitness benefit under specific environments. What mechanisms underlie this advantage, and what are the consequences of mistranslation for adaptive evolution? To understand this, we induced mistranslation in *E. coli* via genetic as well as chemical methods and exposed bacteria to DNA damaging agents. We found that mistranslation rapidly triggered the bacterial SOS response, allowing fast repair of damaged DNA and thus increasing survival. This novel link between mistranslation and the SOS pathway suggests convergent cellular responses to disparate physiological stresses. Interestingly, under longer-term exposure to DNA damaging antibiotics, mistranslating strains also fixed distinct beneficial mutations compared to wild-type *E. coli*. Thus, our work elucidates the molecular mechanism underlying the fitness benefits of mistranslation, and shows that mistranslation may significantly alter the evolutionary trajectory of populations.

Evolution of phenotypes: understanding diversity and the role of plasticity in adaptation to new environments
SMBE-PO-118

Evolution of chromatin accessibility associated with traits of cichlid phenotypic diversity

T. Mehta [1,*], A. Man [1], D. Joyce [2], W. Salzburger [3], W. Haerty [1], F. Di-Palma [1]
[1]Earlham Institute, Norwich, [2]Biological Science, University of Hull, Hull, United Kingdom, [3]Zoological Institute, University of Basel, Basel, Switzerland

Abstract: In vertebrates, the East African cichlid radiations represent arguably the most dramatic examples of adaptive speciation. In the great lakes Victoria, Malawi and Tanganyika and within the last few million years, one or a few ancestral lineages of haplochromine cichlid fish have given rise to over 1500 species exhibiting an unprecedented diversity of morphological and ecological adaptations. Explaining the wide range of phenotypic diversity that has arisen in adaptive radiations such as the haplochromine cichlid fish lineages of East Africa, has been an ongoing challenge. We recently reported striking cases of gene regulatory network (GRN) rewiring for adaptive trait genes, and confirmed that discrete transcription factor binding sites (TFBS) mutations disrupt regulatory edges across species and segregate according to lake species phylogeny and ecology. To further analyse gene regulatory network evolution and gene regulatory activity associated with traits of cichlid phenotypic diversity, we developed and optimised the assay for transposase-accessible chromatin using sequencing (ATAC-seq) from cichlid fish tissues. This allowed us to characterise and compare tissue-specific activity of regulatory regions, including the integration of TFBS and GRN predictions in representative species of the great lake radiations.

Evolution of phenotypes: understanding diversity and the role of plasticity in adaptation to new environments
SMBE-PO-123

Rapid evolution of color vision during the land-sea transition of elapid snakes

B. F. Simoes [1 2,*], N. S. Hart, J. C. Partridge, D. M. Hunt, B. S. Chang, D. Pisani, D. J. Gower, K. L. Sanders
[1]School of Earth Science, University of Bristol, Bristol, United Kingdom, [2]School of Biological Sciences, The University of Adelaide, Adelaide, Australia

Abstract: Sea snakes are among the most ecologically specialised of all squamate reptiles. This group diverged from terrestrial Australian elapid snakes approximately 6-16 MY and radiated into at least 60 species found from open ocean waters to coastal and mangrove habitats. Sea snakes are highly adapted to their marine lifestyle, being viviparous, having paddle-shaped tails, sealed nostrils, and respiratory traits that allow them to remain active underwater for several hours. However, how their visual system adapted during their ecological transition to a marine environment is still largely unknown. The visual pigments of sea snakes emerged to be highly adaptable when compared to those of all other terrestrial snakes. In particular, species in the *Hydrophis*clade have undergone numerous shifts in colour sensitivity, with blue-shifts caused by mutations in the OPN1SW opsin gene, and green-shifts caused by mutations in the OPN1LW opsin gene. We found that several *Hydrophis* sea snakes are polymorphic for visible-spectrum and ultra-violet encoded by alleles of the OPN1SW opsin, suggesting that they might have expanded spectral sensitivity, or that similarly to some primates they might have allelic trichromacy. Our results suggest that sea snakes are a remarkable case of visual evolution in vertebrates.

Evolution of phenotypes: understanding diversity and the role of plasticity in adaptation to new environments
SMBE-PO-117

The genomic and phenotypic evolution of Ralstonia solanacearum plant-pathogenic bacterium on a global scale.

E. M. Farnham*

Abstract: *Ralstonia solanacearum* is a pathogenic bacterium that can infect over 200 different plant species, including many important crop plants. *R. solanacearum* can occupy multiple different habitats including the soil and water and can easily transmit from these environmental reservoirs to agricultural environments driving disease outbreaks. This wide host range and environmental versatility means *R. solanacearum* has a global distribution covering most of the world making it extremely difficult to control.

In collaboration with FERA Science Ltd we are studying the worldwide diversity of *R .solanacearum* from a collection of 384 strains spanning 66 countries and a wide variety of hosts sampled from 1945 to 2018. This includes 170 strains from the UK with fine scale geographic and temporal sampling during and between infections. We will first present primary data on the variation of key life-history traits based on high throughput phenotyping. This will give us an insight in how certain phenotypes are distributed geographically and temporally and how the pathogen has adapted in time focussing on changes associated with virulence and survival in environmental reservoirs.

This data will then be combined with whole genome sequencing data to identify genes associated with virulence in a genome-wide association study (GWAS). We also aim to link genetic patterns with certain environmental backgrounds to better understand pathogen adaption across natural and agricultural environments. Investigating the distribution of genotypes and phenotypes in space and time will provide an insight into bacterial adaptation and increase our understanding of *R. solanacearum* transmission dynamics.

Evolution of phenotypes: understanding diversity and the role of plasticity in adaptation to new environments
SMBE-PO-106

Genomes of kiang and domestic donkey provides insight into their evolutionary history and adaptation

L. zeng [1,*]
[1]Kunming Institute of Zoology, Chinese Academy of Sciences, Kunming, Yunnan, China

Abstract: Kiang (Equus kiang) is famous in biodiversity conservation, and is absorbing for its origin, evolution and potential admixture with domestic donkey. Here, we firstly assemble a de novo genome of kiang, and re-sequenced genomes of another 5 kiangs. Genetic innovations including structural variation and rapidly evolved genes were recapitulated in kiang. Particularly, the rapid evolution of immune genes might be related to the evolution of immune system, and hard selective sweep of EPAS1 gene is likely coupled to high altitude adaptation. However, we didn't find frequent occurrence of hard selective sweep in the kiang. To investigate relationship between kiang and domestic donkey, we further re-sequenced the whole genomes from 93 domestic donkeys across the world. In contrast to previous viewpoint, rare genetic introgression was found between kiang and domestic donkey, indicating minor genetic contribution of kiang to domestic donkey. We found common occurrence of soft selective sweep and revealed rapid evolution of HIF-pathway (e.g. ELGN1) for high altitude adaptation of Tibetan domestic donkey. The results substantiate rare hard selective sweep, but common soft selective sweep for high altitude adaptation, and the data resource will provide insight into the scenarios of relationship between kiang and domestic donkey, and their evolution.

Evolution of phenotypes: understanding diversity and the role of plasticity in adaptation to new environments
SMBE-PO-112

New polygenic associations explain skin pigmentation diversity in Papua New Guinea

A. Morez [1,*], N. Brucato [1 2], K. Suseki [3], R. Tsang [3], J. Kariwiga [3], T. Beni [3], L. Saag [4], J. Muke [5], K. Miamba [6], A. Kuaso [6], T. Kivisild [4 7], M. P. Cox [8], M. Metspalu [4], W. Pomat [9], M. Leavesley [3 10], F.-X. Ricaut [1 2]

[1]CNRS UMR 5288, Laboratoire d´Anthropobiologie Moléculaire et Imagerie de Synthèse (AMIS), [2]CNRS UMR 5174, Laboratoire Evolution & Diversité Biologique (EDB), Université de Toulouse (Paul Sabatier), Toulouse, France, [3]Archaeology, School of Humanities and Social Sciences, University of Papua New Guinea, Port Moresby, Papua New Guinea, [4]Institute of Genomics, University of Tartu, Tartu, Estonia, [5]Social Research Institute, [6]National Museum and Art Gallery, Port Moresby, Papua New Guinea, [7]Laboratory for Human Evolutionary Genetics, Department of Human Genetics, Faculty of Medicine, University Catholic of Leuven, Leuven, Belgium, [8]Statistics and Bioinformatics Group, School of Fundamental Sciences, Massey University, Palmerston North, New Zealand, [9]Papua New Guinea Institute of Medical Research, Goroka, Papua New Guinea, [10]College of Arts, Society and Education, James Cook University, Cairns, Australia

Abstract: Papuan populations have an exceptional biological diversity. For decades, anthropological studies have established the large spectrum of phenotypes present in Papua New Guinea such as stature, with individuals showing 'pygmy' features, or hair color, with blond-haired individuals. This diversity is a result of a long-standing geographical isolation, since their first arrival 50,000YBP, a strong cultural structure, with the highest linguistic diversity of the world, and unique admixture patterns, notably with an archaic hominin, Denisova. Particularly, high skin pigmentation diversity has long been documented in Melanesia but the genetic pattern involved and the underlying evolutionary processes are so far poorly understood.

Combining new whole genome sequences to skin color (melanin index), we conducted genome-wide association analyses on the skin color diversity in Papua New Guinea. Four regions, never associated to pigmentation, were identified (PRDM2, SOCS5, THSB2 and GPR19). PRDM2 and SOCS5 have previously been found to be highly expressed in skin. None of the major genes previously linked to human skin pigmentation are implicated in the observed color diversity. Neither significant natural selection index (iHS, nSL, XP-EHH, Tajima D) nor skin color correlation to environmental factors (i.e. solar radiation, precipitation…) can explain the four newly associated regions. Important cultural or demographic processes have more probably shaped the skin color in Papuans, as previously suggested. Our results are in line with the recent findings that human skin pigmentation is a polygenic trait and its diversity implied multiple and complex convergent evolution.

Evolution of phenotypes: understanding diversity and the role of plasticity in adaptation to new environments
SMBE-PO-107

Nucleotide divergence pattern between L and M opsin genes shows human uniqueness among catarrhine primates possibly correlated with color vision variation in humans

Y. Matsushita [1,*], A. D. Melin [2], S. Kawamura [1]
[1]Integrated Biosciences, Graduate School of Frontier Sciences, The University of Tokyo, Kashiwa, Japan, [2]Anthropology and Archaeology & Medical Genetics, University of Calgary, Calgary, Canada

Abstract: Trichromatic color vision in catarrhine primates (Old World monkeys, apes, and humans) is achieved by juxtaposition of spectrally differentiated L and M opsin genes on the X chromosome and autosomal S opsin gene. The L and M opsin genes are susceptible to gene conversion each other and homogenization between them. Natural selection, on the other hand, acts against the homogenization and maintains the spectral difference between L and M opsins. Combined effect of the two opposing forces is expected to create a pattern of L-M nucleotide divergence low in introns and peripheral exons (exons 1 and 6) irrelevant to L-M spectral difference and high in central exons, especially exons 3 and 5, relevant to the spectral difference. Our previous study showed that the gibbon L and M opsin genes followed this pattern, where the L-M nucleotide difference was highest in exons 3 and 5, with the height ratio, (exon 3)/(exon 5) (designated ex3/ex5), roughly 1. Spectral effect from exon 3 is known to be smaller than that from exon 5 in primate L and M opsin genes. Peculiarly, ex3/ex5 ratio is smaller in the L and M opsin genes retrieved from the human reference genome (HRG), possibly reflecting relaxation of the selective constraint in humans to maintain the L-M spectral difference. However, generalization of the gibbon result to other non-human catarrhines and of HRG to diverse humans remains unexamined. In this study, we examined one individual each from five species of non-human hominoids and nine species of Old World monkeys, as well as ca. 400 humans from diverse ethnicity and subsistence by employing target capture and massive parallel sequencing methods to the L and M opsin genes. While the ex3/ex5 ratio was overall close to 1 in non-human catarrhines, that of humans was overall smaller than 1 and variable among individuals irrespective of ethnicity and subsistence. Importantly, the lower ex3/ex5 ratio than 1 is also the norm in individuals with "normal" trichromacy. These results suggest that selective constraint to maintain L-M divergence and normal trichromacy already relaxed in the hunter-gatherer common ancestor of all modern humans in Africa.

Evolution of proteins and molecular pathways/networks under higher level selective effects
SMBE-PO-130

Evolution and functional diversity of short-chain alcohol dehydrogenase and related proteins in animals

A. Julian-Sanchez [1], O. A. Santillan-Lopez [1], H. Riveros-Rosas [1,*]
[1]Bioquimica, Fac. Medicina, Universidad Nacional Autonoma de Mexico, Mexico City, Cd. Mx., Mexico

Abstract: In animals, three non-homologous NAD(P)$^+$-dependent ADH protein families are reported. These arose independently throughout evolution and possess different structures and mechanisms of reaction: type I (medium-chain) ADHs are zinc-containing enzymes and comprise the most studied group in vertebrates; type II (short-chain) ADHs lack metal cofactor and have been extensively studied in Drosophila; and type III ADHs are iron-dependent enzymes initially identified in microorganisms. The presence of ADHs in animals has been assumed to be a consequence of chronic exposure to ethanol. By far the most common natural source of ethanol is fermentation of fruit sugars by yeast, and available data support that fruit traits evolved in concert with the characteristics of their frugivorous seed dispersers. Therefore, if the presence of ADHs in animals evolved as an adaptive response to dietary ethanol exposure, then it can be expected that the enzymogenesis of these enzymes began after the appearance of angiosperms with fleshy fruits, because substrate availability must precede enzyme selection. However, we demonstrated that origin of type I and type III ADHs predates the origin of angiosperms, and evolved independently of ethanol availability (Hernández-Tobias et al., Chem-Biol. Interact. 2011, 191: 14; Gaona-Lopez et al., PLoS ONE 2016, 11: e0166851). In contrast, induction of type II *Adh* by ethanol exposure, a positive correlation between ADH activity and ethanol resistance, ethanol utilization as a nutrient, and the fact that flies and type II *Adh* diversification occurred in concert with angiosperm diversification, strongly suggest that type II ADHs were recruited to allow Drosophila larval flies to exploit new restricted niches with high ethanol content. However, type II *Adh* locus has undergone several duplication events, and can be found in other invertebrate phyla not exposed to natural ethanol sources. This suggest that type II ADHs performs besides ethanol oxidation other metabolic roles. In fact, the identification of orthologous proteins outside *Drosophila* genus is very difficult because a complex pattern of duplications occurred. Since additional complete genomes and sequences are now available, we performed a new phylogenetic analysis of type II (short chain) ADH in animals, in parallel with a gene synteny analysis used as an additional criterion to identify true orthologous genes. Results showed that 15-hydroxyProstaglandin dehydrogenase (PGDH) exhibits the closest homology to type II ADHs, and can be found in all animal phyla. In contrast, type II ADHs proteins are restricted to insects of the order Diptera (flies). Thus, it can be assumed that the origin of PGDH predates the origin of Diptera type II ADHs; and therefore, are derived by duplication from PGDHs. *Adh* locus in *Drosophila* has undergone at least five duplication events, and all them occurred only within the Diptera. However, bona fide type II ADHs (participating in ethanol metabolism) are encoded by an *Adh* gene nested in the *outspread* (*osp*) gene, i.e., true *Adh* genes are completely included within the second intron of the *osp* gene, and this occurred only in species that belong to *Drosophila* genus. Therefore, type II ADHs found in no-*Drosophila* species are involved in metabolic roles not related to ethanol metabolism, and are true orthologs of other type II ADH proteins, like Fat Body Protein 2, and retinol dehydrogenases.
Supported by UNAM-DGAPA-PAPIIT grants IN225016 & IN218819

Evolution of proteins and molecular pathways/networks under higher level selective effects
SMBE-PO-157
Recapitulating virus-host coevolution through in vitro reconstitution of antibody:antigen interactions
C. W. Pugh [1 2,*], A. G. Schmidt [1 2]
[1]Department of Microbiology, Harvard Medical School, Cambridge, MA, [2]Ragon Institute of MGH, MIT and Harvard, Massachusetts General hospital, Boston, MA, United States

Abstract: The dynamics of host-pathogen interactions give a window into the evolutionary arms race between host and pathogen. Understanding these interactions on a structural level can guide therapeutic and vaccine development, while simultaneously informing us about viral (and protein?) evolution. A current goal for a "universal" influenza vaccine is to elicit broadly neutralizing antibodies (bnAbs) that target conserved epitopes. It is hypothesized that by targeting these sites, the virus cannot readily escape without compromising viral fitness. To test this, we developed a novel *in vitro* reconstitution assay that allows for continuous directed evolution of a germline precursor of a broadly neutralizing antibody in the presence of a replicating virus. We created a mutagenized antibody library based on a bnAb germline precursor (Ab_0) to mimic the *in vivo* affinity maturation process. Panning this library against replicating influenza virus identified escape-mutant bearing hemagglutinin (mtHA'). This mtHA' was used to select for an evolved antibody (eAb_1) that now recognizes the escaped virus. Repeating this selection and counter-selection recapitulates the evolutionary landscape of the adaptive host response and subsequent viral adaptability, encompassing both divergent and convergent adaptive pathways. We continue to structurally characterize the accumulated viral escape mutations and the resulting counter-resistant antibody mutations; doing so creates a "molecular movie" of this dynamic "arms race". This assay describes the overall evolutionary plasticity of a viral protein and finiteness of evolvability at a conserved epitope. These results have broad implications for understanding evolution of rapidly evolving pathogens under immune selection.

Evolution of proteins and molecular pathways/networks under higher level selective effects
SMBE-PO-132
Unique silence mechanisms used in marsupial X chromosome inactivation
S. A. Waters, K. L. McIntyre , P. D. Waters*

Abstract: Marsupial and eutherian X chromosome inactivation (XCI) share a common origin. This explains the overlapping histone code present on the inactive X in each group, but it fails to explain other fundamental molecular dissimilarities, such as the startling difference of X chromosome DNA methylation patterns. In eutherians, the long noncoding RNA (lncRNA) *XIST*associates with epigenetic machinery and ribonuclear proteins to direct XCI. In marsupials, an independently evolved lncRNA – called *Rsx* – is responsible for mediating silencing, but little is known about how itfunctions. Here we demonstrate that *Rsx*interacts with proteins responsible for epigenetic silencing, which it probably recruits to the X chromosome to be silenced. Marsupial XCI is often considered the primitive cousin of eutherian XCI; however, our data highlight layers of independently evolved RNA mediated XCI complexities that are far from primitive.

Evolution of proteins and molecular pathways/networks under higher level selective effects
SMBE-PO-133
Identifying promoter DNA regulatory motifs associated with drought related genes
J. Li*, C. Casola

Abstract: The synergy between transcriptional and functional regulation assists plants in their growth, reproduction, and responses to abiotic and biotic stresses. The promoter regions contain regulatory motifs, regulatory motifs tend to evolve relatively quickly, whereas there is some conservation of these sequences across distantly related species. However, little is known regarding the regulation of gene expression in gymnosperms, including the composition of gymnosperm promoter regions and the conservation of regulatory motifs between gymnosperms and angiosperms. In this study, we identified regulatory motifs using transcriptomic data from a drought-simulation experiment in loblolly pine (*Pinus taeda* L.). We retrieved the putative promoter region of thousands differentially expressed genes between roots of control and "drought" and obtained candidate "proximal" (-1,000,+100bp) and "distal" (-2000,+100bp) promoter regions surrounding the putative transcription start site (TSS) of these genes. We found 4 and 7 regulatory motifs with 0-order among the differentially expressed transcripts in the 1100bp and 2100bp promoter regions by MEME program. We discovered the most significant motif in the root clone2 up-regulated transcripts 1100bp and 2100bp promoter regions containing the ABRE (abscisic acid responsive element) motif ACGTG[G/T]C (Narusaka et al. 2003). This is a strong sequence conservation of ABRE across seed-plants. Besides, there are two sequences CGGCCC and ACGGCC found in other studies to be regulatory motifs in plants with unknown functions (Yamamoto et al. 2007 and Korkuc et al. 2014). Conservation of these motifs with other conifers will be performed using syntenic regions of the sequenced genomes of sugar pine, Douglas fir and Norway spruce.

Evolution of proteins and molecular pathways/networks under higher level selective effects
SMBE-PO-159

Did the targeting signals for mitochondria and plastids evolve from ancient antimicrobial peptides?

M. Burdukiewicz [1], K. Sidorczuk [2], S. Rödiger [3], P. Gagat [2,*]
[1]Faculty of Mathematics and Information Science, Technical University of Warsow, Warsaw, [2]Department of Bioinformatics and Genomics, University of Wroclaw, Wroclaw, Poland, [3]Faculty of Natural Sciences, Brandenburg University of Technology Cottbus–Senftenberg , Senftenberg , Germany

Abstract: Antimicrobial peptides (AMPs) are ancient and evolutionarily conserved molecules widespread in all living organisms that participate in host defence and/or microbial competition. They are short, 12 to 50 amino-acid long, do not display any consensus sequences but do share some common features, such as positive charge, hydrophobicity and amphipathicity. These structural characteristics enable AMPs to preferentially disrupt negatively-charged bacterial membranes and do not adversely affect the eukaryotic cells. AMPs are also hypothesized to have greatly contributed to the establishment of bacteria-derived organelles, i.e. mitochondria and plastids, by facilitating two key processes of endosymbiont-to-organelle transformation: (i) endosymbiont gene transfer due to bacterial cell lysis and (ii) evolution of efficient protein import machinery by becoming N-terminal targeting signals for nuclear-encoded, mitochondria/plastid-targeted proteins. Indeed mitochondria and plastid transit peptides (mTP and pTP) seem to share some characteristics with AMPs and some do exhibit antibacterial activities. In order to test the hypothesis, we developed AmpGram, a new tool based on N-grams, reduced alphabets and random forest methods that significantly outperforms all the available AMP prediction algorithms. We tested the hypothesis on mTPs and pTPs extracted from the experimentally verified nuclear encoded, mitochondria/plastid targeted proteins downloaded from the UniProt database. As a control, we used annotated signal peptides, i.e. presequences responsible for protein targeting to the endomembrane system. AmpGram did recognize as AMPs a significant number of mTPs and pTPs but not signal peptides. Based on our results, we propose a new model for evolution of protein import into bacteria-derived organelles with AMPs playing a central role.

Evolution of proteins and molecular pathways/networks under higher level selective effects
SMBE-PO-161

Proteins functional annotation based on their alignment and genes expression

R. Galimullina [1,*], R. Deviatiiarov [1], O. Gusev [2], E. Shagimardanova [1]
[1]Kazan Federal University, Kazan, Russian Federation, [2]KFU-RIKEN "Translational genomics" Unit, RIKEN, Yokohama, Japan

Abstract: Anhydrobiosis is a condition when due to water loss all biological processes of organism slowed to complete stop. This is a perspective topic for protein, tissue, cell storage research because dry storage technologies have advantages in material safety and energy costs. Nowadays the larvae of an African midge *Polypedilum vanderplanki is* the largest and phylogenetically closest organism to mammals, which can survive complete desiccation in natural conditions.Acknowledgements: This work was supported by Russian Science Foundation, grant №17-44-07002
Comparative genomic analysis has shown that the genome of *Polypedilum vanderplanki* contains specific clusters of paralogous genes, while the phylogenetically closest species *P.nubifer* doesn't have and can't survive similar extreme conditions. Also, these new genes were shown to be active in response to different stress-factors and dehydration. Despite the huge amount of transcriptomic data the mechanisms and role of this multi-paralogs still unknown.
In the current research, we created a unique algorithm based on a comparison of protein sequences alignment and activity of its genes for functional domains prediction. Such data provides additional information about the roles of paralogous genes regarding adaptations to stresses or tissue-specific structures. Thus, we found that C-end of Lil and TRX proteins correlates with gene expression in experiments with desiccation, while there is no correlation with tissue-specificity.
Moreover, this method was successfully applied to finding main domains in PAX transcription factors family for *Mus musculus*, which makes this approach principally new and universal for protein functional annotation.

Evolution of proteins and molecular pathways/networks under higher level selective effects
SMBE-PO-127

Genetic architecture and sex-specific selection govern modular, male-biased evolution of doublesex

S. Baral [1,*], G. Arumugam [1], R. Deshmukh [1], K. Kunte [1]
[1]Biodiversity Lab, National Center for Biological Sciences, TIFR, Bangalore, India

Abstract: *doublesex* is a transcription factor that regulates early embryonic sex differentiation in all holometabolous insects, along with the development of species-, sex- and morph-specific adaptations during pupal stages. It splices into male- and female-specific isoforms in response to sex-determination cues, and triggers a sex-specific developmental cascade. How does a highly conserved gene with a critical developmental role also remain functionally dynamic enough to gain ecologically important adaptations that are divergent in sister species? We analyzed patterns of exon-level molecular evolution and protein structural homology of *doublesex* from 145 species of four insect orders representing 350 million years of divergence. This analysis revealed that evolution of *doublesex* was governed by a modular architecture: functional domains and many female-specific regions were highly conserved, whereas male-specific sequences and protein structures evolved up to a thousand-fold faster, with sites under pervasive and/or episodic positive selection. This pattern of sex bias was reversed in Hymenoptera. The modular architecture holds true at multiple levels of taxonomic organization ranging from few species in Chalcid wasps to more than 60 species in the Lepidoptera, although their exon usage is highly divergent and order-specific. These findings suggest a widespread role of *dsx* beyond currently known functions in producing sexual dimorphisms and polymorphisms of ecological significance. This is irrespective of diverse mechanisms of sex determination, heterogamety and exon usage that are evident across insect orders. Thus, highly conserved yet dynamic master regulators such as *doublesex* may partition specific conserved and novel functions in different genic modules at deep evolutionary timescales.

Evolution of proteins and molecular pathways/networks under higher level selective effects
SMBE-PO-129
How many tRNA genes do bacteria need?
P. K Raval [1,*], D. Agashe [1]
[1]National Centre for Biological Sciences, TIFR, Bangalore , India

Abstract: Translation is a critical process in all cells. In bacteria, translation rate is strongly correlated with maximum growth rate, and should be constrained both by available nutrients and components of the translation machinery. Hence, translating faster should be selectively favoured when nutrients are plenty. Specifically, we predicted that available tRNAs (that deliver amino acids during translation) should limit translation rate, and altering tRNA pools should therefore have the largest fitness impact in nutrient-rich conditions. Indeed, we find that in E. coli, multi-copy tRNA genes are beneficial in nutrient rich media where fast growth is under selection, but impose a significant cost in nutrient poor media. In contrast, deleting single-copy tRNAs is costly only if tRNA modifying enzymes (MEs, which expand wobble pairing) are also deleted. Thus, MEs buffer against the loss of single-copy tRNA genes; though it remains unclear why many bacteria have both MEs and their target tRNAs. Overall, our results indicate that E. coli tRNA pools have evolved under selection for fast growth, presumably because E. coli frequently encounters nutrient-rich niches. Our work constitutes the first experimental demonstration of the strong imprint of growth-mediated selection on bacterial tRNA content. We further suggest that the large diversity of tRNA pools observed across bacterial genomes stems from selection for rapid vs. slow growth in the diverse ecological niches that lineages occupied in their evolutionary history.

Evolution of proteins and molecular pathways/networks under higher level selective effects
SMBE-PO-131

The evolution of termite digestion system

S. He [1,*], J. J. G. Plaza [1], A. Chakraborty [1], J. Šobotnik [1]
[1]Faculty of Forestry and Wood Sciences, Czech University of Life Sciences Prague, Prague, Czech Republic

Abstract: Termites are known as decomposers to break down the dead organic matter in tropical and subtropical forests. Evolved from wood-feeding cockroaches, termites, especially the abundant higher termite species, have diverse feeding materials from wood, grass to soil. In order to acquire sufficient nutrition, their digestion systems adapt to different food resources, which may effect the evolution of termites. It has been widely acknowledged that digestion ability of termites mainly depend on their symbionts, though termites also express digestion-related proteins. But how the termite digestion systems evolved from wood feeding cockroaches is unknown. We hypothesis the non-wood feeding termites have high investment in digestion and their symbionts shift to fit feeding materials. In our study, we will use genomics to analysis the evolution of digestion-related genes in termites and their symbionts. Termites with different feeding materials will be compared: two soil-feeders, *Anoplotermes banksi* and *Cubitermes intercalatus,* two wood-feeders, *Nasutitermes octopilis* and *Microcerotermes biroi,* a grass-feeder*, Trinervitermes disparatus, and* a bacteria-farming termite species, *phaerotermes sphaerothorax*. After genome and metagenome sequencing, we will predict the digestion related gene families among termites and symbionts. Subsequently, evolution rates and gene gain/loss of these gene families will be analyzed across different termite species and their corresponding symbionts. In addition, the transcriptome and metatranscriptome sequencing will be used to determine the expression of identified digestion-related genes in order to identify the main functional genes and symbionts. These analysis could reveal the difference of digestion systems in hosts and symbionts from termites with different feeding materials. This study could help us to understand the effect of feeding material in the evolution of termite digestion system and provide further evidence in the study of wood digestion and termite evolution.

p { margin-bottom: 0.1in; direction: ltr; color: rgb(0, 0, 0); line-height: 115%; text-align: left; }p.western { font-family: "Liberation Serif", serif; font-size: 12pt; }p.cjk { font-family: "Noto Sans CJK SC Regular"; font-size: 12pt; }p.ctl { font-family: "Lohit Devanagari"; font-size: 12pt; }

Evolution of proteins and molecular pathways/networks under higher level selective effects
SMBE-PO-135
Duplication history of the GPCR signalling pathway reveals the ancestry of its disease genes
A. Barradas [1,*], D. L. Robertson [1 2]
[1]Division of Evolution and Genomic Sciences, University of Manchester, Manchester, [2]MRC-University of Glasgow Centre for Virus Research, University of Glasgow, Glasgow, United Kingdom

Abstract: Human disease genes are enriched in duplicate copies originating from small-scale duplication (SSD) and whole-genome duplication (WGD) events. Two rounds of WGD in early vertebrate evolution have especially affected signalling systems, however, it is unclear how specific duplication events impacted individual signalling pathways and to what extent they have contributed to the development of disease. Here, we compiled high-quality manually curated gene sets corresponding to the topological layers of the G protein-coupled receptor (GPCR) signalling pathway and investigated their duplication history since their common opisthokont ancestor. We find that receptors and downstream signalling proteins have experienced duplication events more often than upstream activating ligands and also that all signalling components have distinct evolutionary origins of duplication. We also find important differences between duplication patterns and the expansion of disease in the network topology. In particular, signalling components where WGD-retained genes (ohnologues) have subsequently undergone SSD tend to have an evolutionary association with Mendelian disease, suggesting that genes involved in mediating the intracellular transduction cascade are dosage-sensitive. In contrast, GPCR and ligand disease genes tend to cause more genetically complex disease. While ohnologues are overrepresented in GPCR disease genes, ligand disease genes do not show any duplicate retention bias, indicating that their tendency to underlie complex phenotypes is a consequence of greater pleiotropic effects resulting from their upstream locations. Our findings provide new insights into the role of duplication in the evolution of signalling pathways and further our understanding of molecular mechanisms giving rise to disease in a system-wide context.

Evolution of proteins and molecular pathways/networks under higher level selective effects

SMBE-PO-136

Functional diversification of the receptor ligand interaction controlling self incompatibility in Arabidopsis: an in planta resurrection approach.

M. Chantreau, C. Poux, M. Lensink, G. Brysbaert, X. Vekemans, V. Castric*

Abstract: Protein-protein interactions regulate a variety of fundamental cellular processes, yet the rate and mechanisms by which interacting protein partners accumulate evolutionary novelties and functionally diversify remain poorly understood. We performed *in planta* resurrection of an ancestral protein to decipher the diversification process of the receptor-ligand interaction controlling the highly diverse self-incompatibility system in the plant *Arabidopsis halleri*. We demonstrate that two allelic variants currently segregating as functionally distinct receptor-ligand combinations result from the long-term maintenance of the ancestral recognition specificity for one of the two descendent lineages, and at the same time functional divergence of the other allelic lineage through acquisition of a new recognition specificity. This asymmetrical evolution suggests a diversification process in which strong balancing selection is maintaining a set of allelic specificities essentially unchanged over the long term, and the rare emergence of novel allelic specificities, possibly through non-interacting evolutionary intermediates. Computational prediction of the structure of the receptor-ligand complex shows that molecular docking of the ancestral receptor resembles more closely that of the derived receptor that has retained identical binding specificity than to that of the derived receptor, providing insights into the structural determinants of the shift in binding specificity and suggesting allosteric changes as an important ultimate source of evolutionary novelty in this system.

Evolution of proteins and molecular pathways/networks under higher level selective effects

SMBE-PO-137

SERK Receptor Kinases phylogenetics reveals five main clades, each harbouring mixed function.

S. Hosseini [1,*], E. Schmidt [2], F. T. Bakker [3]

[1]Plant protection, University of Rafsanjan, Rafsanjan, Iran, Islamic Republic Of, [2]UMCG Groningen, Groningen, [3]Biosystematics Group, Wageningen University, Wageningen, Netherlands

Abstract: Receptor Like Kinases represent the largest group of cell surface receptors in plants, of which monophyletic LRR-RLK subfamily II is considered to contain the somatic embryogenesis receptor Kinases known to be involved in both developmental processes as well as cellular immunity in plants. We used almost all proteins annotated as SERK proteins in GenBank in order to estimate phylogeny within the LRRII-RLK clade, using nematode serine/threonine-protein kinase Pelle as outgroup. We reconstruct five main clades with the pattern of land plant relationships re-occurring in each, confirming previous hypotheses that several duplication events happened in the HDR gene family prior to the divergence among land plant lineages.

According to our results, different annotated LRR-RLKs names as occurring in the literature and in data bases are not congruent and cannot be considered synonyms. We offer a new classification and nomenclature of LRR-RLKs based on the five main clades found, naming them 'Hetero-dimerising receptor' (HDR) I – V. We also present the motif structure of each HDR type based on 1328 available sequences, showing that structures and intron-exon boundaries of all types are well conserved in evolution.

Evolution of proteins and molecular pathways/networks under higher level selective effects

SMBE-PO-134

Ancestral micro-RNA biogenesis drives normal development in a modern plant

K. Aadland [1,*], H. Jia [2], B. Kolaczkowski [1 3]

[1]Microbiology and Cell Science, University Of Florida, Gainesville, [2]Biology, University of North Carolina at Chapel Hill, Chapel Hill, NC, [3]Genetics Institute, University Of Florida, Gainesville, United States

Abstract: A central goal of evolutionary biology is to better understand how the evolution of molecules encoded in the genome impacts the evolution of the phenotype. Ancestral sequence reconstruction (ASR) is a robust methodology for determining how substitutions in specific molecules affect their functions. However, current approaches have limited ability to characterize the effects of molecular-functional changes on the organism. Here we combine ASR with plant transgenics and quantitative RNA-Seq to directly assess how ancestral micro-RNA biogenesis impacted gene-regulatory processes and resulting whole-organism phenotype. We show that ancestral Hyponastic Leaves 1 (ancHYL1) -a dsRNA-binding protein contributing to miRNA biogenesis– could bind dsRNA targets and modern-day partner proteins with high affinity before the monocot-dicot split ~150mya, and rescues the knockout phenotype *in plantae*. RNA-sequencing suggests knocking-out HYL1, or replacing it with ancHYL1, has little impact on which miRNAs are produced. However, ancHYL1 does affect the relative abundance of miRNAs, and which strand of the miRNA duplex is preferentially loaded onto the RNA-induced silencing complex for RNA-interference. While ancHYL1 typically generated miRNA abundances closer to those of the wild type, there were significant differences in abundances between ancestral and modern HYL1 proteins. Together, these results suggest ancHYL1 sufficiently orchestrates normal organism development when expressed in a modern plant, although normal miRNA biogenesis is incompletely recovered. These observations suggest that plant development is relatively robust to variation in miRNA biogenesis and is the first direct evidence suggesting the miRNA biogenesis machinery originated early and has been largely conserved throughout the evolutionary history of plants.

Evolution of proteins and molecular pathways/networks under higher level selective effects

SMBE-PO-142

Ancient proteins and the thrifty gene hypothesis: Uric acid's contribution to primate evolution

J. E. Farrar [1,*], Z. Li [2], L. Tran [2], E. A. Gaucher [2]

[1]Biological Sciences, Georgia Institute of Technology, [2]Biology, Georgia State University, Atlanta, GA, United States

Abstract: The protein uricase, which breaks down the insoluble molecule uric acid, is found ubiquitously throughout each domain of life. However, certain animal lineages, including birds, reptiles, and apes, have experienced distinct evolutionary events leading to the pseudogenization of uricase. In order to understand the evolutionary process behind the inactivation of the uricase gene in the primate lineage, we have synthesized ancestral uricases by employing evolutionary models to predict their ancestral sequences. We discovered that uricase gradually lost activity prior to its inactivation. Ancestral sequence reconstruction on two uric acid transporters, URAT1 and ABCG2, also showed that the function of the ancient transporters changed concomitantly with the ancient uricases. Further, phylogenetic analysis of the enzyme that converts xanthine to uric acid (xanthine oxidoreductase, XOR) suggests that functional constraints acting on this enzyme changed at the same time that uricase and its transporters experienced a change in function. It is unknown why these events have occurred, particularly because the buildup of uric acid in the body causes hypertension, renal disease, liver damage and gout. However, with our collaborators, we have shown that increased uric acid levels are important for the conversion of fructose into triglycerides. These results support a recent hypothesis that frugivory is responsible for the evolution of large brain size in primates. Higher levels of uric acid would have facilitated the digestion of the fructose-rich diets of these primates, allowing more energy to be used for an increased encephalization quotient in primates.

Evolution of proteins and molecular pathways/networks under higher level selective effects
SMBE-PO-148
Evolution of the yeast microtubule organizing center
A. Cavanaugh [1,*], A. Grazzini [1]
[1]Biology, Creighton University, Omaha, United States

Abstract: Despite our growing understanding of molecular evolution, there is still much to be learned about how proteins that are part of large multiprotein complexes can undergo dramatic evolutionary changes without disruption of the larger structure. One major protein complex that faces this dilemma is the microtubule organizing center (MTOC). Although the function of MTOCs is highly conserved among eukaryotes their structure varies widely throughout the tree of life. In *S. cerevisiae* the MTOC, or spindle pole body (SPB) consists of repeating units of 18 distinct proteins that come together to form a rigid and highly structured organelle. Due to the SPB's involvement in appropriately segregating yeast chromosomes, 16 of the 18 SPB proteins are essential to the survival of *S. cerevisiae*. However, preliminary studies show that SPB proteins are not well conserved even among closely related yeast species. Not only is there a large amount of dissimilarity in protein sequence, but the presence of each SPB component in the genome varies wildly among yeast species. This finding is both surprising and useful for studying how large protein complexes can tolerate evolutionary changes in the proteins that comprise them. Furthermore, our findings suggest that the proteins involved directly in microtubule nucleation are the most broadly conserved, while components that make up the core of the *S. cerevisiae* SPB can only be found among other yeasts of the same genus. Additionally, even the most broadly conserved proteins are diverging more rapidly among yeasts than they are in other fungal species. This finding is particularly pronounced among species lacking a flagella, which has been postulated to have an evolutionary stabilizing effect on MTOC structure. Our data supports the hypothesis that lifting this organizational restriction has allowed for diversification among MTOCs.

Evolution of proteins and molecular pathways/networks under higher level selective effects

SMBE-PO-144

Conservation, Extensive Heterozygosity, and Convergence of Signaling Potential All Indicate a Critical Role for KIR3DL3 in Higher Primates

L. A. Leaton [1 2,*], J. Shortt [1], K. M. Kichula [1 2], S. Tao [1 2 3], N. Nemat-Gorgani [4 5], A. J. Mentzer [6], S. J. Oppenheimer [7], Z. Deng [8], J. A. Hollenbach [9], C. R. Gignoux [1], L. A. Guethlein [4 5], P. Parham [4 5], M. Carrington [10 11], P. J. Norman [1 2]

[1]Division of Biomedical Informatics and Personalized Medicine, [2]Department of Immunology and Microbiology, University of Colorado, Aurora, CO, United States, [3]Blood Center of Zhejiang Province, Hangzhou, China, [4]Department of Structural Biology, [5]Department of Immunology and Microbiology, Stanford University, Stanford, CA, United States, [6]Wellcome Trust Centre for Human Genetics and Jenner Institute, [7]Institute of Social and Cultural Anthropology, School of Anthropology and Museum Ethnography, University of Oxford, Oxford, United Kingdom, [8]Immunogenetics Laboratory, Shenzhen Blood Center, Shenzhen, China, [9]Department of Neurology, University of California, San Francisco, San Francisco, CA, [10]Basic Science Program, Frederick National Laboratory for Cancer Research, Frederick, MD, [11]Ragon Institute of the Massachusetts General Hospital, Massachusetts Institute of Technology and Harvard University, Boston, MA, United States

Abstract: Human natural killer (NK) cell functions are modulated by polymorphic killer cell immunoglobulin-like receptors (KIR). Among 13 *KIR* genes, which vary by presence and copy number, *KIR3DL3* is ubiquitously present in every individual across diverse populations. No ligand or function is known for KIR3DL3, but limited knowledge of expression suggests involvement in reproduction, likely during placentation. With 157 human alleles, KIR3DL3 is also highly polymorphic and we show heterozygosity exceeds that of HLA-B in many populations. The external domains of catarrhine primate KIR3DL3 evolved as a conserved lineage distinct from other KIR. Accordingly, and in contrast to other KIR, we show the focus of natural selection does not correspond exclusively to known ligand binding sites. Instead, a strong signal for diversifying selection occurs in the D1 Ig domain at a site involved in receptor aggregation, which we show is polymorphic in humans worldwide, suggesting differential ability for receptor aggregation. Meanwhile in the cytoplasmic tail, the first of two inhibitory tyrosine motifs (ITIM) is conserved, whereas independent genomic events have mutated the second ITIM of KIR3DL3 alleles in all great apes. Together, these findings suggest that KIR3DL3 binds a conserved ligand, and a function requiring both receptor aggregation and inhibitory signal attenuation. In this model KIR3DL3 resembles other NK cell inhibitory receptors having only one ITIM, which interact with bivalent downstream signaling proteins through dimerization. Due to the extensive conservation across species, selection, and other unusual properties, we are employing evolutionary and molecular analyses to elucidate the ligand and function of *KIR3DL3*.

Evolution of proteins and molecular pathways/networks under higher level selective effects
SMBE-PO-143
Reconstruction of ancestral protein sequences accounting for protein folding stability with ProtASR
M. Arenas [1,*], C. C. Weber [2], D. A. Liberles [2], U. Bastolla [3]
[1]Biochemistry, Genetics and Immunology, University of Vigo, Vigo, Spain, [2]Biology, Temple University, Philadelphia, United States, [3]Centre for Molecular Biology Severo Ochoa, Consejo Superior de Investigaciones Científicas, Madrid, Spain

Abstract: Reconstructed ancestral proteins have useful applications in biomedicine and biotechnology. A key step to obtaining these proteins is computational ancestral sequence reconstruction (ASR). Current ASR methodologies are based on a single matrix of rates of change among amino acid states and a single set of amino acid frequencies, usually obtained from empirical data, that are applied to all protein sites. However, it is known that protein evolution is highly heterogeneous due to variation in selective constraints among sites. In order to address this property, we developed an evolutionary framework called *ProtASR*, which reconstructs ancestral protein sequences accounting for selection on protein folding stability. This framework infers site-specific substitution matrices through a structurally constrained substitution model of protein evolution called mean-field (MF). It considers both unfolding and misfolding stability and outperforms traditional empirical models in terms of likelihood and correctly inferring amino acid distributions across sites. Applying extensive computer simulations we found that *ProtASR* yields reconstructed proteins with less biased stabilities than those obtained assuming a constant empirical matrix across all sites. Recently, we improved *ProtASR* by considering site-specific exchangeability matrices following the Halpern-Bruno model instead of a global exchangeability matrix as in the previous version. We also analyzed the evolution of the protein folding stability of several protein families, revealing fluctuations over time. *ProtASR* is freely available from https://github.com/miguelarenas/protasr and includes source code, detailed documentation and ready-to-use examples. The program runs in seconds/minutes depending on the analyzed alignment size.

Evolution of proteins and molecular pathways/networks under higher level selective effects
SMBE-PO-161A

The complex interplay between neutrality, innovability and robustness determines which genetic mechanisms can be successful in creating new molecular phenotypes

E. Bornberg-Bauer*

Abstract: Two fundamental question in molecular evolution ask
(i) how proteins can adapt to new functions while maintaining an existing one and
(ii) how new proteins with a selectable function, akin to tiny patches, can be found in a vast sequence space full of presumably non-functional and non-folding proteins.
Several theoretical models have been put forward to explain these apparent paradoxa.
The most popular models include neofunctionalization, subfunctionalization (SUBF) by degenerative mutations, and dosage models.
All of these models focus on adaptation after gene duplication.
"Escape from Adaptive Conflict" (EAC) includes adaptive processes before and after gene duplication that lead to multifunctional proteins, and divergence (SUBF).
Support for the importance of multifunctionality (promiscuity) for the evolution of new protein functions comes from two experimental observations.
First, many enzymes have highly evolvable promiscuous side activities.
Second, different structural states of the same protein can be associated with different functions.
How these observations may be related to the EAC model, under which conditions EAC is possible, and how the different models relate to each other is important to understand their selective effects.
We developed a theoretical framework that uses biophysical principles to infer the roles of functional promiscuity, gene dosage, gene duplication, point mutations, and selection pressures in the evolution of proteins.
We find that selection pressures can determine whether neofunctionalization or SUBF is the more likely evolutionary process.
Multifunctional proteins, arising during EAC evolution, allow rapid adaptation independent of gene duplication. This becomes a crucial advantage when gene duplications are rare.
We also propose that an increase in mutational robustness, not necessarily functional optimisation, can be the sole driving force behind SUBF.
Overall, this is the first model in which all three processes are unified and it is demonstrated that, given a certain rate of gene duplications and point mutations, the selection pressure determines which processes is most likely to be successful.
Furthermore, by mapping both RNA and protein-like models on a unified landscape with tunable neighbourhood properties, we recently demonstrated that the relationship between robustness and evolvability depends critically on the relationship of viable mutations which are neutral (coding for the same phenotype) and innovative (coding for a new phenotype) -- which has been ignored in all models hitherto.

Evolution of proteins and molecular pathways/networks under higher level selective effects
SMBE-PO-155
The influence of host and environment on protein content in two parasitoid wasp genomes
A. Dennis*

Abstract: Parasitoid wasps are a diverse group of insects whose development is usually fatal to their host. These abundant taxa are highly diverse, often with tight host specialization. Parasitoids that target aphids are abundant in the wild and also used in agricultural biocontrol. To successfully develop inside an aphid, parasitoids must overcome both aphid defenses and the products of the aphid's defensive microbiome. This long-held coevolutionary relationship has driven diversification and specialization. We have assembled *de novo* the full genomes of two parasitoid wasps that target aphids: *Lysiphlebus fabarum* and *Aphidius ervi*. The two genomes are highly syntenic and possess similar biases; most notably their coding genes have the lowest GC content of any arthropod yet sequenced. We have used comparisons of codon usage, GC content, Carbon and Nitrogen content, and gene expression to explore the relationship between this biased genome content and their host aphids. We hypothesize that the protein content in these parasitoid genomes is driven by biases in the host aphid genome, which in are turn influenced by their nutrient poor, sugar-rich diet.These results demonstrate how the environment can impact, and potentially constrain, genome-wide protein content, and how this impact may continue over multiple trophic levels.

Evolution of proteins and molecular pathways/networks under higher level selective effects
SMBE-PO-128
Evolutionary couplings detect side-chain interactions
A. J. Hockenberry [1,*], C. O. Wilke [1]
[1]Integrative Biology, The University of Texas at Austin, Austin, United States

Abstract: Patterns of amino acid covariation in large protein sequence alignments can inform the prediction of *de novo* protein structures, binding interfaces, and mutational effects. While algorithms that detect these so-called evolutionary couplings between residues have proven useful for practical applications, less is known about how and why these methods perform so well, and what insights into biological processes can be gained from their application. Evolutionary coupling algorithms are commonly benchmarked by comparison to true structural contacts derived from solved protein structures. However, the methods used to determine true structural contacts are not standardized and different definitions of structural contacts may have important consequences for interpreting the results from evolutionary coupling analyses and understanding their overall utility. Here, we show that evolutionary coupling analyses are significantly more likely to identify structural contacts between side-chain atoms than between backbone atoms. We use both simulations and empirical analyses to highlight that purely backbone-based definitions of true residue-residue contacts (i.e., based on the distance between Cα atoms) may underestimate the accuracy of evolutionary coupling algorithms by as much as 40% and that a commonly used reference point (Cβ atoms) underestimates the accuracy by 10-15%. These findings show that co-evolutionary outcomes differ according to which atoms participate in residue-residue interactions and suggest that accounting for different interaction types may lead to further improvements to contact-prediction methods.

Evolution of proteins and molecular pathways/networks under higher level selective effects

SMBE-PO-158

Expression divergence and duplicate retention: Insights from the nuclear pore complex

S. Aubé [1 2 3 4,*], A. Marchant [2 3 4], A. K. Dubé [1 2 3 4], F. D. Rouleau [1 3 4], I. Gagnon-Arsenault [1 2 3 4], D. Ascencio [2 3 4], P. C. Després [1 3 4], C. R. Landry [1 2 3 4]

[1]Département de biochimie, de microbiologie et de bio-informatique, [2]Département de biologie, [3]Institut de biologie intégrative et des systèmes, [4]PROTEO, Le réseau québécois de recherche sur la fonction, la structure et l'ingénierie des protéines, Université Laval, Québec, Canada

Abstract: Gene duplication is a frequent event throughout evolution and is thus one of the main sources of new proteins. Traditionally, the maintenance of duplicated genes is thought to involve functional divergence, either through the acquisition of new functions (neofunctionalization) or through the partition of ancestral functions (subfunctionalization). However, dosage constraints may also play an important role in shaping the fate of duplicates. It has indeed been postulated that, after an event of whole-genome duplication (WGD), the retention of both copies of a gene is favored until the expression level of either one becomes low enough for its loss to be effectively neutral. Yet, the factors underlying the retention of duplicated genes remain to be fully elucidated. In order to provide insights on the determinants of the evolutionary trajectory followed by duplicates, we combined CRISPR-Cas9 genome edition with the high-throughput screening of protein-protein interactions to reconstruct the history of paralogous nucleoporins from the nuclear pore complex (NPC) of the yeast *Saccharomyces cerevisiae*. Our results show that most functional change following the duplication happened symmetrically in both copies and suggest that expression divergence is in great part responsible for the retention of duplicates, particularly in the case of the *NUP53*/*NUP59* couple, which is almost exclusive to the *Saccharomyces* genus. This work thus supports the idea that stochastic regulatory changes are a major driver of duplicated genes' maintenance and loss.

Evolution of proteins and molecular pathways/networks under higher level selective effects

SMBE-PO-160

Detoxification systems expression patterns of Chromera velia in response to Mercury stress condition.

A. Sharaf [1,*], R. De Michele [2], A. Sharma [3], M. Oborník [3]

[1]Institute of Parasitology, Biology Centre Ascr Institute Of Parasitology, Vat No.: Cz60077344, České Budějovice, Czech Republic, [2]nstitute of Biosciences and Bioresources (IBBR), National Research Council (CNR) of Italy, Palermo, Italy, [3]Institute of Parasitology, Biology Centre, Czech Academy of Sciences, České Budějovice, Czech Republic

Abstract: The heavy metal pollution in ecosystems is of increasing global concern. The main sources of the metal toxicity are the industrial waste, mining and using of pesticides; which includes heavy metals. Out of all heavy metals, mercury (Hg) is considered to be the one, easily accumulated in the aquatic organism. Hg can induce more severe oxidative stress by triggering production of reactive oxygen species (ROS) and damage macro-molecules. ROS serve not only as dangerous molecules that damage proteins, lipids and DNA but also as signalling molecules in the regulation of biological processes such as biotic and abiotic stress responses. Xenobiotics detoxification pathway “Green Liver-like” is another defence system, which can activate due to heavy metal stress. This study aims to explore the change in the *Chromera velia* transcriptome due to Hg stress; *C. velia* is the recently discovered closest known relative of Apicomplexans. The results show that 1239 genes differentially expressed; Differentially Expressed Genes (DEGs) showed overall up-regulation (1,070 genes) while only 169 genes were down-regulated. Moreover, 145 ROS-mediated related unigenes identified; of them, 125 (86.2 %) unigenes were showed similar expression patterns while 20 showed Differential Expression (DE) patterns. Out of 126 xenobiotics metabolism related identified unigenes; only fifteen (11.9 %) unigenes showed DE patterns while 111 (88.1 %) unigenes showed similar expression patterns. Our study presents the first deep transcriptomic analysis of *C. velia*, focusing on the expression level of genes involved in various detoxification defence systems in response to heavy metal stress.

Evolution of proteins and molecular pathways/networks under higher level selective effects
SMBE-PO-126

Prevalent epistatic interactions between amino acid sites in Schizophyllum commune

A. V. Stolyarova [1,*], A. V. Bezmenova [1], A. S. Kondrashov [2], G. A. Bazykin [1]
[1]Skolkovo Institute of Science and Technology, Moscow, Russian Federation, [2]University of Michigan, Ann Arbor, United States

Abstract: Many lines of evidence indicate that amino acid sites of protein-coding genes are involved in tight networks of epistatic interactions, but revealing the identity of these interactions at a genome scale is challenging. The basidiomycete fungus *S. commune* is the most variable eukaryotic species known, with ~20% of neutral sites differing between any two individuals. This provides the possibility to study the forces affecting the polymorphism level with unprecedented resolution. Here, we sequence complete genomes of 54 individuals from two populations of *S. commune.* In each population, we describe a high prevalence of both short- and long-range linkage disequilibrium (LD) between nonsynonymous, but not synonymous, sites. The increase in LD is particularly pronounced for pairs of nonsynonymous sites within the same gene, but is also observable between genes. Simulations show that the elevated nonsynonymous LD cannot result from differences in negative selection on the considered sites, or from selective sweeps or background selection affecting at neighboring sites. Instead, it is evident of abundant epistasis between sites. The high polymorphism level in *S. commune* provides the first opportunity to use population sequencing data to infer the network of intra- and intergenic interactions.

Evolution of proteins and molecular pathways/networks under higher level selective effects
SMBE-PO-145

The impact of protein architecture on adaptive evolution

A. F. Moutinho [1,*], J. Yann Dutheil [1 2]
[1]Evolutionary Genetics, Max Planck Institute for Evolutionary Biology, Plön, Germany, [2]Institut des Sciences de l'Evolution, CNRS, IRD, EPHE, Université de Montpellier, Montpellier, France

Abstract: Understanding the molecular basis of adaptation is a challenge that requires to assess the frequency and nature of mutations along the genome. Several studies have addressed the factors influencing the molecular adaptive rate within genomes, providing evidence for the key-role of variables such as mutation and recombination rate. At the intra-genic level, however, little is known about the drivers of adaptive evolution. To address this, we analysed the impact of protein architecture on the rate of adaptive substitutions, aiming to understand how protein biophysics influences fitness and adaptation. Using a population genomics approach, we fitted models of distribution of fitness effects and estimated the rate of adaptive evolution at the protein and amino-acid residue level. We found that the relative solvent accessibility is a major driver of protein adaptation, with exposed residues evolving much faster than residues at the core of proteins. Such effect was previously reported at the divergence level (*dN/dS*), but here we show that the faster evolution at the protein's surface is explained both by a relaxed purifying selection and an increased rate of adaptive evolution. Moreover, we observe that the rate of adaptive substitutions differs between protein functional classes, with genes involved in protein regulation and signalling pathways exhibiting the fastest rates of adaptation. Our results therefore suggest that the greater accumulation of adaptive mutations at the surface of proteins is mainly driven by inter-molecular interactions, either at the intra-cellular level via network rewiring, or at the inter-organism level through host-pathogen coevolution.

Evolution of proteins and molecular pathways/networks under higher level selective effects
SMBE-PO-147
Functional adaptations of dim-light vision in aquatic environments
Y. L. I. Chiu [1,*], B. S. Chang [1 2]
[1]Cell & Systems Biology, [2]Ecology & Evolutionary Biology, University of Toronto, Toronto, Canada

Abstract: Dim-light dwelling species have acquired various visual adaptations including molecular changes within their visual system to accommodate for their lifestyle as vision is one of the key senses for survival. The drastic decrease in light intensity as water depth increases makes the aquatic environment an ideal system to study dim-light adaptation. In this project, we investigated the evolution of visual pigments in dim-light adapted aquatic species using a combination of computational methods to analyse selective pressures and *in vitro* expression and spectroscopic assays to measure wavelength sensitivities and kinetic properties. We have found evidence for adaptive evolution in the visual transduction proteins of dim-light dwelling species which appear to correlate with their ecological habitat. Our results provide insight into possible molecular mechanisms underlying the functional adaptations of dim-light vision.

Evolution of proteins and molecular pathways/networks under higher level selective effects
SMBE-PO-153

Regulatory rewiring events in Pseudomonas fluorescens Gene Regulatory Network Evolution

M. J. Shepherd [1,*], J. Horton [1], T. Taylor [1]
[1]Biology and Biochemistry, University of Bath, Bath, United Kingdom

Abstract: Gene Regulatory Networks (GRNs) are the fundamental control circuits that determine a significant proportion of an organism's phenotype and responses to its environment. Consequently, GRNs are hotbeds for phenotypic evolution as mutational change to a small number of master regulators can achieve drastic change to complex multi-gene behaviours. Using experimental evolution within a thoroughly characterised model system we can study the drivers and nuances of GRN evolution. This study focusses on the soil bacterium *Pseudomonas fluorescens* with the flagellum master regulator *fleQ* deleted; under selection motility is rescued in a highly repeatable manner by co-opting the homologous nitrogen metabolism regulator *ntrC*. Alternatives to the functionally promiscuous regulator *NtrC*, which are also capable of undergoing evolutionary rewiring to rescue flagellar motility were identified by deletion of *ntrC*. Strains of *ΔntrC P. fluorescens* are forced to take alternative evolutionary pathways to rescue motility. Fitness assays reveal these alternative pathways do not appear to carry a severe fitness trade-off, however motile mutants take far longer to appear. This suggests that some regulators have a greater potential for driving the evolution of gene regulatory networks via rewiring events than others, as they can be re-purposed faster than the alternatives.

Evolution of proteins and molecular pathways/networks under higher level selective effects
SMBE-PO-150

Mammalian APLP1 adaptation to neural system : Phylogenetic analysis on functional divergence of APP family

W. Onodera [1,*], A. Toru [1 2], N. Sawamura [2]
[1]Science and Engineering, [2]Research Organization for Nano & Life Innovation, Waseda university, Tokyo, Japan

Abstract: APP (Amyloid precursor protein) family maintains a normal neuronal network in developing and matured brains. Various functions including cell adhesion, synaptogenesis, and receptor-like signaling are associated with its maintenance. Within the APP family, APLP1 shows differentiated localization and augmented properties in mammalian brain suggesting specialized role in CNS. However, evolutionary trait of APLP1 during adaptation towards CNS has not been described. Here we show the transition of evolutionary pressure on vertebrate APP family, and further analyzed coevolutions with spatially and functionally APLP1-related biomolecules. Using evolutionary rate (dN/dS) analysis, increased dN/dS was detected at heparin binding domain (HBD) of APLP1. Coevolution with neuron specific proteins were confirmed that may explain how mammalian APLP1-HBD has evolved for CNS. HBD was reported as heparin-binding site which leads to dimerization and functionalization of APP family excluding APLP1. As such, we assumed high dN/dS at heparin-binding site of mammalian APLP1 may be a result of relaxation of functional constraints. Interestingly, docking simulation revealed alternative heparin binding site of mammalian APLP1 suggesting distinct heparin-induced dimerization mode consistent with *in vitro* studies of APLP1. Our results indicate reinforced molecular adaptation of APLP1 toward CNS, directed to development of mammal-specific brain structure. Focusing on functional divergence of APP family would support confirming their specific physiological role that cannot be detected *in vitro* and *in vivo* studies.

Evolution of proteins and molecular pathways/networks under higher level selective effects
SMBE-PO-149

Regulatory potential conferred by the nucleotide sequences of high-frequency short amino acids fragments in Arabidopsis thaliana

Y. Li [1,*], Z. Li [1], Y. Yan [1], R. Yang [1]
[1]College of Life Sciences, Northwest A&F University, Yangling, China

Abstract: The birth and death of gene families create dynamic reservoir of protein domains that contributes to the phenotypic evolution of organisms. In comparison with the domains that are usually composed of dozens of amino acids and have full molecular function, short amino acid fragments with 3~8 consecutive amino acids in length are largely unexplored with regard to their function and evolution. These short fragments vary in the composition and the frequency in the proteome of an organism, and some of them have extremely larger numbers than expected. Most previous studies have mainly investigated them at the protein level. While recent studies revealed a high proportion of protein-coding sequences in animals and plants with additional regulatory capacity. Thus, toward a better understanding of the function and evolution of short amino acid fragments, 100 four-amino acid fragments (*4*-mers) with the highest frequency in *Arabidopsis thaliana* were focused on and a comprehensive investigation on them at the protein, RNA and DNA levels was performed. Strikingly, 92 *4*-mers were preferentially located outside of protein domain regions, suggesting considerable regulatory potential encoded in the nucleotides of high-frequency short fragments. Analysis on the similarities between the position weight matrices (PWMs) of transcription factors (TFs) and the PWM derived from the nucleotide instances of a given *4*-mer provided compelling evidence for 94 *4*-mers as partial or full DNA providers of transcription factor binding sites (TFBSs) targeted by at least one of the 619 inspected TFs. In contrast, only 25 *4*-mers tended to possess regulatory function at the RNA level. But it may be seriously underestimated since only 41 RNA-binding proteins were analyzed. The effect of such regulatory potential on codon usage was examined. The codon usage of all the 15 amino acids that constitute the 100 *4*-mers displayed significant difference for the DNA sequences between within and outside the instance of these high-frequency *4*-mers. To determine the codon usages of which amino acids are more likely to be constrained by the regulatory code than the genetic code, the preferred codons corresponded to the *4*-mer instances that happen to overlap with putative TFBSs and the trinucleotide combinations favored by TFs in non-coding regions were further identified. As a result, threonine and proline were discovered that their respective preferred codons used in the high-frequency TFBS-overlapped *4*-mer instances and the trinucleotide combinations were the same, which is different from that in non-high-frequency *4*-mer instances. Therefore, our analysis suggested that the dual functions of the high-frequency *4*-mer instances have affected the codon choice of amino acid in the proteome. Lastly, we estimated the divergence degree of homologous DNA of *4*-mer instances between *Arabidopsis thaliana* and *Arabidopsis lyrata* and found 11 *4*-mers evolved slower than background sequences, which was probably due to the additional selective forces from the potential regulatory role. Overall, using *Arabidopsis* as a model, our findings provide novel insights into proteome evolution in plants and multiple information accommodation in small-piece coding DNA sequences. These results support that the nucleotide sequences encoding high-frequency *4*-mers may have great potential as regulatory elements to recruit trans-acting factors at the DNA or RNA levels.

Evolution of proteins and molecular pathways/networks under higher level selective effects
SMBE-PO-152

Pre-adaptation and positive selection drive evolution of a novel function of J-protein co-chaperone of Hsp70.

M. Stolarska [1,*], B. Tomiczek [1], O. Shrestha [2], R. Sharma [2], W. Lee [2 3], M. Tonelli [2 3], G. Cornilescu [2 3], J. Markley [3], L. Nierzwicki [4], J. Czub [4], S. Ciesielski [2], E. Craig [2], J. Marszalek [1 2]

[1]Laboratory of Evolutionary Biochemistry, Intercollegiate Faculty of Biotechnology, University of Gdansk and Medical University of Gdansk, Gdansk, Poland, [2]Biochemistry, University of Wisconsin-Madison, [3]National Magnetic Resonance Facility at Madison, Madison, United States, [4]Physical Chemistry, Gdansk University of Technology, Gdansk, Poland

Abstract: The J-protein Zuotin is a multi-domain eukaryotic Hsp70 co-chaperone. Though it is primarily ribosome-associated, where it promotes folding of nascent polypeptide chains, Zuotin also has off-ribosome functions. Domains of Zuotin needed for 60S association and interaction with Hsp70 are conserved in eukaryotes. However, whether the 4-helix bundle (4HB) domain is conserved remains an open question. We undertook evolutionary and structural approaches to clarify this issue. We found that the 4HB segment of human Zuotin also forms a bundle of 4 helices. The positive charge of Helix I, which in *Saccharomyces cerevisiae* is responsible for interaction with the 40S subunit, is particularly conserved. However, the C-termini of fungal and human 4HBs are not homologous. In fungi the C-terminal segment forms a plug that folds back into the bundle; in *S. cerevisiae* it plays an important role in bundle stability and, off the ribosome, in transcriptional activation. In human, C-terminal helix IV of the 4HB is extended, protruding from the bundle. This extension serves as a linker to the regulatory SANT domains, which are present in animals, plants and protists, but not fungi. Further analysis of Zuotin sequences revealed that the plug likely arose as a result of genomic rearrangement upon SANT domain loss early in the fungal lineage. In the lineage leading to *S. cerevisiae*, the 4HB was subjected to positive selection with the plug becoming increasingly hydrophobic. Eventually, these hydrophobic plug residues were coopted for a novel regulatory function - activation of a recently emerged transcription factor, Pdr1.

Evolution of proteins and molecular pathways/networks under higher level selective effects

SMBE-PO-151

Pre-adaptation and positive selection drive evolution of a novel function of J-protein co-chaperone of Hsp70.

M. Stolarska [1,*], B. Tomiczek [1], O. K. Shrestha [2], R. Sharma [2], W. Lee [2 3], M. Tonelli [2 3], G. Cornilescu [2 3], L. Nierzwicki [4], J. Czub [4], J. L. Markley [2 3], S. J. Ciesielski [2], E. A. Craig [2], J. Marszalek [1 2]

[1]Laboratory of Evolutionary Biochemistry, Intercollegiate Faculty of Biotechnology, University of Gdansk and Medical University of Gdansk, Gdansk, Poland, [2]Department of Biochemistry, University of Wisconsin-Madison, [3]National Magnetic Resonance Facility at Madison, Madison, United States, [4]Department of Physical Chemistry, Gdansk University of Technology, Gdansk, Poland

Abstract: The J-protein Zuotin is a multi-domain eukaryotic Hsp70 co-chaperone. Though it is primarily ribosome-associated, where it promotes folding of nascent polypeptide chains, Zuotin also has off-ribosome functions. Domains of Zuotin needed for 60S association and interaction with Hsp70 are conserved in eukaryotes. However, whether the 4-helix bundle (4HB) domain is conserved remains an open question. We undertook evolutionary and structural approaches to clarify this issue. We found that the 4HB segment of human Zuotin also forms a bundle of 4 helices. The positive charge of Helix I, which in *Saccharomyces cerevisiae*is responsible for interaction with the 40S subunit, is particularly conserved. However, the C-termini of fungal and human 4HBs are not homologous. In fungi the C-terminal segment forms a plug that folds back into the bundle; in *S. cerevisiae*it plays an important role in bundle stability and, off the ribosome, in transcriptional activation. In human, C-terminal helix IV of the 4HB is extended, protruding from the bundle. This extension serves as a linker to the regulatory SANT domains, which are present in animals, plants and protists, but not fungi. Further analysis of Zuotin sequences revealed that the plug likely arose as a result of genomic rearrangement upon SANT domain loss early in the fungal lineage. In the lineage leading to *S. cerevisiae*, the 4HB was subjected to positive selection with the plug becoming increasingly hydrophobic. Eventually, these hydrophobic plug residues were coopted for a novel regulatory function - activation of a recently emerged transcription factor, Pdr1.

Evolution of proteins and molecular pathways/networks under higher level selective effects
SMBE-PO-156

Pleiotropic impacts of regulatory mutations and their relationship to fitness

P. Vande Zande [1,*], F. Duveau [2], B. Metzger [2], P. V. Wittkopp [1 2]
[1]Department of Molecular, Cellular, and Developmental Biology, [2]Department of Ecology and Evolutionary Biology, University of Michigan, Ann Arbor, United States

Abstract: Phenotypic evolution is often the result of changes in gene expression. Despite their importance, the constraints and biases that influence how gene expression evolves are not fully understood. For example, *cis*-regulatory divergence has been shown to explain increasing amounts of expression divergence over evolutionary time, but the reasons for this pattern remain to be explained. One possible explanation involves pleiotropy, or the impact of a mutation on multiple independent traits. If mutations with higher pleiotropy are on average more detrimental to fitness than those with low pleiotropy, they may be preferentially removed from populations over time by natural selection. In addition, if *trans*-regulatory mutations are more pleiotropic than *cis*-regulatory mutations, they may be purged more frequently, increasing the proportion of gene expression divergence explained by *cis*-regulatory divergence over evolutionary time. To test the relationship between fitness and pleiotropy for both *cis*- and *trans*-regulatory mutations, RNA-sequencing was performed on a set of yeast strains containing regulatory mutations affecting expression of the *TDH3* gene in both *cis and trans.* These data are then used in conjunction with measures of fitness to explore how regulatory mutations' affects throughout genetic networks relate to organismal fitness. Understanding how these pleiotropic effects impact fitness helps explain observed trends in regulatory evolution and inform models of gene expression evolution.

Evolution of proteins and molecular pathways/networks under higher level selective effects

SMBE-PO-146

Simplification of ribosomes in bacteria with tiny genomes

D. Nikolaeva [1,*], S. Garushyants [2], M. Gelfand [2]

[1]Department of Bioengineering and Bioinformatics, Lomonosov Moscow State University, [2]IITP RAS, Moscow, Russian Federation

Abstract: The ribosome is an essential cell organelle performing protein biosynthesis. Its structure and composition are highly conserved in all species. However, some bacteria have been reported to have an incomplete set of ribosomal proteins. We have analyzed ribosomal protein composition in 214 small bacterial genomes (< 1 Mb) and found that although the ribosome composition is fairly stable, some ribosomal proteins may be absent, especially in bacteria with dramatically reduced genomes. The protein composition of the large subunit is less conserved than that of the small subunit. We have identified the set of frequently lost ribosomal proteins and demonstrated that they tend to be situated on the ribosome surface and have fewer contacts to other ribosome components. Moreover, loss of some proteins is correlated at least in one case leading to the loss of a complete functional block. Additionally, the reduction of rRNA is also common in bacteria with short genomes, and deletions mostly occur in free loops. The loss of ribosomal protein L24 occurs simultaneously with the loss of the corresponding binding structure in rRNA. Finally, the loss of the anti-Shine-Dalgarno sequence is not correlated with the genome size or the loss of particular ribosomal proteins.

Evolution of proteins and molecular pathways/networks under higher level selective effects
SMBE-PO-154

Proteomic thermal stability profiling for elucidating protein-ligand interactions and downstream pathway analysis

F. Feyertag [1 2,*], J. Ward [1 2], K. V. M. Huber [1 2]
[1]Structural Genomics Consortium, [2]Target Discovery Institute, University of Oxford, Oxford, United Kingdom

Abstract: One of the key challenges in drug discovery is gathering a comprehensive understanding of the molecular targets of a compound on a global cellular level. A well-established notion is that proteins tend to increase in thermal stability when bound with a ligand. Over recent years, quantitative mass-spectrometry has been utilised to characterise the thermal stability of proteins in the presence of a ligand on a proteomic scale, termed thermal profiling (TP). This gives ample opportunity to identify potential on- and off-targets of a drug by identifying proteins whose stability is altered upon compound treatment. Furthermore, identifying networks of proteins that are stabilised or destabilised may provide insights into the downstream effects of a drug, including the rewiring of protein interaction networks and metabolic pathways within a cell.

We developed the Thermal Profiling Meltome Analysis Package (TPMAP) software to support systems-biology level analysis of the thermal characteristics of proteins and their pathways from TP datasets. TPMAP allows a user to easily identify proteins that are either stabilised or destabilised in TP datasets, and investigate protein complexes and protein-protein interactions associated with these proteins. We demonstrate TPMAP on a TP dataset characterising a compound that emerged from a phenotypic drug screen, potentially targeting the Vacuolar-type H^+-ATPase (V-ATPase) complex with an apparent distinct mode of action from existing V-ATPase inhibitors. We anticipate TPMAP will help researchers to analyse TP datasets, and broaden its application to better understand drug phenotypes or other environmental factors, and associated metabolic regulation of cellular pathways.

Evolutionary genetics and genomics of metabolic networks

SMBE-PO-166

Sex-specific transcriptomic responses to changes in the nutritional environment

F. Camus [1,*], M. Piper [2], M. Reuter [3]

[1]Genetics, Evolution & Environment, University College London, London, United Kingdom, [2]School of Biological Sciences, Monash Univeristy, Melbourne, Australia, [3]University College London, London, United Kingdom

Abstract: Males and females have divergent reproductive strategies and thus need to consume diets that maximise sex-specific fitness. While much research has focused on identifying nutritional requirements for each sex, little is known about the molecular mechanisms underlying this response. Here we first identify male- and female-optimal diets using nutritonal geometry techniques. We then examine the transcriptional changes that occur within each sex when going between these two optimal diets. We find that although most of the sex-specific transcriptome responds similarly to dietary changes, a subset of genes diverge in their response. We then characterise the regultary network underpinning both sexually concordant and antagonistic genes. Finally, we show that although core metabolic genes show sexually concordant changes, male- and female-specific reproductive genes respond to diet in opposing directions. Thus, sex-limited genes are almost invariably up-regulated in the conditions that maximises each sex's fitness. This implies that the shared nutrient-sensing signal is inverted to produce diametrically opposed regulation of reproductive genes in males and females, thereby allowing each sex to achieve high fitness at their optimal nutritional environment. Our work provides deeper insights into the link between sex-specific nutrition and reproduction.

Evolutionary genetics and genomics of metabolic networks
SMBE-PO-162
Using network clustering to investigate the evolution of structured RNA motifs regulating gene expression
M. Crum [1,*], N. Ram Mohan [1], M. M. Meyer [1]
[1]Biology, Boston College, Chestnut Hill, United States

Abstract: Structured RNA motifs play vital roles in all kingdoms of life. They are essential in protein translation (tRNAs and rRNA), perform catalytic functions (ribozymes), and regulate gene expression (riboswitches). Understanding the evolution and conservation of structured RNAs provides insight into the cellular processes in which they are involved. However, accounting for RNA secondary structure provides different challenges than those typically encountered in protein evolution investigations. To overcome this, we implement network clustering analysis, in which each vertex corresponds to an RNA sequence and edges are weighted based on a distance metric (e.g. RNA structure, sequence similarity, or some combination of both). Using this approach, we investigated the two forms of the glycine riboswitch, an RNA motif found in bacterial mRNA that regulates gene expression in response to glycine concentration. The singleton form contains one glycine-binding domain, and the tandem form contains two homologous glycine-binding domains. While the tandem conformation is well studied biophysically, it is not clear how/why it arose and became conserved across bacteria.
Initial phylogenetic analysis showed riboswitches grouping based on the regulated gene (genomic context), as opposed to their taxonomic origin. We then utilized network clustering to investigate the conservation of sequence and secondary structure of glycine riboswitches. Using these networks, we identified communities of closely related glycine-binding domains. This clustering revealed that genomic context effects which tandem glycine-binding domain is more highly conserved: the first domain is more highly conserved in riboswitches regulating the glycine cleavage system (GCV), while the second is more highly conserved in those regulating transport proteins (TP). It also revealed that singleton riboswitches are more similar to the first or second tandem domain based on genomic context: singletons regulating GCV are more similar to the first domain of tandems regulating GCV, while singletons regulating TP are more similar to the second domain of tandems regulating TP. Taken together, these findings paint a picture of tandem glycine riboswitches degrading into functional singletons, with the genomic context dictating which glycine-binding domain is conserved.

Evolutionary genetics and genomics of metabolic networks

SMBE-PO-175

Honor by association, leveraging co-expression networks for gene discovery in specialized metabolism

J. Wisecaver [1 2,*], R. Auber, A. Pendleton

[1]Department of Biochemistry, [2]Center for Plant Biology, Purdue University, West Lafayette, United States

Abstract: Specialized metabolites serve myriad biological functions that allow organisms to interact with and manage their environment (e.g., resist abiotic stress, combat negative ecological interactions and promote beneficial ones). These metabolites are synthesized in response to dynamic ecological pressures, and as a consequence, the pathways involved in metabolite biosynthesis are often fast-evolving, lineage-specific, and remain uncharacterized at the genetic level. This hampers our ability to understand metabolic gene innovation at the level of individual pathways. Critically, this also limits the potential utility of specialized metabolites in agriculture, pharma, and biotech applications. To address this challenge, we developed the mutual ranks to modules workflow, a method for identifying small, overlapping modules of co-expressed genes in global co-expression networks. These modules serve as the basis for high-throughput prediction of specialized metabolic pathways. Using the model plant Arabidopsis, modules accurately recovered the enzymatic genes of functionally characterized specialized pathways as well as genes involved in pathway regulation and metabolite transport. Importantly, a co-expression network approach can straightforwardly be applied to any species, model and non-model, so long as the organism's transcriptome can be sampled under a range of ecologically relevant conditions. Currently, we are utilizing Oxford Nanopore long-read sequencing technology to construct highly contiguous genome assemblies of eukaryotes with expanded metabolic repertoires to identify gene candidates for pathways of interest and characterize the genes' location(s) within genomes. With these data we aim to evaluate the degree to which specialized metabolic pathways form of biosynthetic gene clusters in different eukaryotic lineages. The utility of this approach is illustrated by ongoing work in our lab to characterize various pathways for the production of specialized metabolites, from plant allelochemicals to algal phycotoxins.

Evolutionary genetics and genomics of metabolic networks

SMBE-PO-177

Screens for convergent gene inactivation uncover changes in biological processes associated with herbivory and carnivory in mammals

N. Hecker[1 2 3,*], V. Sharma[1 2 3 4 5], M. Hiller[1 2 3]

[1]Max Planck Institute of Molecular Cell Biology and Genetics, [2]Max Planck Institute for the Physics of Complex Systems, [3]Center for Systems Biology Dresden, [4]Deutsche Forschungsgemeinschaft-Center for Regenerative Therapies Dresden, [5]Technische Universitaet Dresden, Dresden, Germany

Abstract: Genomic changes associated with the adaptation to different ecological niches like the dietary specialization into herbivores and carnivores remain poorly understood. To identify genes which were inactivated due to dietary specialization, we employed a systematic genome-wide screen for the detection of convergently inactivated protein-coding genes in 16 obligate herbivores and 15 obligate carnivores. Herbivores exhibit the convergent inactivation of the pancreatic triglyceridelipase inhibitor PNLIPRP1 and of the pancreatic zymogen secretion enhancing SYCN gene which likely reflects their constant feeding behavior and fat-poor diet. In carnivores that feed on irregular intervals and may perform constant gluconeogenesis, convergently inactivated genes include the ligand receptor pair, INSL5-RXFP4, which regulates glucose homeostasis and appetite. Several carnivore genomes also show the convergent inactivation of two receptors for the detection of xenobiotics, NR1I2 and NR1I3, which is likely related to a reduced exposure to toxic plant-derived compounds. Furthermore, our screen discovered the convergent inactivation of the antimicrobial NADPH oxidase NOX1 in carnivores that may be related to the less diverse gut microbiomes of carnivores. Our work highlights that phenotype-to-genotype screens for convergently inactivated genes can help us to uncover molecular biological processes that coincide with evolutionary changes.

Evolutionary genetics and genomics of metabolic networks

SMBE-PO-178

Structure of genes and regulatory elements related to the biosynthesis of organic acids and sugars in Citrus spp

C. S. Argolo [1], F. Micheli [1], R. R. da Fonseca [2,*]

[1]Department of Biology, Universidade Estadual de Santa Cruz, Ilhéus, Brazil, [2]Department of Biology, University of Copenhagen, Copenhagen, Denmark

Abstract: The culture of citrus has great relevance in the global socioeconomic scenario, being present in countries of temperate and tropical climate. The most relevant characteristics for the quality of citrus fruit are color and taste; the latter being dependent on the balance between the acid and sugar contents. One key limitation of citrus production in Brazil is the fact that the varieties grown in tropical climate have a higher acid content than that obtained in temperate climates. In this project, I am investigating the pathways responsible for the metabolism of sugars and organic acids in various citrus species of agronomic interest. I started by identifying the genes involved in the citric acid pathway in model plants, with well annotated genomes, including algae, bryophytes, pteridophytes, monocotyledons and dicotyledons. I then proceeded to annotate them in 10 citrus cultivars and did a comparative genomics analyses to assess the rates of evolution within citrus and between all the plant species in the data set. I then characterized the genomic context around these genes in the different genomes. This then allowed us to identify upstream and downstream cis-elements that could be responsible for balance in sweetness and acidity in citrus. Overall, this project will provide fundamental knowledge regarding the regulation of the genes directly involved in obtaining a high-quality fruit, and how this relates to the diversity of each cultivar in the genus Citrus. Ultimately, we aim at designing improvement strategies for the genetic enhancement of fruit quality in tropical regions.

Evolutionary genetics and genomics of metabolic networks
SMBE-PO-169
Wing pheromones in Heliconius butterflies: physiology, behavior, and genetics
K. J. Byers [1,*]
[1]Department of Zoology, University of Cambridge, Cambridge, United Kingdom

Abstract: Butterflies in the genus *Heliconius* (Nymphalidae) have been extensively studied as key examples of speciation, mate choice, and Müllerian mimicry via their bright colour patterns. Male chemical signaling via wing pheromones has recently been demonstrated in *H. melpomene* and shown to have an effect on female choice. Our understanding of chemical signaling in reproductive isolation between *Heliconius* species is still in its infancy. Using electroantennography, I examined responses of female *H. melpomene* and *H. cydno* to natural and synthetic male pheromones of both species, as well as to individual dominant compounds. Surprisingly, *H. melpomene* and *H. cydno* females both react more strongly to wing pheromones from *H. cydno*. Of the major components of the wing pheromone in both species, only octadecanal (26% of the *H. melpomene* male pheromone) provoked a significant response in both species, despite its absence in *H. cydno*. When female *H. melpomene* were presented with a choice between a control male and one augmented with additional octadecanal, they showed a slight preference for the control male. Preliminary QTL analysis shows a locus on chromosome 20 with no linkage to the known major wing colour or mate choice loci. QTL for production of both octadecanal and its likely precursor octadecanol are strongly overlapping, in agreement with the likely biosynthetic route of octadecanal from stearic acid. The candidate region contains 15 likely fatty acyl-CoA reductase genes and two alcohol dehydrogenase genes, one copy of which contains a deletion in *H. cydno* that likely interrupts a splice site and may confer nonfunctionality. This tight grouping of candidate loci seems to have undergone a lineage-specific duplication in *Heliconius*. This work is the first to show female physiological responses to male wing pheromones in *Heliconius*, and argues for their importance in both female mate choice and reproductive isolation in combination with other mate choice signals. It also highlights the potential role of octadecanal, a relatively simple but rarely used pheromone component, in mate choice and reproductive isolation.

Evolutionary genetics and genomics of metabolic networks
SMBE-PO-168
Adaptations to deadly diets: how poison frogs metabolize pumiliotoxin
A. Alvarez-Buylla [1,*], L. Coloma [2], L. O'Connell [1]
[1]Biology, Stanford University, Stanford, United States, [2]Centro Jambatu, Quito, Ecuador

Abstract: All animals have mechanisms to deal with ingested toxic small molecules. For humans, these molecules include commercially available oral pharmaceuticals, which once ingested are often metabolized or cleared, reducing their efficacy. South American poison frogs have evolved to sequester the small molecule alkaloids present in their normal insect diet and accumulate them in skin glands to serve as a chemical defense against predation. The toxin repertoire carried by different species or populations of frogs is variable, and some poison frog species are further able to metabolize toxins into more potent forms. One example of toxin metabolism is that of the alkaloid Pumiliotoxin (PTX), which certain poison frog species enantioselectively convert into the more toxic Allopumiliotoxin (aPTX). This project used thermal proteome profiling to screen poison frog proteins that may interact with PTX, and identified several Cytochrome P450s that may be responsible for this metabolism. In parallel, the ability to metabolize PTX was assessed in the lab using tissue lysate from different species and populations of poison frogs. We then performed a P450 sequence comparison of these different species who can and cannot convert PTX into aPTX to map enzyme function to protein sequence variation. Together, these findings shed light on the enzymatic mechanism responsible for the metabolism of PTX in some poison frogs.

Evolutionary genetics and genomics of metabolic networks

SMBE-PO-171

Ribosome provisioning in Pseudomonas fluorescens activates a bistable switch coupled to fast exit from stationary phase

P. Remigi[1,2], G. C. Ferguson[3,*], E. McConnell[4], S. De Monte[5,6], D. W. Rogers[4], P. B. Rainey[2,4,7]

[1]INRA/CNRS, Laboratoire des Interactions Plantes Microorganismes, University of Toulouse, Toulouse, France, [2]New Zealand Institute for Advanced Study, [3]School of Natural and Computational Sciences, Massey University, Auckland, New Zealand, [4]Department of Microbial Population Biology, Max Planck Institute for Evolutionary Biology, Plön, Germany, [5]Institut de biologie de l'Ecole normale supérieure (IBENS), Ecole Normale Supérieure, Paris, France, [6]Department of Evolutionary Theory, Max Planck Institute for Evolutionary Biology, Plön, Germany, [7]Laboratoire de Génétique de l'Évolution, ESPCI Paris Tech, Paris, France

Abstract: When challenged with repeated selection cycles through two contrasting environments, the bacterium *Pseudomonas fluorescens* SBW25 evolved, *de novo*, the ability to switch repeatedly between two bistable states - one where cells secrete an extracellular colanic acid-like polymer (Cap+), and an ancestral, uncapsulated state (Cap-). Subsequent characterisation showed the switch to be epigenetic, though underpinned by mutation at the start of the pyrimidine (UTP) biosynthesis pathway. Exploration of the genotype-phenotype map through a combination of transposon mutagenesis screens, revolution and genome re-sequencing, transcriptome analysis, and gene deletion revealed a complex network of 'players' whose activities modulate Cap+/Cap- switching. Among these are the Gac/Rsm signalling pathway and various ribosome components. Several lines of evidence indicate that expression of ribosome components is up-regulated in the switchers (Cap+) compared to wild-type cells. We show that pyrimidine limitation triggers an increase in ribosome biosynthesis, with switching caused by competition between ribosomes and CsrA/RsmA proteins for the mRNA transcript of a positively autoregulated activator of colanic acid biosynthesis. We additionally show that in the ancestral bacterium the switch is part of a programme that determines stochastic entry into the semi-quiescent capsulated state, ensures that such cells are provisioned with excess ribosomes, and enables provisioned cells to exit rapidly from stationary phase under permissive conditions.

Evolutionary genetics and genomics of metabolic networks
SMBE-PO-172
Development of SNP markers for geographical prigin discrimination of mung bean
J. A. Choi*

Abstract: Tracing of the geographical origin of agricultural products is critical to providing accurate information about the agricultural product to the consumers.
Recently, the import volume of mung bean for food is being increased and some local vendors mislabeled cultivation of origin deliberately for economic purposes.
To develop scientific methods for origin discrimination of mung bean a DNA-based approach has been used.
Whole genome re-sequencing of mung bean with NGS (Next generation sequencing) technique enables us to survey DNA polymorphisms for origin discrimination.
Based on the NGS data from 24 mung bean varieties, a total of 700,000 single nucleotide polymorphisms (SNPs) and insertions/deletions (InDels) were identified.
Among the DNA variants, 380 SNP loci were selected as candidate markers.
Validation of the candidate markers has been conducted by using AS-PCR (allele-specific PCR) on the 24 mung bean varieties from 4 countries.
With NGS data, mung bean has 5 class of the phylogenetic analysis. 2 class has Korea, other 3 has different country.
The result showed that the selected 12 SNP markers can be used for the country of origin discrimination and variety identification of mung bean.
Thus the SNP analysis gives reliable information to discriminate the geographical origin of mung bean. [Key words: mung bean, NGS, SNP, DNA marker]

Evolutionary genetics and genomics of metabolic networks
SMBE-PO-170
Comprehensive transcriptome analysis provides new insights into evolutionary history of C4 in Family Molluginaceae.

L. A. Munshi[*], P.-A. Christin [1]
[1]the university of Sheffield, SHEFFIELD, United Kingdom

Abstract: C4 photosynthesis is a novel adaptation that increases plant productivity in tropical conditions. It is a complex trait requiring the coordination of many biochemical, anatomical and genetic components. Despite its complexity, the C4 trait evolved more than 60 times independently in flowering plants. However, the history of genetic modifications leading to the C4 trait remains poorly understood. The Molluginaceae family includes closely related species that span C4 photosynthesis as well as a variety of non-C4 types. Capitalizing on this system, we used comparative transcriptomics to track the changes in gene expression levels and coding sequences along the phylogenetic tree. Genes encoding core C4 enzymes are upregulated in C4 accessions, as expected. However, we show that these genes were already upregulated in some non-C4 ancestors, which likely facilitated transitions to a C4 biochemistry. By sampling multiple populations per species, we were further able to demonstrate that amino acid replacements adapting the proteins for the C4 function followed the upregulation of genes, and continued when the plants were already C4. Our comparative work therefore shows how a complex biochemical pathway can be gradually assembled through multiple rounds of changes in gene expression and coding sequences.

Key words: adaptive evolution, C4 photosynthesis, phylogenetics, transcriptomics.

Evolutionary genetics and genomics of metabolic networks

SMBE-PO-165

Using core and accessory gene co-expression networks to reveal genospecies-specific bacterial adaptations

B. Fields [1,*], M. I. A. Cavassim [2], S. Moeskjær [2], V.-P. Friman [1], J. P. W. Young [1], S. Uggerhøj Andersen [2]

[1]Department of Biology, University of York, York, United Kingdom, [2]Department of Molecular Biology and Genetics, Aarhus University, Aarhus, Denmark

Abstract: *Rhizobium leguminosarum* biovar *trifolii* (*Rlt*) bacterium forms a symbiotic relationship of agronomic importance with white clover (*Trifolium repens*). *Rlt* isolates display wide genomic variation and can be categorized into five genospecies (A-E) based on core genes. However, considerable variation exists in accessory genome content within genospecies and it is unclear to what extent this large genomic variation influences transcriptomic variation. To investigate this, we generated whole-genome transcriptomes of 30 genome-sequenced *Rlt* isolates obtained from clover nodules across Europe using RNA-Seq. Our results show that genospecies displayed significant differences in the transcription of individual core genes, with expression profiles of the 10% most variable core gene expression corresponding to genospecies clustering. Genospecies C and E displayed the most significant variation in their gene expression with 9.03% of core genes differentially expressed. These differences were mainly linked with regulating species-specific signalling pathways involved in secretion, biofilm formation and cellulose production, which could indicate differences in root and rhizosphere colonisation. Using co-expression analyses we identified core and accessory genes that were highly co-expressed within genospecies. We hypothesise that evolution and maintenance of these genospecies-specific regulatory interactions could be explained by *Rlt* strains adapting to specific environmental niches within the plant rhizosphere and by selection in symbiotic interaction with certain clover genotypes. Ongoing transcriptomics work will focus on linking the transcriptomic variation with symbiotic efficiency at the level of *Rhizobium* and clover genotypes with the aim to elucidate the importance of genospecies boundaries and symbiosis at a more mechanistic level.

Evolutionary genetics and genomics of metabolic networks
SMBE-PO-163
Functional characterization of human and chimpanzee enhancers with a massively parallel reporter assay
J. Pizzollo[*], C. Babbitt[1]
[1]Biology, University of Massachusetts Amherst, Amherst, United States

Abstract: Human and chimpanzee genomes are very similar, but our neural phenotypes are significantly different. Changes in gene expression can account for phenotypic differences between species, and adaptation in DNA regulatory elements like distal enhancers may help explain some of these differences. Although putative enhancers can be identified with molecular or bioinformatic methods, the function of these elements remains unknown. To understand how changes in enhancers can affect human neural phenotypes, we have identified putative enhancers that show signs of accelerated nucleotide substitution in humans or have epigenetic marks associated with enhancers that are specifically active during human brain development. Using these human enhancers and their chimpanzee homologs we have designed a massively parallel reporter assay to test the function of these enhancers in neurons and neural progenitor cells that we have differentiated from human and chimpanzee induced pluripotent stem cells. By expressing reporter constructs harboring these candidate enhancers in relevant cell types and using next generation sequencing to quantify enhancer-driven transcription, this study allows us to functionally characterize enhancers that may have roles in evolution of uniquely human phenotypes.

Evolutionary genetics and genomics of metabolic networks

SMBE-PO-176

Gene co-expression network analysis identifies gene candidates for shikonin biosynthesis pathway in the medicinal plant Lithospermum erythrorhizon

R. Auber [1 2,*], T. Suttiyut [2 3], J. W. Cook [2 3], J. R. Widhalm [2 3], J. Wisecaver [1 2]

[1]Biochemistry, Purdue University, [2]Purdue Center for Plant Biology, [3]Horticulture, Purdue University, West Lafayette, United States

Abstract: Specialized metabolites (SMs) are of great interest due to their unique properties, particularly in the field of medicine and in commercial use. These compounds are usually of an intricate nature that is difficult and inefficient to create via chemical synthesis. The biological machinery that assembles these metabolites are just as complex and are difficult to characterize, especially due to their high level of evolutionary divergence. One SM, shikonin, is produced by a Chinese plant by the name of *Lithospermum erythrorhizon*. Shikonin is known for its use in dyes, cosmetics, and most notably, its suppressive activity against human immunodeficiency virus (HIV). Only two genes of this SM pathway have been elucidated. One encoding a prenyltransferase, involved in the committing step of shikonin synthesis from 4-hydroxybenzoic acid (4HBA), and the other encoding a hydroxylase involved in the second reaction of the pathway.In this study, we performed whole genome sequencing coupled with gene expression analysis to construct a list of gene candidates likely to be involved in shikonin biosynthesis.Similar to many other SM pathways, shikonin production is exclusive to specific tissue types and environmental conditions. This provided an opportunity to design a series of RNA-seq experiments using samples known to have variant levels of shikonin. We hypothesized genes that coexpressed with known shikonin pathway genes across conditions with perturbed shikonin production would serve as strong candidates. We also performed a differential expression analysis as an orthogonal approach to identify genes of interest. RNA-seq samples selected for this study included four tissue types and two different culture lighting conditions. Using a hypothesized comparative expression profile and GO enrichment analysis, we were successful in recovering known pathway genes and we isolated potential novel candidates. We also investigated the genomic location of known and putative pathway genes to evaluate clustering of SM genes in this organism. In future work, gene candidates identified in this study will be functionally tested by molecular cloning in *L. erythrorhizon* hairy root cultures. In using the latter stated approach, we hope to further uncover the shikonin biosynthetic pathway and apply this powerful approach to other unexplored SM pathways.

Evolutionary genetics and genomics of metabolic networks
SMBE-PO-173
Natural selection on the level of molecular crowding in cells
T. Y. Pang*, M. Lercher

Abstract: Proteins, metabolites and other molecules solved inside cellular compartments occupy a substantial fraction of the available volume. Cells can adjust the level of this "molecular crowding" by importing or exporting water. While increased substrate concentrations in crowded cells may increase cellular efficiency by increasing biochemical fluxes, crowding may also affect cellular efficiency by slowing down diffusion and by changing Gibbs free energies. We hypothesize that maximal cellular efficiency is achieved at an intermediate crowding level. The resulting natural selection would be consistent with the observed constancy of crowding across growth conditions in *E. coli*, where solutes account for roughly half of the cytosol volume. To understand the effects of molecular crowding on cellular efficiency at the network level, we simulated the maximal growth of a simple cell at different cytosol densities. Our mathematical model accounts systematically for (1) the volume exclusion effect that changes Gibbs free energy and (2) the slowdown of diffusion caused by molecular crowding. We find that optimal crowding depends strongly on the kinetic properties of the cellular reactions. From our results, we conclude that cellular efficiency is indeed optimized at an intermediate cytosol density, but that the environment causes small variations of the optimal crowding level.

Evolutionary Insights from Metagenonics

SMBE-PO-181

M1CR0B1AL1Z3R – A user-friendly web server for the analysis of large-scale microbial genomics data

O. Avram [1], D. Rapoport [1], S. Portugez [1], T. Pupko [1,*]

[1]The School of Molecular Cell Biology & Biotechnology, Tel Aviv University, Tel Aviv, Israel

Abstract: Large-scale mining and analysis of bacterial datasets contribute to the comprehensive characterization of complex microbial dynamics within a microbiome and among different bacterial strains, e.g., during disease outbreaks. Studying large-scale bacterial evolutionary dynamics possesses many challenges. These include data-mining steps, such as gene annotation, orthologs detection, sequence alignment and phylogeny reconstruction. These steps require the use of multiple bioinformatics tools and ad-hoc programming scripts, making the entire process cumbersome, tedious, and error-prone due to manual handling. This motivated us to develop the M1CR0B1AL1Z3R (pronounced: microbializer) web server, a “one-stop shop” for conducting such microbial genomics data analyses via a simple graphical user interface. An example of features which are implemented in M1CR0BIALIZ3R web server include: (1) Extracting putative Open Reading Frames and comparative genomics analysis of gene content; (2) Extracting orthologous sets and an analysis of their size distribution; (3) Analyzing presence-absence patterns of genes; (4) Reconstructing a phylogenetic tree based on the extracted orthologous set; (5) Inferring GC content variation among lineages. M1CR0BIALIZ3R will facilitate mining and analyzing dozens of bacterial genomes using advanced techniques, with a click of a button. M1CR0BIALIZ3R is freely available at https://microbializer.tau.ac.il/.

Evolutionary Insights from Metagenonics

SMBE-PO-183

Reconstructing the health landscape of a medieval hospital cemetery: a holistic interdisciplinary approach.

C. L. Scheib [1,*], M. Guellil [1], R. Hui [2], A. W. Wohns [3], X. Ge [4], S. J. Griffith [1], J. Bates [5], P. Maheshwari-Aplin [2], B. Haines [2], S. A. Inskip [6], J. Dittmar [6], C. Cessford [7], F. Key [8], A. Rose [2], M. Metspalu [1], T. C. O'Connell [2], P. Mitchell [2], J. Krause [8], J. E. Robb [2], T. Kivisild [9]

[1]Estonian Biocentre, University of Tartu, Tartu, Estonia, [2]Department of Archaeology, University of Cambridge, Cambridge, [3]Big Data Institute, University of Oxford, Oxford, [4]Department of Biology, Medicine and Health, University of Manchester, Manchester, [5]Department of Archaeology, University of York, York, [6]McDonald Insitute, [7]Cambridge Archaeological Unit, University of Cambridge, Cambridge, United Kingdom, [8]Department of Archaeogenetics, Max Planck Institute for the Science of Human History, Jena, Germany, [9]Human Genetics, KU Leuven, Leuven, Belgium

Abstract: Established for the care of the poor and infirm, with the exception of pregnant women, lepers, the wounded, crippled and insane [1], the assemblage of 400 internments (c. 1204-1511) from the Hospital of St. John, Cambridge [2], provides a laboratory for testing osteological/aDNA methods and for understanding the health and diseases of a broadly representative British medieval population as it experienced the spread of Black Death from continental Europe during 1347 – 1349. Four hundred skeletons from the cemetery site were assessed for osteological traits and paleopathology. A subset was sampled for isotopic and ancient DNA (aDNA) analysis and compared to a time-transect of sites in the local area. DNA was extracted from teeth, built into double-stranded Illumina-compatible libraries and sequenced on the NextSeq500 75-cycle single-end platform. The raw libraries were screened in collaboration with the Jena Max Planck Institute for the Science of Human History using MALT [3]. Surprisingly, though the cemetery's prime usage overlaps the main plague period, no skeletons (n = 86, 0%) tested positive for the presence of *Y. pestis* in the teeth, while a contemporaneous site less than 500 meters away contained at least 5 individuals (n = 26, 19%) whose teeth tested positive for *Y. pestis* aDNA. In line with the charter, within the defined graves of St. John Hospital cemetery, no signs of active *M. leprae* infection (osteologically or aDNA) were found; however, osteological evidence of possible *M. tuberculosis* infection is abundant (post-cranial remains are currently being extracted and screened for pathogen aDNA) as well as a number of trauma. Ancient DNA scans do, however, find evidence of opportunistic viral and bacterial infections as well as oral microbiome signatures implicated in periodontal disease; more indicative of 'diseases of the elderly' that effect immunocompromised individuals in today's hospital settings and analyses of the human genome provide insight into heritable disease risk present in the population. By combining the aDNA with dietary isotopic information and skeletal analysis, this collection provides a valuable experiment in holistic Bioarchaeology and an opportunity to explore the relationships between diet, culture, and heritable disease risk on the discovery and expression of heritable and communicable disease in the archaeological record.

1. Rubin, M. 1987. Charity and Community in Medieval Cambridge, Cambridge: Cambridge University Press.
2. Cessford, C., The St. John's Hospital Cemetery and Environs, Cambridge: Contextualizing the Medieval Urban Dead. Archeological Journal 172 (1), 52 – 120 (2015).
3. Herbig, A., Maixner, F., Bos, K. I., Zink, A., Krause, J., Huson, D. H., MALT: Fast alignment and analysis of metagenomic DNA sequence data applied to the Tyrolean Iceman. (https://doi.org/10.1101/050559) (2016).

Evolutionary Insights from Metagenonics

SMBE-PO-182

31,000-year-old virus genomes reveal deep divergence and long-term host association of Human Adenovirus C

S. Holtsmark Nielsen [1,*], A. G. Pedersen [2], M. E. Allentoft [1], L. Vinner [1], A. Margaryan [1], E. Pavlova [3 4], V. Chasnyk [5], P. Nikolskiy [4 6], V. Pitulko [4], M. Sikora [1]

[1]Department of Biology, University of Copenhagen, [2]DTU Bioinformatics, Technical University of Denmark, Copenhagen, Denmark, [3]Polar Geography Department, Arctic & Antarctic Research Institute, [4]Palaeolithic Department, Institute for the History of Material Culture RAS, [5]St Petersburg Pediatric Medical University, St Petersburg, [6]Geological Institute, Russian Academy of Sciences, Moscow, Russian Federation

Abstract: Ancient pathogen genomics, the reconstruction of whole genomes of infectious disease agents from ancient DNA, is a transformative tool for understanding the evolution of pathogens and diseases. We performed screening for pathogen DNA in shotgun sequencing data generated from milk teeth of two Upper Paleolithic individuals from Yana RHS, an early human site in the Russian Arctic dated to ~31,000 years before present. We detected ancient DNA sequences of human adenovirus C (HAdV-C), a highly prevalent double-stranded DNA virus and significant human pathogen, responsible for respiratory infections common in children and potentially severe disease in immunocompromised individuals. After stringent ancient DNA processing and authentication, we reconstructed two genomes with ~1X and ~6X average genomic coverage respectively, the oldest human pathogen genomes sequenced to date and the first direct evidence of infectious disease in Upper Paleolithic hunter-gatherers. Phylogenetic analyses revealed that the two virus genomes belonged to different species types (HAdV-C1 and HAdV-C2, respectively), demonstrating deep divergence and co-circulation of multiple types already during the Upper Paleolithic. Our results demonstrate the power of ancient pathogen genomics to elucidate the epidemiology and evolution of diseases prevalent among early modern humans.

Evolutionary Insights from Metagenonics

SMBE-PO-180

The Significance of Robustly Identifying Microbes in Archaeological Samples of Domesticated Animals

E. A. Dimopoulos[*], E. Irving-Pease[1], I. Velsko[2], L. Frantz[3 4], G. Larson[3]

[1]RLAHA, University of Oxford, Oxford, United Kingdom, [2]Department of Archaeogenetics, Max Plank Institute for the Science of Human History, Jena, Germany, [3]PalaeoBarn, University of Oxford, Oxford, [4]QMUL, London, United Kingdom

Abstract: The accurate identification of microbial species from archaeological samples has the potential to provide an unprecedented perspective on the evolutionary history of major diseases that have affected humans and domestic animals. Limited DNA survival and DNA damage, however, have hindered the search for ancient pathogens. Here, we present a novel, hypothesis driven, method for identifying microbes in ancient DNA libraries that complements established metagenomic pipelines.

By applying this method to individual libraries constructed on extracts from 1,066 ancient pig and 1,838 ancient dog samples, we identified the causative agents of a range of bacterial and viral diseases including: Kennel cough, Rabies and Salmonellosis. We also interrogated published human metagenomic datasets and we were able to identify microbes that are capable of infecting both humans and domesticated animal hosts including the causative agents of tuberculosis and diphtheria.

This method can therefore be employed to test specific hypotheses related to the role of domesticated animal species as pathogen reservoirs. Since many of the causative agents of these zoonotic diseases became prevalent in a post-industrial context, the results of this approach will be able to offer insights into how human activity has reshaped the landscapes of both the natural environment and animal diseases.

Evolutionary Insights from Metagenonics

SMBE-PO-179

The evolution of the honey bee gut microbiota correlates with host demography and divergence

K. M. Ellegaard [1,*], S. Suenami [2], R. Miyazaki [2 3], P. Engel [1]

[1]Fundamental Microbiology, University of Lausanne, Lausanne, Switzerland, [2]The national institute of advanced industrial science and technology, [3]Computational bio big data open innovation laboratory, Tsukuba, Japan

Abstract: gghjkjhgThe European honey bee, Apis mellifera, harbors a specialized gut bacterial community composed of a modest number of species, but with high strain-level diversity. This makes the honey bee a promising model for studying strain-level evolution of microbial communities with shotgun metagenomic data. We have recently conducted a metagenomic analysis of the European honey bee gut microbiota, showing that individual bees from the same colony sampled in Switzerland harbor functionally and taxonomically distinct microbial communities when analyzed at the strain-level, despite having highly similar species-level profiles.

In a follow-up study, we have collected 40 metagenomic samples from the European honey bee and the Eastern honey bee (Apis cerana) in Japan. These two species are closely related and sympatric in the sampling area, but they differ in their overall global distribution and mean colony size. Thus, the dataset allows us to gain first insights into the evolution of the honey bee gut microbiota in response to host sympatry, migration, and colony size.

Our current results show that each host species harbor related, but clearly distinct microbial communities, with only a few cases of transfer among host species detected. Interestingly, the gut microbiota of A. cerana harbors substantially lower levels of strain diversity for all community members compared to A. mellifera, as evidenced from both depth-normalized SNV profiling, metagenomic assemblies and phylogenetic analyses on key gene families. Taken together, these results suggest that the honey bee gut microbiota evolution is shaped by both host specialization and demography.The European honey bee, *Apis mellifera*, harbors a specialized gut bacterial community composed of a modest number of species, but with high strain-level diversity. This makes the honey bee a promising model for studying strain-level evolution of microbial communities with shotgun metagenomic data. We have recently conducted a metagenomic analysis of the European honey bee gut microbiota, showing that individual bees from the same colony sampled in Switzerland harbor functionally and taxonomically distinct microbial communities when analyzed at the strain-level, despite having highly similar species-level profiles. In a follow-up study, we have collected 40 metagenomic samples from the European honey bee and the Eastern honey bee (*Apis cerana*) in Japan. These two species are closely related and sympatric in the sampling area, but they differ in their overall global distribution and mean colony size. Thus, the dataset allows us to gain first insights into the evolution of the honey bee gut microbiota in response to host sympatry, migration, and colony size. Our current results show that each host species harbor related, but clearly distinct microbial communities, with only a few cases of transfer among host species detected. Interestingly, the gut microbiota of *A. cerana* harbors substantially lower levels of strain diversity for all community members compared to A. mellifera, as evidenced from both depth-normalized SNV profiling, metagenomic assemblies and phylogenetic analyses on key gene families. Taken together, these results suggest that the evolution of the honey bee gut microbiota is shaped by both host specialization and demography.

Evolutionary Insights from Metagenonics

SMBE-PO-191

Viral screening from archeological remains uncovers an ancient Human Parvovirus of African origin associated to a Colonial epidemic in Mexico City.

A. A. Guzmán-Solís [1,*], V. Villa [1], M. Bravo [1], M. Sandoval-Velasco [2], J. K. Wesp [3], J. Gómez-Valdés [4], D. Blanco-Melo [5], M. C. Ávila-Arcos [1]

[1]International Laboratory for Human Genome Research., Universidad Nacional Autónoma de México., Querétaro., Mexico, [2]Center for GeoGenetics., Natural History Museum of Denmark., Copenhaguen, Denmark, [3]Department of Sociology and Anthropology., North Carolina State University., North Carolina., United States, [4]Escuela Nacional de Antropología e Historía., Mexico City, Mexico, [5]Department of Microbiology., Icahn School of Medicine at Mount Sinai., New York City, United States

Abstract: The introduction of viral pathogens during colonization of the Americas caused a dramatic population collapse by the multiple epidemics on the immunologically naïve indigenous inhabitants. The development of next generation sequencing (NGS), allows the retrieval of whole genomes from ancient pathogens as molecular fossils. In order to understand the role of pathogens introduced upon colonization, we extracted and sequenced ancient DNA (aDNA) from archaeological skeletal remains from a Colonial hospital in Mexico City, with evidence of having experienced a mass epidemic. We obtained aDNA from teeth and performed NGS to screen for DNA viruses in 21 individuals. Furthermore, to increase the yield of viral DNA, we designed an in-solution capture-enrichment assay targeting the genomes of 28 viral pathogens of clinical importance. With this assay we were able to enrich viral DNA and reconstruct whole human Parvovirus genomes, closely related to the B19V modern references, from the remains of two individuals (16th c.-18th c.). We further corroborated the ancient origin of the viral DNA by the presence of deamination damage at the terminal bases of the reads. Of notice, one of the individuals displayed skeletal indicators associated with anemia, such as porotic hyperostosis on the cranial vault and cribra orbitalia, which could be caused by parvovirus infection, as the virus infects precursor of the erythroid lineage and has been found in individuals with severe anemia. When assessing its similarity to modern strains, we found these genomes are more related to African sequences (genotype 3), in agreement with the inferred genetic ancestry of the human hosts from which the samples were retrieved. To our knowledge, these are the first molecular evidences of a virus brought to the Americas likely during the transatlantic slave trade, the first genotype 3 parvoviruses recovered from archeological remains, as well as the first report of this genotype in Mexico.

Evolutionary Insights from Metagenonics
SMBE-PO-184
The genomic basis of three common aquatic bacterial lifestyles
M. L. Schmidt*, J. T. Evans, V. J. Denef

Abstract: Aquatic systems harbor two distinct, dynamically interacting habitats: particulate matter and the aqueous matrix surrounding and interconnecting these particles. Organic particles form hotspots of microbial activity and interactions, and often harbor cell densities exceeding those in the surrounding water. The surrounding water is typically more homogeneous and is marked by lower substrate and nutrient concentrations. Here, using a genome-centric metagenomic approach, we tested whether or not bacteria inhabiting these two aquatic habitats had key genomic features and genetic traits that underpinned habitat specialization. We constructed hundreds of metagenome-assembled genomes from eight particle-associated and eight free-living samples from an ecologically important freshwater estuarine lake, Muskegon Lake, in Michigan USA. Next, genomes were grouped based on differential abundance as particle-associated specialist, free-living specialist, or generalist. We found several key differences between particle-associated and free-living specialist and generalist bacteria: free-living specialists (and to a lesser extent, generalists) harbor signatures of genome streamlining. Whereas the genomes of free-living specialists had lower GC contents, there was a bimodal distribution within the generalists. In contrast, particle-associated bacteria had large genomes, lower percent coding region of the genome, higher GC content, and larger nitrogen requirements for coding genes. The average ratio of carbon-to-nitrogen use in particle-associated specialists was lower, suggesting that particle-associated bacteria require more nitrogen for translation. Moreover, while particle-associated bacteria possess more unique genes, which may allow them to take advantage of variable environments, free-living bacteria “pack more” genes per region of the genome, a feature that is consistent with living in nutrient poor conditions.

Evolutionary Insights from Metagenonics

SMBE-PO-188

Exploiting sedaDNA to trace the impact of the Storegga slide tsunami on the Doggerland paleolandscape

R. Ware*, B. Cribdon, R. Everett, M. Bates, R. Bates, S. Fitch, B. Gearey, T. Hill, T. Kinnaird, P. Murgatroyd, D. Smith, V. Gaffney, R. Allaby

Abstract: Metagenomic sequencing of ancient sediment DNA (sedaDNA) is a valuable tool for reconstructing the dynamics of palaeoenvironmental change.

Doggerland was a fertile land-mass in what is now the southern North Sea. This submerged landscape retains valuable evidence about local Mesolithic communities, prior to its submersion and subsequent excision of Britain from Europe. This submersion is considered a consequence of sea level rises in the mid-Holocene, with the catastrophic tsunami triggered by the Storegga landslides speculated to have played a pivotal role. Whilst there is extensive evidence for this tsunami observed across the northern areas of the North Sea, there is little of evidence from the south, despite this area having been predicted to within the expected range of the impact of the tsunami.

Sediment cores were taken from across Doggerland, guided by extensive seismic mapping. Evidence based on established methods of traditional paleoenvironmental analysis, geochemical analysis, coupled with sedaDNA analysis, has enabled the first identification of tsunami deposits. This is supported by dating (radiocarbon and OSL) which places these deposits as contemporary to the Storegga slide. Sequencing of sedaDNA identified microbial signatures of sulphur metabolism within the tsunami deposits. We also identified sweeping shifts in plant communities, with an influx of incongruous plant taxa within the deposit, consistent with the catastrophic effects of a tsunami.

We will discuss how these data enable us to identify potential tsunami deposits, assess the environmental impacts, and establish how the morphology of the landscape impacted the tsunami's progression.

Evolutionary Insights from Metagenonics

SMBE-PO-187

Comparative genomic analysis of Yersinia pestis and Yersinia pseudotuberculosis and a de novo assembly of a second pandemic plague genome provide insights into the evolution of Yersinia pestis

A. Andrades Valtueña [1,*], M. A. Spyrou [1], E. A. Nelson [1], N. Carty [2], R. Hartle [2], M. Henderson [2], E. L. Knox [2], D. Walker [2], K. I. Bos [1], A. Herbig [1]

[1]Archaeogenetics, Max Planck Institute For The Science Of Human History, Jena, Germany, [2]Museum of London Archaeology (MOLA), London, United Kingdom

Abstract: *Yersinia pestis* (*Y. pestis*), the causative agent of plague, has afflicted human populations as early as the Neolithic. Through the accumulation of comparatively few known genetic changes, it evolved from the less virulent environmental pathogen *Y. pseudotuberculosis*. Previous research on *Y. pestis* evolution has largely focused on virulence factors where the gain of plasmids (pPCP1 and pMT) and genes (e.g. *pla* and *ymt*), combined with the loss of genes related to mobility and downregulators of biofilm production shaped its emergence as a highly virulent pathogen. While the genus *Yersinia* is characterised by high genome variability, few studies have explored plasticity of the pan and core genome of *Y. pestis*.

Here we investigate the genomic evolution of *Y. pestis* by combining the analysis of an ancient *Y. pestis de novo* assembled genome with a phylogeny-wide comparative analysis of gene content. Despite the challenges common to ancient DNA data such as low and uneven genomic coverage and short fragment length, we were able to assemble an exceptionally well-preserved *Y. pestis* genome from a 17th century plague epidemic in London (UK). This *de novo* reconstructed genome allowed us to explore the genome architecture of an extinct *Y. pestis* during the second pandemic. This was not possible from mapped-reference based genome reconstructions available from previous ancient data. We additionally queried a *Y. pestis*-*Y. pseudotuberculosis* pan-genome to detect differences between species and subspecies clades in terms of gene content, functional gene groups, and pseudogenization of core genes. Integrating these analyses, we attempt to explore hidden patterns of parallel evolution across the entire *Y. pestis* phylogeny on a temporal scale that could inform on changes in virulence. Furthermore, we integrate more than 50 ancient genomes such that genome evolution can be considered in three time transects: 1) Late Neolithic and Early Bronze Age (LNBA, around 5,000-3,700 years ago), 2) the “first pandemic” (Justinianic plague, 5th-8th century) and the “second pandemic” (14th-18th century). These temporal categories permitted an evaluation of changes in genomic content over time, with special interest given to deletions of genes with similar functions that characterise these distinct clades. Utilising ancient data in a novel way, we are able to investigate *Y. pestis* evolution on a much deeper and under-explored resolution in both time and space.

Evolutionary Insights from Metagenonics
SMBE-PO-190
Metagenomic sedaDNA evidence for the Storegga Slide tsunami
B. Cribdon [1,*], R. Ware [1], R. Everett [1], M. Bates [2], R. Bates [3], S. Fitch [4], B. Geary [5], T. Hill [6], T. Kinnaird [3], P. Murgatroyd [4], D. Smith [7], V. Gaffney [4], R. Allaby [1]
[1]University of Warwick, Coventry, [2]University of Wales Trinity Saint David, Lampeter, [3]University of St Andrews, St Andrews, [4]University of Bradford, Bradford, United Kingdom, [5]University College Cork, Cork, Ireland, [6]Natural History Museum, London, [7]University of Birmingham, Birmingham, United Kingdom

Abstract: Sedimentary ancient DNA (sedaDNA) is increasingly being used for palaeoenvironmental reconstruction despite numerous challenges. We study sedaDNA as part of a multidisciplinary project reconstructing Doggerland, a palaeolandscape now submerged under the North Sea. A likely contributor to the inundation of Doggerland was the Storegga Slide tsunami. We have identified deposits consistent with the tsunami in sediment cores.

SedaDNA from these deposits contained unexpected taxa, including plants alien to the local environment and sulphur-metabolising bacteria associated with corresponding sulphur spikes. This is consistent with the tsunami hypothesis and with data from traditional palaeoenvironmental proxies. However, deposits dated shortly after the event continue to show a clear terrestrial signal, suggesting that the tsunami did not contribute significantly to inundation in this area.

This work addresses two key challenges of ancient metagenomics: taxon assignation and age authentication. Taxon assignation, notoriously difficult with shotgun-sequenced metagenomic data, requires extra stringency to filter false positives, especially assignments to model organisms. Therefore, we filtered standard BLAST results using our Phylogenetic Intersection Analysis (PIA). This confidently assigns reads down to genus level with >81% accuracy. Age authentication was based on DNA damage patterns, particularly terminal or near-terminal cytosine deamination lesions, modelled using the thermal ages of the samples.

Overall, the analysis presents convincing evidence of the Storegga slide tsunami in southern Doggerland, but suggests that it may not have been a significant contributor to inundation. The PIA and age authentication could be applied in a variety of contexts, increasing the reliability of sedaDNA in future studies.

Evolutionary processes and consequences of animal and plant domestication
SMBE-PO-199

Impact of demography and admixture on domesticated cole crops (Brassica oleracea)

S. Turner-Hissong [1,*], M. Mabry [2], J. C. Pires [2], T. Beissinger [3], J. Ross-Ibarra [1]
[1]Department of Plant Sciences, University of California-Davis, Davis, CA, [2]Division of Biological Sciences, Bond Life Sciences Center, University of Missouri, Columbia, MO, United States, [3]Department of Crop Science & Center for Integrated Breeding Research, Georg-August-Universität, Göttingen, Germany

Abstract: The vegetables comprising the plant species *Brassica oleracea* are valued for their flavor, culinary, and nutritional attributes. These crops present an especially interesting model of domestication since the wild progenitor(s) were selected to enrich different plant organs, producing distinct morphotypes like kale, cabbage, Brussels sprouts, kohlrabi, broccoli, and cauliflower. Despite widespread use of these crops to demonstrate the power of human selection on crop domestication, the influence of domestication on *B. oleracea* remains poorly understood. We present a fine-scale analysis of diversity, admixture, and demography in *B. oleracea* using published resequencing data from 119 accessions. Findings suggest population size contraction in cabbage, broccoli, cauliflower, and kohlrabi, which is consistent with recent selection and improvement of these morphotypes compared to kale. Further, variation in demographic trajectories across morphotypes indicates the possibility of multiple, independent origins. This work expands our understanding of how humans have influenced the domestication of these crops, provides insight into Brassica evolution and, by detailing patterns of recent changes in population size, facilitates future crop improvement efforts.

Evolutionary processes and consequences of animal and plant domestication

SMBE-PO-193

Canine transmissible venereal tumor genome reveals ancient introgression from coyotes to pre-contact dogs in North America

X. Wang [1], G.-D. Wang[*]

[1]Kunming Institute of Zoology, Kunming, China

Abstract: Canine transmissible venereal tumor (CTVT), the oldest known somatic cell line, is a living fossil, originating from cancer cells transmitted from a host to other canids during the mating process. Recent studies revealed that CTVT originated from pre-contact dogs in North America, but the genetic ancestry of the ancient founder of CTVT is still unknown. We used monomorphic sites in 5 CTVTs globally dispersed to decipher the genetic ancestry of the CTVT founder, comparing with a high-quality canine reference panel. Our results show that the CTVT founder did originate from pre-contact dogs in North America, but the founder also possessed introgressed regions from another species, the coyote. We also establish that monomorphic sites belonging to CTVT can be used to study the early stages of CTVT somatic evolution, and the evolution of dogs as a living fossil.

Evolutionary processes and consequences of animal and plant domestication
SMBE-PO-200

Runs of homozygosity, recessive disease genotypes, and inbreeding depression in domestic dogs.

A. J. Sams [1,*], E. T. Chu [2], A. R. Boyko [1]
[1]Embark Veterinary, Inc., Ithaca, [2]Embark Veterinary, Inc., Seattle, United States

Abstract: Inbreeding leaves distinct genomic traces, most notably long genomic tracts that are identical by descent and completely homozygous. These runs of homozygosity (ROH) can contribute to inbreeding depression if they contain deleterious variants that are fully or partially recessive. Several lines of evidence have been used to show that long (> 5 megabase) ROH are disproportionately likely to harbor deleterious variation, but the extent to which long vs. short tracts contribute to autozygosity at loci known to be deleterious and recessive has not been studied.
In domestic dogs, nearly 200 mutations are known to cause recessive diseases, most of which can be efficiently assayed using SNP arrays. Here, we examine genome-wide data from over 200,000 markers, including 150 recessive disease variants and built high-resolution ROH density maps for nearly 2,500 dogs, recording ROH down to 500 kilobases. Additionally, we utilized reproductive fitness-related phenotype data from the Morris Animal Foundation's Golden Retriever Lifetime Study and Doberman Pinscher longevity data from the Doberman Diversity Project.

We observed over 678 homozygous deleterious recessive genotypes in the panel across 29 loci, 90% of which overlapped with ROH inferred by GERMLINE. Although most of these genotypes were contained in ROH over 5 Mb in length, 14% were contained in short (0.5 - 2.5 megabase) tracts, a significant enrichment compared to the genetic background, suggesting that even short tracts are useful for computing inbreeding metrics like the coefficient of inbreeding estimated from ROH (F_{ROH}). In our dataset, FROH differed significantly both within and among dog breeds. Finally, we utilize these F_{ROH} estimates to examine the extent of inbreeding depression in several common dog breeds, including Doberman Pinscher and Golden Retriever and find that both lifespan and fecundity are substantially associated with increases in F_{ROH} in these breeds.

Our results suggest that even short (> 0.5 Mb) ROH should be considered when calculating the coefficient of inbreeding in domestic dogs and that elevated FROH values lead to a notable reduction of fitness in domestic dog breeds.

Evolutionary processes and consequences of animal and plant domestication
SMBE-PO-198
Ancient genomic insights into the goat herds from the earliest phases of domestication.
K. G. Daly[*], D. G. Bradley[1]
[1]Smurfitt Institute of Genetics, Trinity College Dublin, Dublin, Ireland

Abstract: As one the first animals domesticated, the domestic goat (*Capra hircus*) has partook in a mutualistic relationship with humans spanning the last 10,000 years. Previous zooarchaeological and ancient DNA analyses have shed light on the dynamics of this long history, but comparatively little is known about the goat populations involved in the initial stages of domestication. We present mitochondrial and nuclear genome data from early Neolithic archaeological sites, including those in the Zagros Mountains of Iran with some of the earliest evidence of goat management. We investigate how these relate to later and modern goat groups, the population dynamics during the initial spread of goat herding, and the role played by gene flow from distinct wild populations in shaping modern goat genetic diversity.

From the rings of life to the new animal phylogeny: James A. Lakes's contributions to evolutionary biology
SMBE-PO-209

Molecular palaeobiology of Ecdysozoa

M. Giacomelli [1,*], R. Howard [2], J. Lozano Fernandez [3], M. Wills [4], J. Vinther [5], G. D. Edgecombe, P. Donoghue [5], D. Pisani [1]
[1]School of Biological Sciences, University of Bristol, Bristol, [2]Department of Earth Science, The Natural History museum, London, United Kingdom, [3]Department of Evolutionary Biology, Ecology and Environmental Sciences, University of Barcelona, Barcelona, Spain, [4]Department of Biology and Biochemistry, The University of Bath, Bath, [5]School of Earth Sciences, University of Bristol, Bristol, United Kingdom

Abstract: Twenty years after its proposal by Aguinaldo, Lake and colleagues, the Ecdysozoa hypothesis is now well accepted among evolutionary biologists. This superclade unites eight phyla of protostome animals: Priapulida, Kinorhyncha, Loricifera (Scalidophora), Nematoda, Nematomorpha (Nematoida) Tardigrada, Onychophora and Euarthropoda (Panarthropoda), which together constitute the overwhelming majority of extant Metazoa, either in terms of diversity and abundance. However, the evolutionary relationships among these phyla remain controversial, with conflicts both between and within morphological and molecular hypotheses. In particular the clade Cycloneuralia, a group uniting Scalidophora and Nematoida, remains in use despite generally lacking molecular support. Furthermore, the timing of ecdysozoan lineages divergences in geological time is under debate, with molecular clock analyses recovering a precambrian origin (619-546 Ma) and thus predating the first fossils unequivocally attributed to this group, which are not identifiable until early Cambrian strata. To address these obstacles, we assembled a large phylogenomic dataset (up to 228 genes, 66 ecdysozoan and 14 outgroup taxa) adding data from ecdysozoan taxa with poor molecular data. We then applied Bayesian methods to infer the phylogeny and divergence times of ecdysozoan lineages. Our phylogenetic analysis support a sister group relationship between nematoids and panarthropods (rendering thus Cycloneuralia a paraphyletic clade). Moreover, our fossil-calibrated relaxed molecular clock analyses, supported by several senistivity tests, corroborate a Neoproterozoic origin of Ecdysozoa (609-582 Ma), followed by its divergence into the three main ecdysozoan sublades: Scalidophora, 601-573 Ma, Nematoida, 563-478 Ma, and Panarthropoda 600-574 Ma. Furthermore, we recovered that crown Priapulida originated 562-489 Ma and Euarthropoda 567-544 Ma. Our study shows that the lag between the genetic isolation of Ecdysozoa and its clades and their representation in the fossil record is not as considerable as has been previously purported.

From the rings of life to the new animal phylogeny: James A. Lakes's contributions to evolutionary biology
SMBE-PO-201

Odontotaenius disjunctus microbiome structure is consistent with the host geographic distribution

A. Waldrop [1], J. Sinsheimer [2], M. C. Rivera [3,*]
[1]RTI International, Research Triangle Park, North Carolina, [2]Human Genetics, University of California, Los Angeles, Los Angeles, CA, [3]Biology, Virginia Commonwealth University, Richmond, VA, United States

Abstract: The digestive tracts of most animals host complex, dynamic, and specialized microbial communities or microbiomes that contribute to enhance the fitness of the host. Despite the increasingly recognized importance of microbiomes to a growing number of host species, the mechanisms underlying their evolution and ecology are not well understood.

Odontotaenius disjunctus ("bess beetle") is a large (up to 3.5 cm), sub-social (brood caring) wood-feeding beetle widely found throughout eastern North America. *O. disjunctus* is able to survive on a diet consisting solely of decaying coarse wood, primarily oak. *O. disjunctus* lives in family units and practices parental care of the offspring, by which the adult beetles provide predigested wood and nutrients for the subsistence of the larvae. Evidence suggests *O. disjunctus* relies upon a specialized gut microbiome to supplement its nutrient-poor diet and that parental care behavior may have evolved as a mechanism to maintain and transmit these host-microbiome partnerships. Due to its unique life history characteristics and putatively symbiotic gut microbiome, *O. disjunctus* provides an intriguing natural system in which to explore the evolutionary and ecological processes that shape the composition of its microbiome.

The aim of our work is to assess the variability of the *O. disjunctus* microbiomes, in both species assemblages and abundance profile, and to uncover the environmental, geographic, and other processes and factors influencing the observed variability. To understand these underlying processes, we characterized the microbiomes of over 200 *O. disjunctus* individuals, sampled from 23 populations across the species' range. Statistical analysis of the environmental and geographical covariates suggests the composition and variability of the *O. disjunctus* microbiome diverges as the distances between the hosts' increases. It is unclear whether the observed compositional variability of the *O. disjunctus* microbiome is due to differences in individual host genetics, the historical climate shifting of the region, the differences in environment and/or to the demographic history of the studied populations.

From the rings of life to the new animal phylogeny: James A. Lakes's contributions to evolutionary biology
SMBE-PO-202

Using molecular phylogenetics to link microbial communities to ecological function in the underground aquifer

J. R. Garey[*], R. J. Scharping[1], M. C. Davis[1]
[1]Cell Biology, Microbiology and Molecular Biology, University of South Florida, Tampa, United States

Abstract: My laboratory has been studying microbial communities in the coastal Floridan Aquifer for the past decade. The Floridan Aquifer is about 260,000 square km and extends through Florida, Alabama, Georgia, Mississippi and South Carolina. Because of growing human population in this region, there is a growing demand on the aquifer to provide freshwater, especially along the coast where it causes deeper salt water to intrude into the fresh water layers. We access the aquifer through surface karst features such as springs and sinkholes via scuba diving and also collect samples from test wells maintained by local government agencies. The Floridan Aquifer is composed of a number of water layers, separated by confining layers of clay. The shallower layers are freshwater, while the deeper parts are saline and typically high in sulfur compounds. Using Illumina-based 16S rDNA sequencing along with hydrochemical and hydrological analyses we have identified numerous microbial communities with diverse functions that are affected by salt-water intrusion caused by overuse of the aquifer. At each sampling location, the microbial communities within different waters layers have different functions, typically centered around sulfur reduction/oxidation, methane cycling, or iron oxidation to fix inorganic carbon into organic carbon. We find that the microbial communities differ greatly at each site, but the functions remain similar. Our results indicate that the deeper saline regions of the Floridan aquifer produce significant amounts of organic carbon that is discharged through coastal karst estuaries.

From the rings of life to the new animal phylogeny: James A. Lakes's contributions to evolutionary biology
SMBE-PO-204

Addressing tree reconstruction artifacts in animal phylogenetics

D. Pisani*

Abstract: Undertanding animal phylogenetics has proven difficult. Howerver, animal phylogenetics has also long been a preferred playground for theoretical phylogeneticists. Critical to the central role of animal phylogenetics as a model to understand phylogenetics more broadly is the fact that animal relationships are generally well understood, but for a relatively small number of key nodes that proved hard to resolve. This allowed the development of an essentially experimental approach to animal phylogenetics. Perhaps the best example of the role plaied by the study of animals to the development of phylogenetics, is found in the Ecdysozoa versus Coelomata debate, now resolved in favour of Ecdysozoa. More recently, a new, similarly iconic problem has emerged in the debate about the relative relationships of the comb jellies. Here, I will present analyses aiming to illustrate how an "experimental approach", were multiple analyses are performed to test specific hypotheses, can be taken to address animal phylogenetics and clarify early animal evolution.

From the rings of life to the new animal phylogeny: James A. Lakes's contributions to evolutionary biology
SMBE-PO-205

Comparison of finite and infinite mixture models for capturing compositional heterogeneity across sites

T. J. Bujaki*, N. Rodrigue [1]
[1]Biology, Carleton University, Ottawa, Canada

Abstract: Phylogenetic modelling of the variation of the evolutionary process across sites from multi-species sequence alignments has garnered increasing attention over the last few decades. One of the main approaches, sometimes known as random effects modelling, adopts the view that the heterogeneity across observations is a result of the data set having been emitted from several different models, each drawn from a distribution. When little is known about the form of the across-site heterogeneity, finite mixture models provide discretizations of the unknown distribution into a pre-determined set of sub-models, or components. Choosing a level of discretization that is sufficiently fine-meshed to reflect the underlying heterogeneity is typically done from a set of likelihood-based model comparisons using different numbers of components. In the infinite mixture framework, accounting for the uncertainty regarding the number of components is another layer built into the model formulation (i.e., a hierarchical modelling framework), providing a rich non-parametric fitting of the distribution of across-site heterogeneity.
Here, we use Bayesian cross-validation to compare a wide range of finite mixture models, along with the infinite mixture modelling approach known as 'CAT', and gamma-distributed rates-across-sites approach. We apply the comparison to both simulated and real multi-gene alignments. Our findings indicate that the potential improvement in model-fit, as reported by cross-validation score, from finite mixture models is attained when the number of components of the mixture is between 20 and 60. In all cases that we considered, the fit of the CAT-GTR+Γ model matched or exceeded the best-fitting finite mixture model. When testing these models on data sets prone to long branch attraction our findings indicate that finite mixture models with low numbers of components result in topologies that align with those found from analyses using simple models. Finite mixture models, with many components at their disposal, on the other hand, produce topologies that match those obtained with the infinite mixture model CAT-GTR.

From the rings of life to the new animal phylogeny: James A. Lakes's contributions to evolutionary biology
SMBE-PO-206

What do endosymbiosis, preDNA life, complexity have in common?

J. Sinsheimer [1,*], M. Rivera [2]
[1]Human Genetics, University of California, Los Angeles, Los Angeles, [2]Biology, Virginia Commonwealth University, Richmond, United States

Abstract: Coming from a background in Physics, Prof. James A. Lake, has tackled questions in evolutionary biology in unconventional ways, which has led to some rather surprising insights. In this talk, we briefly survey his work. We then consider some of his most important contributions in the light of their underlying statistical and mathematical models.

From the rings of life to the new animal phylogeny: James A. Lakes's contributions to evolutionary biology
SMBE-PO-208

Illuminating the twilight zone with phylogenetic analysis of protein structures

M. S. Fullmer [1,*], A. J. Malik [2], J. R. Allison [1], A. M. Poole [1]
[1]School of Biological Sciences, University of Auckland, Auckland, New Zealand, [2]A*Star Agency for Science, Singapore, Singapore

Abstract: Over large evolutionary distances sequences can accumulate so many changes that evolutionary analysis becomes challenging. In some instances this sequence change is so great that homologous sequences are no longer recognizably related. In this "twilight zone" the three-dimensional structures of proteins may nevertheless retain clear similarities. This grants the opportunity to plumb greater evolutionary depths and ask broader questions about relationships by comparing structure instead of sequence[1]. Prior work has demonstrated that the evolutionary relationships within a protein superfamily can be recovered from structural analyses[2]. While credible phylogenies can be inferred for well-characterized protein groups using structural comparisons there has been no method to access the robustness of the data to the topology. We present a novel method of generating a parametric support set of protein topologies using molecular dynamics simulations. This method allows assessment for the reliability of all branches of the inferred best topology. We also present our progress in using structural phylogenetics for uncovering the evolutionary history of protein superfamilies. These developments offer the potential to further develop and refine the organization of protein repositories such as CATH and SCOP, which group proteins hierarchically into, e.g. superfamilies, but do not assess the evolutionary relationships between superfamily members.

1. Lake, J.A., Henderson, E., Oakes, M. & Clark, M.W. Eocytes: a new ribosome structure indicates a kingdom with a close relationship to eukaryotes. *PNAS* (1984)
2. Lundin, D., Poole, A.M., Sjoberg, B.M. & Hogbom, M. Use of structural phylogenetic networks for classification of the ferritin-like superfamily. *J.Biol.Chem.* (2012)

Genetic conflicts in molecular evolution

SMBE-PO-214

Testing the functional conservation of the rapidly evolving germline stem cell gene, bag of marbles, in diverse Drosophila

J. Bubnell [1,*], C. F. Aquadro [1]

[1]Cornell University, Ithaca, United States

Abstract: Although germline stem cell maintenance and differentiation are conserved processes across *Drosophila*, many of the genes involved show lineage specific patterns of adaptive evolution. One of these genes, *bag of marbles* (*bam*) initiates gametogenesis by promoting germline stem cell differentiation. We are currently testing the hypothesis that *bam* has acquired new functions in lineages with bursts of positive selection. We are using CRISPR-Cas9 to generate *bam* null alleles in diverse *Drosophila* species and evaluating germline stem cell function. We have found that *bam's* core role in gametogenesis is conserved in lineages with signals of adaptive evolution in the melanogaster subgroup - *D.melanogaster, D. simulans,* and *D. yakuba.* However, in the outgroup species *D. ananassae* where *bam* is not evolving under positive selection*,* function is only conserved in females. In males, *bam* is not necessary for germline stem cell differentiation or fertility. This suggests that either *bam's* role in spermatogenesis is novel in the melanogaster subgroup or that *bam* has lost this function in the *ananassae* lineage. However, we observe lineages with and without signals of adaptive evolution at *bam* in the melanogaster subgroup, so a novel role in spermatogenesis is not likely the only factor driving bam's evolution. If *bam* function is conserved between species with different patterns of evolution, the episodic signals of positive selection imply adaptation to lineage specific factors. A prime candidate we are evaluating is the bacteria *Wolbachia* that transiently infects *Drosophila*, manipulates reproduction and genetically interacts with *bam.*

Genetic conflicts in molecular evolution

SMBE-PO-215

The dependence of homologous recombination rate on the level of heterozygosity in hypervariable fungus Schizophyllum commune

A. V. Bezmenova [1,*], E. A. Zvyagina [2], A. Fedotova [3], M. D. Logacheva [4], G. A. Bazykin [1], A. S. Kondrashov [5]

[1]Center of Life Science, Skolkovo Institute of Science and Technology, Moscow, [2]Surgut State University, Surgut, [3]Lomonosov Moscow State University, [4]Skolkovo Institute of Science and Technology, Moscow, Russian Federation, [5]University of Michigan, Michigan, United States

Abstract: Basidiomycete *Schizophyllum commune* is a unique organism with the highest known genetic diversity that can reach 20% in a population. Thus, *S. commune* is a very promising object to study different evolutionary processes with resolution previously unachievable. In particular, one can study homologous recombination with great accuracy as the points of recombination can be detected with high resolution. In the light of a very high polymorphism of this species, it is not quite clear how homologous recombination may act, as homologous chromosomes are still very different from each other compared to other species. It was previously shown that in *S. commune* crossing-over events preferentially occur in more conservative regions – in particular, in genes. We have developed an experimental system that allows us to directly study the dependence of the homologous recombination rate on the level of heterozygosity of a chromosome region. We perform back crosses of F1 offsprings that carry chromosomes with shoulders came from one parent and central part came from another parent, with both parents. By shoulders genotyping of backcross progeny we were able to estimate the number of crossing-over events in the central part of chromosome when this region was heterozygous (when crossing with one parent) and also homozygous (when crossing with another parent).

Genetic conflicts in molecular evolution
SMBE-PO-211
Domestication of a large B chromosome in the tilapia genome
M. A. Conte [1,*], F. E. Clark [1], T. D. Kocher [1]
[1]Department of Biology, University of Maryland, College Park, United States

Abstract: B chromosomes are selfish genetic elements that are found in roughly 15% of eukaryotes. They are dispensable for normal growth and exist in addition to the normal set of "A" chromosomes. They are highly repetitive, heterochromatic, and evolve various drive mechanisms to ensure their selfish propagation.

B chromosomes in the cichlid fish of lakes Victoria and Malawi have been characterized by cytogenetics and genomics. B chromosomes in Lake Victoria cichlids vary from 0-3 copies per individual and are present in high frequency (~85% of individuals). In at least one Lake Victorian species, increasing numbers of B chromosomes bias the sex ratio toward females. In Lake Malawi, B chromosomes are found less frequently (6% of individuals) and in only 1 copy. Lake Malawi B chromosomes are also only found in females, where the B acts as a WO sex determination system.

Cytogenetic studies of the cichlid tribe Oreochromini, including the Nile tilapia (*Oreochromis niloticus*) revealed a karyotype of 22 chromosome pairs, including one large (~70Mbp) and highly repetitive chromosome, referred to as linkage group 3 (LG3). Most other African cichlid species do not harbor a similarly large LG3. This chromosome has more than doubled in size in the last 20 million years. Unlike the African lake cichlids, no B chromosomes have been reported in the Oreochromini tribe.

We hypothesize that this size of LG3 in the Oreochromini is the result of the fusion of a B chromosome with an ancestral autosome. This fusion may have resolved genetic conflicts between the B chromosome and the A genome. We analyze tilapia LG3 in a comparative genomic framework and characterize this domesticated B chromosome using several new chromosome-scale genome assemblies and the genomes of many African cichlids both within and outside of the Oreochromini tribe. The large LG3 in tilapia has characteristics similar to other cichlid B chromosomes and little sequence similarity to other cichlid chromosomes.

Genetic conflicts in molecular evolution

SMBE-PO-234

Testis specific genes show signs of positive selection in Great Apes with an enrichment for X-linked genes

M. Riera Bellés [1,*], S. Soraggi [1], M. H. Schierup [1]

[1]Aarhus University, Bioinformatics Research Center, Aarhus C, Denmark

Abstract: Compared to mice, human males produce markedly fewer mature spermatozoa, a low percentage of which appear to be fully functional. Nevertheless, genes involved in male reproduction show evidences of positive selection after the human and mouse split. We have sought to investigate if testis specific genes show signs of positive selection in Great Apes. Using PAML, we have performed likelihood ratio tests to assess whether a model that allows for dN/dS > 1 fits the data significantly better than a null model with dN/dS ≤ 1. Out of 865 genes, we found 211 genes with site-specific and/or branch-specific dN/dS > 1. A larger proportion of X-lined testis expressed genes are under positive selection compared to autosomal genes. This faster-X effect may be due to X hemizygosity in males, assuming that most new mutations are recessive. Alternatively, the unique gene content of the X chromosome could also contribute to its evolutionary potential. The X chromosome has been reported to harbor megabase-sized regions of low diversity that show overlap between Great Apes and are devoid of archaic admixture. These regions are enriched in ampliconic genes, testis expressed small multicopy adjacent genes that show high similarity between copies. These genes have been shown to present a rapid evolutionary turnover and have been suggested to play a role in meiotic drive, which could explain both the evidences of recurrent selective sweeps and the potential involvement of these regions in speciation as a result of hybrid incompatibilities. If genes on the X chromosome were to have a role in meiotic drive, we would expect them to be expressed after meiosis, where they could favor the transmission of a given sex chromosome. However, as a result of asynapsis during the pachytene, sex chromosomes are silenced during meiosis and appear to continue repressed after meiosis, where spermatids show overall low expression. Nevertheless, some multi-copy genes, miRNA genes and ampliconic genes have been shown to escape this repressed state in spermatids. The study of gene expression in spermatogenesis has been hindered by the complexity of the tissue, involving all spermatogenic cell types in close proximity, together with somatic cells. However, with the advent of single cell RNA sequencing, we can better characterize cell type specific expression and potential differences between cell populations that belong to the same stage. We are now studying single cell sequencing data of spermatogenesis to investigate whether the enrichment in testis-specific X-linked genes with signs of positive selection in Great Apes can be explained by the functional particularities of this chromosome acting in spermatogenesis and will report results from these investigations.

Genetic conflicts in molecular evolution

SMBE-PO-213

Rapidly evolving fertilization genes in threespine stickleback reproductive isolation and speciation

E. Killingbeck[*], D. Wilburn [1], G. Merrihew [1], M. MacCoss [1], C. Peichel [2], W. Swanson [1]

[1]Department Of Genome Sciences, University Of Washington, Seattle, United States, [2]Institute of Ecology and Evolution, University of Bern, Bern, Switzerland

Abstract: In species-specific mating, gamete recognition ensures a single sperm fertilizes the egg while avoiding deleterious polyspermy. Because gamete compatibility must be maintained for the fitness of sexually reproducing organisms, interacting male and female proteins might be expected to be highly evolutionarily conserved – instead, however, these genes are generally among the most rapidly evolving in any taxa. This juxtaposition of rapid evolution and functional constraint is likely driven by coevolution between interacting partners to maintain compatibility. The contribution of rapidly evolving reproductive proteins to reproductive isolation and speciation is not well understood, and we seek to investigate this question in threespine stickleback fish (*Gasterosteus aculeatus*). An adaptive radiation of ancestral marine stickleback into freshwater environments throughout the Northern Hemisphere has resulted in phenotypically divergent forms characterized by varying degrees of reproductive isolation. The role of sperm-egg recognition proteins in this isolation remains unknown, as the consensus in the field has been that behavioral and ecological factors underlie stickleback population divergence. To examine the contribution of reproductive proteins to stickleback reproductive isolation, we have characterized the proteome of stickleback egg coats, the first barrier to sperm during fertilization. We find that they are comprised of homologs to mammalian zona pellucida (ZP) proteins, and evolutionary rate analysis across teleost fish indicates positive selection. In many cases fish ZP homologs are tandemly duplicated in the genome, including in stickleback. Tandem repetitive sequence elements provide a substrate for diversification that can drive rapid evolution, suggesting a potential mechanism underlying sexual conflict and ultimately speciation.

Genetic conflicts in molecular evolution

SMBE-PO-231

Investigating the Evolution of an Abalone Reproductive Gene Using Long-Read Sequencing

A. M. Rivera [1,*], W. J. Swanson [1]

[1]Genome Sciences, University of Washington, Seattle, United States

Abstract: While fertilization is essential, we know little of the mechanisms underlying this process. Studying gametic protein evolution can improve our understanding of reproductive barriers and their relationship to speciation. Abalone and their interacting sperm-egg proteins lysin and VERL are a classical model of fertilization. We currently know little about VERL's evolution and variation, but multiple lines of evidence suggest that specificity in lysin-VERL interactions helps maintain reproductive barriers in sympatric Pacific abalone species. The VERL gene has a 10 kb repeat array (containing 22 450-bp repeats) that has prevented accurate assembly from short reads. For this reason, I am focused on long-read sequencing of VERL in these seven Pacific abalone species and I will quantify its variation and coevolution with lysin both within and between species. My initial attempts at long-read sequencing of PCR amplified VERL, produced numerous artifacts arising from amplification of repetitive DNA. To avoid many of the challenges of PCR, I have begun direct RNA sequencing. VERL is already overrepresented in ovary RNAs, and I can further amplify for VERL using magnetic pulldown methods, as evidenced by quantitative real-time PCR experiments. I have sequenced mRNA samples directly on an Oxford Nanopore device, and have obtained partial 3' VERL sequence. Comparative genetic analysis of VERL in abalone could reveal meaningful differences in repeat number or sequence that may relate to species-specific fertilization. More work is necessary to improve the method, but these data highlight the potential to illuminate the evolutionary history of functionally relevant repetitive regions.

Genetic conflicts in molecular evolution
SMBE-PO-212

Drive against PRDM9 allelic combinations in natural populations suggest functional constraints on meiotic recombination

C. Vara [1], L. Capilla [1], L. Ferretti [2], A. Ledda [3], R. A. Sanchéz-Guillén [4], S. Gabriel [5], G. Albert-Lizandra [1], B. Florit-Sabater [1], J. Ventura [1], J. Searle [6], M. Mathias [5], A. Ruiz-Herrera [1,*]
[1]Universitat Autònoma de Barcelona, Cerdanyola del Valles, Barcelona, Spain, [2]Oxford Big Data Institute, Oxford, [3]Imperial College London, London, United Kingdom, [4]Instituto de Ecología AC (INECOL), Veracruz, Mexico, [5]Universidade de Lisboa, Lisboa, Portugal, [6]Cornell University, Ithaca, United States

Abstract: One of the major challenges in evolutionary biology is the identification of the genetic basis of post-zygotic reproductive isolation. Given its pivotal role in speciation, here we explore the drivers that may account for the evolutionary dynamics of the PRDM9 gene between continental and island systems of chromosomal variation in house mice. Using a dataset of nearly 400 wild-caught mice of Robertsonian systems we identify the extent of PRDM9 diversity in natural house mouse populations, determine the phylogeography of PRDM9 at a local and global scale based on a new measure of pairwise genetic divergence, and analyse selective constraints. Our analysis of genetic diversity and positive selection, together with a phylogeographic reconstruction of PRDM9 allelic variation suggest that the variability observed in island systems might be influenced by the presence of distinct chromosomal fusions resulting from a complex pattern of introgression or multiple colonization events. Importantly, we detect a significant reduction in the proportion of PRDM9 heterozygotes in Rb mice, which showed a high degree of similarity in the amino acids responsible for protein-DNA binding. Our results suggest that despite the rapid evolution of PRDM9 in natural populations, functional constraints could facilitate the accumulation of allelic combinations that maintain recombination hotspot symmetry. We anticipate that our study will provide the basis for examining the role of genetic conflict in reproductive isolation in natural populations.

Genetic conflicts in molecular evolution

SMBE-PO-220

Hybrid incompatibility in Pristionchus pacificus

W. Hu [1,*], R. Sommer [1]

[1]Dept. for Evolutionary Biology, Max-Planck Institute for Developmental Biology, Tuebingen, Germany

Abstract: As a diplogastrid nematode, *Pristionchus pacificus*which has a necromenic association with scarab beetles has been successfully established as a model organism in evolutionary biology. To study the speciation process of *Pristionchus pacificus*, we performed reciprocal crosses between four *P. pacificus*strains (RSC011, RSA016, RS5407 and RSB080)isolated from different locations of La Réunion Island as representive strains of four clades (clade A2, B, C1 and C2) of phylogenetic tree. Surprisingly, different degrees of sterility can be observed in since F2 RILs derived from selfing of F1 hybrids in crosses between RSC011 and other three strains, and this sterility phenotype can be observed even after F6. Among all the crosses, reciprocal crosses between RSC011 and RSB080 shows highest sterility frequencies among all crosses, over 55% and 48% of ~300 RILs became sterile until F6 in RSB080♀×RSC011♂ and RSC011♀×RSB080♂**.** Moreover, sterility of selfing of RILs could still be observed in further generations, then became stable until F10. To elucidate the mechanism behind this hybrid incompatibility between different *P. pacificus*isoaltes, we used next generation sequencing to sequence both sterile and fertile F2 RILs and try to locate the elements responsible for this phenotype by GWAS and LD analysis. This study will explore more information about the genetic basis of hybrid incompatibility within the species.

Genetic conflicts in molecular evolution
SMBE-PO-221
Consequences of hitchhiking between a male killer and a neo-sex chromosome in a butterfly
S. Martin*

Abstract: Male killing endosymbionts in insects provide extreme examples of genetic conflicts. The spread of male killers can be associated with rapid evolutionary change in the host, both as a direct consequence (e.g. mitochondrial hitchhiking), and as a selective response (e.g. evolution of suppressors). In this presentation I will describe the remarkable case of the African monarch (*Danaus chrysippus*) in which a spreading male killer *Spiroplasma* has caused the concurrent spread of not only a single mitochondrial haplotye but also a neo-W sex chromosome that carries a wing patterning supergene that only recently become sex linked. I will describe the immediate consequences of this spread for the evolution of the host, and also present analyses to determine whether the host has begun to evolve in response to this conflict that emerged just a few thousand years ago.

Genetic conflicts in molecular evolution
SMBE-PO-219
Strong hybrid male incompatibilities impede the spread of a selfish chromosome between populations of a fly
R. Verspoor*, T. Price [1]
[1]University of Liverpool, Liverpool, United Kingdom

Abstract: Intragenomic conflict has been proposed to drive speciation. X chromosomes that enhance their transmission by killing Y-bearing sperm are particularly likely to drive differentiation. As males with X chromosome drive produce few or no sons, the fitness costs are severe for the rest of the genome, potentially causing autosomal genes to evolve suppression to the sperm killing mechanism.

This counter-adaptation could create cyclical co-evolution between X-chromosome drive and the suppressors. If this co-evolution occurs in specific populations it could create interpopulation differences in reproduction genes that cause hybrid incompatibilities. Alternatively, meiotic drivers could homogenise populations, because a driver might spread rapidly through a naïve population with no suppressors.
We investigate the consequences of population specific genetic conflict using an X-chromosome meiotic driver in *Drosophila subobscura*. We show that phenotypic suppression exists in the native population, but not in a nearby naive population. However, meiotic drive carrying males are rendered almost completely infertile in crosses with the naive population, suggesting that in this system meiotic drive is causing hybrid incompatibilities that prevent the driver from spreading between populations. So intragenomic conflicts over drive in one population appear to both be creating the early stages of reproductive isolation between populations, and rendering the driving X incapable of leaving the host population.

Genetic conflicts in molecular evolution

SMBE-PO-218

Selection within and between individuals of a mitochondrial deletion in C. elegans

J. Dubie [1,*], U. Bergthorsson [1], V. Katju [1]

[1]Veterinary Integrative Biosciences, Texas A&M, College Station, United States

Abstract: Mitochondria exist within a nested hierarchy of populations, with different evolutionary forces acting upon competition between populations, individuals within populations, and different mitotypes within individual cells. This can lead to conflict between an organism and its mitochondria, as different mitotypes compete to increase their frequency regardless of the effects on the organism's fitness. These selfishly acting mitotypes often contain deletions (ΔmtDNA) and have been implicated in the evolution of sex, mitonuclear coadaptation, and human aging and disease. Despite this, little is understood regards the population dynamics of mitochondrial mutations. We used a selfishly acting mitochondrial DNA molecule containing a 400-bp deletion in *ctb-1* that arose within an experimental evolution *Caenorhabditis elegans* line to analyze the population dynamics of ΔmtDNA with and without competition between individuals. We demonstrate that, in a bottlenecked population (*N*=1) without competition between individuals, the frequency of ΔmtDNA within individuals rapidly (<50 generations) increased (80-98%) after origination but was then maintained for more than 150 generations without fixing. In large homogeneous populations, the ΔmtDNA persists but its average intracellular frequency decreases over time. There is also an increase in the variation of ΔmtDNA intracellular frequencies found within the large populations over time.

Genetic conflicts in molecular evolution

SMBE-PO-227

The long and the short of it: linking genome-wide signatures of selection across evolutionary timescales in birds

T. Sackton [1,*], A. Shultz [2], B. Arnold [1]

[1]Harvard University, Cambridge, [2]Natural History Museum of LA County, Los Angeles, United States

Abstract: A challenge of evolutionary biology is linking short-term evolutionary processes to long-term patterns of biodiversity. Many studies focus on a single evolutionary timescale, rather than studying the same question across timescales, inhibiting the ability to link evolution through time. In this study, we compare population-level genome-wide signatures of positive selection to a recently published set of genes under positive selection across the bird tree of life. We obtained genome-wide signatures of selection across populations of birds from over 30 species, first, by building and executing a standardized pipeline to process publicly-available whole-genome resequencing data using a version of the GATK pipeline optimized for non-model organisms. Then, we used a variation of the McDonald-Kreitman test to detect signatures of selection in coding regions of each species. Preliminary results in several species of songbirds suggest that there is a significantly greater overlap than expected in genes under selection at both short (across populations) and long (across Aves) evolutionary timescales. Genes under positive selection in populations of multiple species have immune, recombination, or reproductive functions, strengthening the hypothesis that a host-pathogen arms race may be one of the most consistent selective pressures across different bird species, and highlighting the role of genetic conflict in shaping genome evolution across a variety of time scales.

Genetic conflicts in molecular evolution

SMBE-PO-225

The role of alternative splicing in the resolution of sexual conflict

T. F. Rogers [1,*], A. Wright [1]

[1]Department of Animal and Plant Sciences, The University of Sheffield, Sheffield, United Kingdom

Abstract: Males and females are often subject to conflicting selection pressures, yet they share an almost identical genome. This genetic correlation can create a significant evolutionary burden on populations, acting as an important force in adaptation. The mechanisms by which sexual conflict can be resolved have been the focus of considerable recent debate. However, the role of alternative splicing, a regulatory process that increases proteomic complexity by generating multiple transcripts from one locus, has been overlooked. We investigated the extent to which sex differences in splicing can resolve sexual conflict, and the relative importance of different gene regulatory mechanisms in the construction of sex-specific architecture. Using male and female transcriptome data across a range of avian species, we combined patterns of alternative splicing with population genomic statistics to test how splicing evolves in response to sexual conflict. First, we show that sex differences in splicing correlate with phenotypic sex differences, indicating that sex-specific isoforms may play a key role in male and female phenotypes. Second, sex differences in splicing are not common in genes with sex-biased expression, suggesting that different regulatory mechanisms may underpin sex-specific regulatory networks. Finally, we used a population genomic approach to test the role of alternative splicing in conflict resolution. Together, our findings show that regulatory evolution is a rapid and efficient route to the resolution of conflict and uncoupling of genetic constraints.

Genetic conflicts in molecular evolution

SMBE-PO-229

Molecular characterization of two hybrid dysgenesis systems in Drosophila melanogaster

F. Schwarz [1,*], F. Wierzbicki [1], O. Cannalonga [1], K.-A. Senti [1], R. Kofler [1]

[1]Institute of Population Genetics, Vetmeduni Vienna, Vienna, Austria

Abstract: Transposable elements (TEs) are among the main causative agents of complex structural variation in genomes and their uncontrolled activity is generally considered detrimental to the host. One of the main defense mechanism against TEs are piRNAs, small RNAs presumed to be mainly produced in repetitive genomic regions termed 'piRNA clusters'. Prominent systems of strongly detrimental phenotypes induced by TE activity are the hybrid dysgenesis (HD) systems in Drosophila melanogaster. Briefly, crosses of a male carrying a particular TE with females not carrying the TE lead to infertile offspring, while the offspring of the reciprocal cross is viable. However, the genomic basis of HD remains poorly characterized. We thus used long-read sequencing (Oxford Nanopore), chromosome conformation capture (Hi-C) and small RNA sequencing to characterize a pair of strains from two different HD systems: the P-element (a DNA transposon) and the I-element (a non-LTR retrotransposon).We generated high-quality genome assemblies, characterized structural variants and TE insertions and identified piRNA-producing insertions (within and outside of piRNA clusters). Finally, we tested the "shotgun-silencing model" which holds that piRNA producing TE insertions are segregating. Under this model, we do not expect fixed cluster insertions for recently active TEs.

Genetic conflicts in molecular evolution

SMBE-PO-226

Ignored heterogeneous codon frequencies among sites biases the estimation of molecular adaptation through dN/dS

R. Del Amparo [1,*], A. Vicens [1], M. Arenas [1]

[1]Biochemistry, Genetics and Immunology, University of Vigo, Vigo, Spain

Abstract: Quantifying selection is fundamental to understand population genetics and evolution. In this concern, a traditional parameter to evaluate the presence of selective constraints in protein-coding genes is the nonsynonymous/synonymous substitution rate ratio (*dN/dS*). However the estimation of this parameter can be biased if some evolutionary processes, such as recombination, are ignored. Here we investigated a new bias in the estimation of *dN/dS* caused by variable codon frequencies among codon sites. Applying extensive computer simulations we found an overall underestimation of *dN/dS* when codon frequencies vary along sequences. Indeed, this underestimation increases with the amount of variability of codon frequencies. Interestingly, we also found that if frequencies only vary at first or second codon positions then *dN/dS* is underestimated, but if they vary at the third codon position then *dN/dS* is overestimated. We interpret these biases as a consequence of the violation of assumptions made in common methods to estimate *dN/dS*. Finally, we propose a methodology based on the independent estimation of *dN/dS* from partitions constituted by codon sites with similar codon frequencies and doing so we could partially reduce this bias. We conclude that accounting for heterogeneous codon frequencies along sequences is crucial to obtain more realistic estimates of molecular adaptation through *dN/dS*.

Genetic conflicts in molecular evolution

SMBE-PO-222

Polyadenylation event by forward-oriented AluYRa1 in cynomolgus macaque genome

H.-M. Cho [1 2,*], Y.-H. Kim [1 2], S.-J. Park [1], B.-W. Sim [3], J.-R. Lee [4], S.-H. Choe [1 2], J.-W. Huh [1 2]

[1]National Primate Research Center, Korea Research Institute Of Bioscience And Biotechnology, Cheongju, [2]Department of Functional Genomics, KRIBB School of Bioscience, University of Science & Technology(UST), Daejeon, [3]Futuristic Animal Resource and Research Center, Korea Research Institute Of Bioscience And Biotechnology, Cheongju, [4]Primate Resource Center, Korea Research Institute Of Bioscience And Biotechnology, Jeongeup, Korea, Republic Of

Abstract: Transposable elements (TEs) are considered as a crucial source of transcript diversification by being involved with alternative splicing (AS) event in different ways. It has been reported that 70% of human mRNA isoforms are produced by alternative polyadenylation (APA). We have previously identified *Alu*YRa1, one of TEs, located at the end of the several genes in cynomolgus macaque through the *in-silico* analysis. To investigate this, further detailed computational analysis of recently registered mRNA sequences and the subsequent experimental validation were performed. Computational analysis revealed that 10 genes of cynomolgus macaque contain *Alu*YRa1 at their end, and 9 *Alu*YRa1 of them are forward-oriented. Furthermore, the ones in 7 genes, *GTPBP4*, *PEX26*, *IRF9*, *TK2, CMBL, SLC16A14* and *PDK4* are expected to have the same consensus sequence of polyadenylation cleavage. Therefore, we carried out genomic PCR to analyze these *Alu*YRa1s, and RT-PCR to check their expression pattern. Our results show that *Alu*YRa1 is distributed in rhesus and cynomolgus macaque and forward-oriented one at 3'UTR tends to provide proper circumstance for polyadenylation. Additionally, this element shows polymorphic insertion pattern in *TK2* and *PDK4.* In conclusion, *Alu*YRa1 is an Old World monkey specific TE, and its forward-oriented insertion diversify gene transcripts mostly by alternative polyadenylation.

Keywords: *Alu*YRa1, alternative polyadenylation, transposable element, exonization, primate evolution

Genetic conflicts in molecular evolution

SMBE-PO-223

Alu-derived isoform creation mechanism during the mammalian evolution

S.-H. Choe [1 2,*], H.-M. Cho [1 2], B.-W. Sim [3], S.-J. Park [1], Y.-H. Kim [1 2], J.-W. Huh [1 2]

[1]National Primate Research Center, Korea Research institute of bioscience & Biotechnology, [2]Department of functional genomics, KRIBB school of bioscience, University of Science & Technology (UST), [3]Futuristic Animal Resource & Research Center, Korea Research institute of bioscience & Biotechnology, Cheongju, Korea, Republic Of

Abstract: *Alu* are primate-specific short interspersed sequence elements (SINEs), ~300 nt in length, integrating within a genome through retrotransposition mechanism. Also *Alu* element can be inserted into mRNA by way of splicing termed 'exonization' event. These events potentially alter splicing process and contributed to create primate specific transcriptomic diversity. We found novel transcripts of specific gene, are derived from the integration of an *Alu* element-*Alu*Sz6 from crab-eating monkey full-length cDNA sequences. We performed PCR-amplification and sequencing analysis with *Alu* consensus sequences in primates including human, chimpanzee, gorilla, macaque, marmoset, lemur for integration timing assessment and phylogenetic analysis. As a results, we found that the *Alu*Sz6 insertion events were occurred in our common ancestor before the divergence of simian and prosimian. Intriguingly *Alu*Sz6 in catarrhini lineage show insertion of G in the *Alu*-exon and subsequent production of a canonical 3' splice site. In addition, we performed global *in silico* analysis in 22 mammalian species to compare the functional proteins from *Alu*-derived isoform. Thus, our results suggested that unique exonization event was derived by the 'G' addition in the *Alu* element, and specific insertion event could make a novel protein coding sequences during primate evolution.

Keywords: Alu, Transposable element, primate evolution, Exonization

Genetic conflicts in molecular evolution

SMBE-PO-235

Dynamics of gene conversion in human MSY palindromes

M. Bonito [1,*], E. D'Atanasio [1], A. Novelletto [2], F. Cruciani [1], B. Trombetta [1]

[1]Sapienza University of Rome, [2]University of Rome Tor Vergata, Rome, Italy

Abstract: The Male Specific region of the human Y chromosome (MSY) is characterized by the presence of 8 near identical 'pseudo-diploid' sequences, called palindromes, which are designated as P1–P8. Palindromes are composed of inverted repeats (palindrome arms), separated by a non-duplicated spacer. Although these structures originated in a non-recombining context, they show evidence of a strong recombinational activity. Palindromes exhibit more than 99.94% sequence identity between arms, due to the homogenizing effect of arm-to-arm gene conversion (GC), a type of recombination which involves the non-reciprocal transfer of genetic information from a "donor" sequence to a highly similar "acceptor" sequence. The independent appearance of these paralogue structures in sex chromosomes of many different species suggests that they may have an important biological meaning. It has been hypothesized that the palindromic organization and the establishment of inter-paralogs gene conversion have a strong adaptive significance since the arm-to-arm GC may allow double-strand break repair and the efficient removal of deleterious mutations. Thus, it has been proposed that this mechanism was acquired to maintain the structural integrity of multi-copy genes involved in the male-fertility. Moreover, it has been hypothesized that gene conversion evolved as a mechanism to retain the ancestral state of sequences: a *de novo* mutation in a palindrome arm is preferentially back mutated to the ancestral state rather than transmitted to the other arm.

In this contest, to gain new insights into the dynamics of gene conversion within human Y chromosome palindromes, we performed next-generation sequencing (depth >50×) of 3 palindromes (P6, P7 and P8) in 157 samples, chosen to represent the most divergent evolutionary lineages of the MSY. In this analysis, we overcame the problem of the inaccurate mapping of the duplicated reads, and performed a sequencing depth analysis to detect deletions or duplications which may result in genotype miscalling. We identified more than 200 paralogue sequence variants and 140 GC events. Mapping these GC events across a stable and non-ambiguous Y chromosome phylogeny enabled the calculation of a precise Y-Y gene conversion rate for each palindrome and the assessment of the direction of the recombinational activity. From these analyses we conclude that MSY palindromes have an evolutionary pattern more complex than previously thought.

Genetic conflicts in molecular evolution

SMBE-PO-230

Transposable Element Misregulation Is Linked to the Divergence between Parental piRNA Pathways in Drosophila Hybrids

V. Romero-Soriano*, L. Modolo, H. Lopez-Maestre, B. Mugat, E. Pessia, S. Chambeyron, C. Vieira, M. P. García Guerreiro

Abstract: Interspecific hybridization is a genomic stress condition that leads to the activation of transposable elements (TEs) in both animals and plants. In hybrids between *Drosophila buzzatii* and *Drosophila koepferae*, mobilization of at least 28 TEs has been described. However, the molecular mechanisms underlying this TE release remain poorly understood. To give insight on the causes of this TE activation, we performed a TE transcriptomic analysis in ovaries (notorious for playing a major role in TE silencing) of parental species and their F1 and backcrossed (BC) hybrids. We find that 15.2% and 10.6% of the expressed TEs are deregulated in F1 and BC1 ovaries, respectively, with a bias toward overexpression in both cases. Although differences between parental piRNA (Piwi-interacting RNA) populations explain only partially these results, we demonstrate that piRNA pathway proteins have divergent sequences and are differentially expressed between parental species. Thus, a functional divergence of the piRNA pathway between parental species, together with some differences between their piRNA pools, might be at the origin of hybrid instabilities and ultimately cause TE misregulation in ovaries. These analyses were complemented with the study of F1 testes, where TEs tend to be less expressed than in *D. buzzatii*. This can be explained by an increase in piRNA production, which probably acts as a defence mechanism against TE instability in the male germline. Hence, we describe a differential impact of interspecific hybridization in testes and ovaries, which reveals that TE expression and regulation are sex-biased.

Genome-wide methods for detecting selection

SMBE-PO-270

The impact of purifying and positive selection in Arabis alpina populations with different mating systems

M. Fracassetti [1,*], B. Laenen [1], J. Gutiérrez [1], F. Kolář [2], T. Slotte [1]

[1]Department of Ecology, Environment and Plant Sciences, University of Stockholm, Stockholm, Sweden, [2]Department of Prague, Charles University in Prague, Prague, Czech Republic

Abstract: In flowering plants the transition from outbreeding to self-fertilization is very common, and mating system variation is considered to be a major driver of genetic diversity. There is accumulating evidence for relaxed purifying selection in highly selfing species, in line with theoretical predictions, but fewer empirical studies have explicitly examined the effects of intermediate levels of outcrossing, so-called mixed mating. The alpine rock-cress (*Arabis alpina*) is a promising model system to address the effects of mixed mating, as it harbors both outcrossing, selfing and mixed mating populations, allowing us to investigate the action of natural selection in populations with different mating systems within the same species. *A. alpina* is a perennial plant that grows mainly in Arctic regions and mountainous areas in Europe, North and East Africa and Middle East. We first examine population structure across 15 sites in Europe, based on whole-genome sequences of 300 individuals of *A. alpina*, and then assess whether purifying and positive selection acts differently on these populations. We compare purifying selection based on estimates of the distribution of fitness effects and genetic load and apply machine learning methods to detect soft and hard sweeps. Our results contribute to an improved understanding of how mating system variation shapes the genomic impact of natural selection.

Genome-wide methods for detecting selection

SMBE-PO-242

Classic and introgressed selective sweeps shape wing colour pattern loci across a butterfly adaptive radiation

J. James [1 2,*], M. Moest [2 3], S. Van Belleghem [2 4], C. Jiggins [2]

[1]Department of Ecology and Evolutionary Biology, University of Arizona, Tucson, United States, [2]Department of Zoology, University of Cambridge, Cambridge, United Kingdom, [3]Department of Ecology, University of Innsbruck, Innsbruck, Austria, [4]Department of Biology, University of Puerto Rico, Rio Piedras, Puerto Rico

Abstract: Characterising genomic signatures of selection is key to our understanding of adaptive evolution. However, in natural populations such signatures may be affected by a large combination of factors including demographic processes, the strength of selection, the type of sweep and the age of the selective event. In our analysis, we have investigated the impact of selection at loci involved in wing colour patterning across the *Heliconius* adaptive radiation, sampling from nearly 600 individuals in 53 populations. We found that sites with the strongest signatures of selection corresponded to the loci which have the greatest phenotypic effects, consistent with these sites having a greater importance for visual identification by predators, and are found in populations with geographically restricted wing colour patterns, indicating recent sweeps in these populations. In contrast, more widespread colour patterns show weaker signatures of selection, suggesting older sweep events. Additionally, we used simulations to compare genomic sweep signatures expected under classic hard sweeps as compared to adaptive introgression, allowing us to characterise the unique signal that adaptive introgression produces in genomic data. We found that introgression results in a distinct 'volcano'-like pattern, with peaks of increased genetic diversity close to the target of selection, consistent with patterns found in some of our *Heliconius* populations. Our results provide insight into the recent history of selection across this species radiation, and illustrate some of the complexities of signatures of selection.

Genome-wide methods for detecting selection

SMBE-PO-272

The interplay between karyotypic and population effects in base composition evolution in vertebrate genomes

Y. Clément [1,*], C. Bon [2], H. Roest Crollius [1]

[1]Institut de Biologie de l'ENS (IBENS), Département de biologie, École normale supérieure, CNRS, INSERM, Université PSL, 75005 Paris, [2]Laboratoire Éco-Anthropologie et Ethnobiologie, UMR 7206 CNRS - Muséum National d'Histoire Naturelle, Université Paris Diderot, Sorbonne Paris Cité, 75016 Paris, France

Abstract: Vertebrate genomes harbor differences in GC content and distributions, from relatively homogeneous distributions in mouse, anole lizard or fish to more heterogeneous distributions in human or, chicken. In vertebrate genomes, GC content evolution is affected by GC-biased gene conversion (gBGC), a neutral process associated with meiotic recombination that favors the fixation of G & C alleles in highly recombining regions. gBGC behaves like natural selection, and various studies have shown that effective population size is linked with gBGC efficiency and GC content evolution. Moreover, the distribution and intensity of meiotic recombination within a genome is driven by chromosome size distribution because larger chromosomes are subject to lower rates of recombination than smaller chromosomes. As a consequence, chromosomal rearrangements affect GC content evolution: fusions or fissions changing chromosomal length affect recombination rates, while changing gene locations within a genome alter their recombination landscape. How modifications in effective population size (population effects) and chromosomal rearrangements (karyotype effects) work together to affect GC content evolution in vertebrates is still an open question.

We first simulated population and karyotype effects on GC content evolution and showed that an increase of gBGC intensity will increase both GC content mean and variance. We show that chromosomal size heterogeneity is also linked with both GC content mean and variance. Finally, we show that gene shuffling within a genome leads to more homogeneous GC content.

Second, we analyzed the evolution of GC content in vertebrate species and ancestral lineages using ancestral sequence reconstruction in coding sequences. We also used high quality ancestral gene order reconstructions in vertebrates to connect changes in GC content with chromosomal rearrangements. We find that the so-called GC-rich isochores inferred to originate at the ancestor of amniotes in fact originate from GC-richer sequences in the ancestor of bony vertebrates (euteleostomi) some 100 million years earlier. We also show that despite all fish species having relatively homogeneous GC content distributions, the evolution of GC content is complex in these species. In murids, the homogenisation of GC content took place in at least two distinct waves. Finally, we could detect the effects of chromosomal rearrangements on GC content evolution in a large number of lineages under scrutiny, which demonstrates the importance of karyotype effects on GC content evolution in vertebrates. With these results and analyses, we propose a model of base composition evolution that incorporates both population and karyotype effects.

Genome-wide methods for detecting selection
SMBE-PO-238
Bootcluster: A fast, tree-based recombination detection and visualization tool
K. Hagihara*, K. Nishimiya, H. Watanabe

Abstract: Detection of genetic recombination is an essential step toward understanding genome evolution, including regional selection after introgression or gene conversion, and hence, a lot of recombination algorithms have been developed. However, as this fact itself indicates, every algorithm has both pros and cons of the performance. In addition, most of the algorithms have been implemented in specific operating systems/environments and tend to be unstable and/or difficult to install properly. One of the recombination detection algorithms is tree-based comparison between different genomic regions. Because of the ease of interpreting the results, tree-based recombination detection tools have been frequently used for a wide variety of studies, including those of the authors of this study. However, the authors' experience has shown that those available tools underperform in many cases. In order to compensate and overcome the available tools, a new fast, tree-based recombination detection algorithm, called Bootcluster, was developed. The Bootcluster program can be installed in most operating systems as a command line interface (CLI) version and even run in a browser as a local application with a graphical user interface (GUI), making it available to virtually every researcher. The performance of Bootcluster will be demonstrated with several different cases, for example, for the window analysis of 64 complete human adenovirus genomes, the Bootcluster program takes 50 seconds while a popular tree-based recombination detection tool takes 100 seconds.

Genome-wide methods for detecting selection

SMBE-PO-246

A likelihood-free framework for classifying natural selection events using convolutional neural networks

U. Isildak [1], M. Fumagalli [2,*]

[1]METU, Ankara, Turkey, [2]Imperial College London, London, United Kingdom

Abstract: The identification of selection signatures in the genome has the potential to unveil the genetic mechanisms underpinning phenotypic variation. Commonly used strategies to detect such signals rely on compressing the genomic information into summary statistics with an inevitable loss of information. Furthermore, few methods are able to quantify the strength of selection. Here we explored the use of deep learning in evolutionary biology and, in particular, the application of convolutional neural networks on population genomic data for the detection and quantification of natural selection.

We represent the information of genomes from multiple individuals as abstract images, depicting the genomic variation at each position for each sample. Each image is created by stacking aligned genomic data and encoding distinct alleles into separate colors. To detect and quantify signatures of positive selection, convolutional neural networks are trained using simulations.

We will discuss how data manipulation and learning strategies can affect the prediction of selection signals. In particular, we will illustrate several options for sorting genomic images and for assessing the error induced by misspecified demographic models during training. We will finally illustrate how this approach can be used to quantify selection coefficients during a sweep and how to characterize signatures of balancing selection.

While the use of deep learning in evolutionary genomics is in its infancy, here we will demonstrate its potential and discuss future directions to detect informative patterns from large-scale genomic data.

Genome-wide methods for detecting selection

SMBE-PO-243

Comparing and applying four MKT methods to detect and quantify natural selection at the genome level

J. Murga-Moreno [1,*], M. Coronado-Zamora [1], S. Casillas [1], A. Barbadilla [1]

[1]Departament de Genètica i de Microbiologia, Universitat Autònoma de Barcelona, Cerdanyola del Vallès, Spain

Abstract: One of the most striking evidence of the power of natural selection is the characteristic footprints that it leaves on the patterns of genetic variation. The McDonald and Kreitman test (MKT) 1 is one powerful and robust methods to detect the action of natural selection at the molecular level. MKT can detect the action of recurrent positive selection by analyzing polymorphism and divergence data altogether. The main drawback of MKT is that it assumes that only neutral mutations contribute to polymorphism, but weak negative selection abounds in genomes, biasing downward the estimated adaptation values. Several methodological MKT extensions have attempted to correct for this bias by taking into account slightly deleterious polymorphism. Here, we perform a comparison of four different MKT methods: (i) the standard (original) MKT 1 ; (ii) the Fay Wickoff and Wu correction 2 ; (iii) the Extended MKT 3 and (iv) the asymptotic MKT 4 . Two population genomic data, real and simulated, are used to assess their performance for different datasets and evolutionary scenarios. Genome-wide DNA variation data come from *Drosophila melanogaste*r and human populations, and simulated data was generated with the SLiM 2 evolutionary framework 5 . We test several conditions including gene-to-gene vs gene concatenating analysis, and recombination effect to assess the power and bias of selection estimates of the different MKT methods. Furthermore, we developed iMKT (acronym for integrative McDonald and Kreitman test), a web-based service performing the four MKT methods. iMKT allows the detection and estimation of four selection regimes (adaptive, neutral, strongly deleterious and weakly deleterious). User's own population genomic data, and pre-loaded *D. melanogaster* and human sequences of protein-coding genes for 16 and 26 populations, respectively, can be analyzed. iMKT is a comprehensive reference site for the study of protein adaptation in massive population genomics datasets, especially in Drosophila and humans. iMKT is a free resource online at https://imkt.uab.cat .

1. McDonald, J. H. & Kreitman, M. Adaptive protein evolution at the Adh locus in Drosophila. Nature 351, 652–654 (1991).
2. Fay, J. C., Wyckoff, G. J. & Wu, C.-I. Positive and Negative Selection on the Human Genome. Genetics 158, 1227–1234 (2001).
3. Mackay, T. F. C. et al. The Drosophila melanogaster Genetic Reference Panel. Nature 482, 173–178 (2012).
4. Messer, P. W. & Petrov, D. A. Frequent adaptation and the McDonald-Kreitman test. Proc. Natl. Acad. Sci. 110, 8615–8620 (2013).
5. Haller, B. C. & Messer, P. W. SLiM 2: Flexible, Interactive Forward Genetic Simulations. Mol. Biol. Evol. 34, 230–240 (2017).

Genome-wide methods for detecting selection

SMBE-PO-236

Are significant results of the branch-site test of positive selection largely spurious due to multinucleotide mutations?

Z. Zou [1,*], J. Zhang [1]

[1]Department of Ecology and Evolutionary Biology, University of Michigan, Ann Arbor, United States

Abstract: The branch-site test has been widely used for detecting positive selection in coding sequence evolution along specific lineages. A recent study reported that positive signals from this test are almost exclusively attributable to codons with multiple nonsynonymous differences (CMNDs), which were claimed to have originated largely from multinucleotide mutations (MNMs). Here we test an alternative hypothesis that most CMNDs are not due to MNMs but are the results of selection for successive nonsynonymous mutations within codons. Upon the control of the MNM effect by focusing on pairs of nonsynonymous differences with a fixed number of intervening nucleotide sites, we found the number of codons showing two nonsynonymous differences per codon significantly greater than that of adjacent codons having one nonsynonymous difference in each codon, strongly suggesting positive selection for successive nonsynonymous substitutions within codons. We have implemented a likelihood method that takes into account MNMs but also allows selection for successive nonsynonymous substitutions within codons, and are testing the performance of this new method using computer simulation as well as the actual data from mammals and fruiflies.

Genome-wide methods for detecting selection
SMBE-PO-239
The whale shark genome reveals how genomic and physiological properties scale with body size
J. Weber*

Abstract: We sequenced and assembled the genome of the endangered whale shark (*Rhyncodon typusi*), the world's largest fish, and compared it to the genomes of 83 animals to characterize the relationships between genome features and biological traits. The relationships between body mass, longevity, and basal metabolic rate (BMR) across diverse habitats and taxa have been researched extensively over the last century, and have led to generalized rules and scaling relationships that explain many physiological and genetic trends observed across the tree of life. Expanding these comparisons to include genomes, we found that major genomic traits, including intron length and gene length, correlate with body size, temperature, and lifespan in most species, and that GC content and codon adaptation index are negatively correlated. In the whale shark genome, specifically, we found that introns are longer than in most other species due to the presence of repetitive CR1-like elements, and that neural genes of several types, including neurodegeneration genes, are much longer than average genes in species with long lifespans. These results suggest that there is an evolutionary relationship between gene size and physiological traits size such as body size, metabolic rate, and lifespan. This holds particularly among genes whose functions are essential for living long lives, such as telomere maintenance and energy production. These results show the power of the comparative evolutionary approach to uncover both general and specific relationships, and suggest ways in which genome architecture is shaped by size and ecology.

Genome-wide methods for detecting selection

SMBE-PO-241

Using GWAS methods to infer patterns of positive selection in Oceania

S. Kaewert [1,*], G. L. Wojcik [2], E. P. Sorokin [2], J. R. Homburger [3], C. D. Quinto-Cortés [4], A. Ioannidis [5], J. B. Portillo [4], J. E. Rodriguez [4], K. Auckland [6], A. Severson [3], J. Blanc [7], G. M. Belbin [8], K. Ferrier [9], S. Oppenheimer [6], T. Parks [6], M. Phipps [10], F. Aghakhanian [10], K. Robson [6], E. E. Kenny [8], A. V. S. Hill [6], A. J. Mentzer [6], C. Bustamante [2], A. Moreno-Estrada [4], C. R. Gignoux [11]

[1]Colorado Center for Personalized Medicine, University of Colorado, Aurora, [2]Department of Biomedical Data Science, [3]Department of Genetics, Stanford University, Stanford, United States, [4]National Laboratory of Genomics for Biodiversity, Irapuato, Mexico, [5]Computational and Mathematical Engineering, Stanford University, Stanford, United States, [6]University of Oxford, Oxford, United Kingdom, [7]University of California, Davis, Davis, [8]Icahn School of Medicine at Mount Sinai, New York, [9]University of Colorado, Aurora, United States, [10]Monash University, Bandar Sunway, Malaysia, [11]Colorado Center for Personalized Medicine, Aurora, United States

Abstract: The human settlement of the Pacific was a complex demographic event involving multiple migrations and exposure to a wide variety of environmental pressures on new islands, including the unique disease dynamics present in isolated, high-density populations with intermittent contact. These challenges resulted in opportunities for local adaptation to new conditions. However, Oceania remains underrepresented in studies of genetic selection. Using data from the Oceanian Genome Variation Project (OGVP), involving 980 individuals from 88 populations measured on the MEGA platform at over 1.7M sites along the genome, we used two methods of finding evidence of positive selection, both of which account for local population structure. The first employs the widely-used framework of linear mixed effects models, originally developed for GWAS, which allow for association with an outcome of interest while controlling for background genomic differentiation. We follow up via the machine learning approach SWIFr to probabilistically combine haplotype-based (iHS, nSL, and xp-EHH) statistics to test for selection in various regions of Oceania as compared with coalescent simulations. Using these methods, we find genome-wide significant levels of potential selection in multiple regions, including in HLA-DRB1 and -DQA1 in West Polynesia (p= 5.02e-10 and 6.87e-09, respectively). We also find significant enrichment in East Polynesia in top scoring genes (min p < 10e-5) involved in monosaccharide metabolism and response to xenobiotic stimulus. The results suggest local adaptation to infectious disease and diet, and provide a demonstration of the framework of GWAS-based methods to efficiently assess evidence for selection across the genome.

Genome-wide methods for detecting selection

SMBE-PO-253

Hierarchical analyses reveal differential adaptation footprints in the edible sea urchin Paracentrotus lividus

C. Carreras [1], A. García-Cisneros [2], O. Wangensteen [3], V. Ordóñez [1], C. Palacín [1], X. Turon [2], M. Pascual [1,*]

[1]Universitat de Barcelona, Barcelona, [2]CEAB-CSIC, Blanes, Spain, [3]UiT The Arctic University of Norway, Tromsø, Norway

Abstract: Species with wide distribution ranges may encounter diverse environmental conditions conferring different selective pressures. The Atlanto-Mediterranean edible purple sea urchin *Paracentrotus lividus* is an exploited keystone species in littoral benthic communities showing early signs of collapse in some areas. We genotyped by sequencing 241 individuals sampled in 11 localities along its distribution range, with a mean of 4 million reads per individual. The 3,348 loci retained after filtering identified significant structuring and a gradient matching the longitudinal position of the localities. Two main genetic clusters (Atlantic and Mediterranean) were detected, and a hierarchical analysis revealed subtler patterns of differentiation within them. Outlier candidate markers identified between and within these two main clusters were mostly different, likely indicating different selective pressures. Regional missing loci further supported the differentiation within and between areas. Adaptation to salinity appeared as an important driver of the transition between Atlantic and Mediterranean basins as revealed by the molecular function of outlier loci and their significant correlation to maximum salinity. Other stressors, such as minimum temperature, seem to define the structuring within the Mediterranean. Our study shows the potential of population genomics to analyse the fine scale genetic structure of marine species, even with long dispersal capabilities as in this case. We also stress the importance of a hierarchical analysis (i.e., defining groups of genetic clusters and analysing them separately) when different levels of geographical structuring are present.

Genome-wide methods for detecting selection

SMBE-PO-244

Phenotypic effects of a per-individual burden score comprised mostly of common and low-frequency deleterious variants

H. M. T. Vy [1,*], A. Dobbyn [1], D. Jordan [1], D. Balick [1 2], R. Do [1]

[1]Personalized Medicine, Icahn School of Medicine at Mount Sinai, NY, [2]Division of Genetics, Brigham & Women's Hospital, Harvard Medical School, Boston, United States

Abstract: Human genetic variation harbors a large number of deleterious variants which collectively can decrease an individual's fitness.Despite complex disease genetic studies showing that polygenic risk scores, which measure the cumulative effect of disease-associated variants typically with common or low-frequencies(>0.5%), can have profound effects on complex disease, little is known about the cumulative effect of deleterious variants are acted upon by purifying selection at the individual level and its effect on disease. In the present study, analogous to polygenic risk score analyses, we developed a per-individual burden scoreby summing the effect of deleterious variants after applying a weight reflecting purifying selection. Using the site frequency spectrum, we evaluated four established methods for our weight: Gerp, phyloP, CADD, and fitcons to see how well each score approximates purifying selection by qualitatively tracking with allele frequency, and observed that phyloP was the best performing score. We next calculated a per-individual burden score, weighted by phyloP, using two sets of variants: 325,057coding sites, and 20,784,990 non-coding sites (majority with>0.5% frequency) restricted to regulatory regions, in 258,492British white individuals from the UK Biobank. We performed association analysis of our burden score on 1,036 medical phenotypes from the UK Biobank. We observed significant associations of the coding burden score on several adiposity and metabolic rate phenotypes. These results suggest that the accumulation of deleterious coding variants at the individual level can have a detectable impact on one's phenotype.

Genome-wide methods for detecting selection

SMBE-PO-252

Mapping ancient admixture and selection via fast, heuristic ancestral recombination graph inference

N. Schaefer [1,*], B. Shapiro [2], R. E. Green [1]

[1]Biomolecular Engineering, [2]Ecology and Evolutionary Biology, University of California, Santa Cruz, Santa Cruz, United States

Abstract: Recent technological advances have made available haplotype-phased genome-wide data sets from large panels of individuals. These data sets enable large-scale population genomic analyses of demographic history, admixture, and natural selection. Inference of the ancestral recombination graph (ARG), a data structure describing how all haplotypes are related at every variable site in the genome, is a powerful tool for such analyses. We present a new algorithm for ARG inference that scales to panels of hundreds or thousands of haplotypes, makes no prior assumptions other than parsimony, and produces results that are always consistent with the input data. We demonstrate the accuracy of our algorithm using simulated data and present a method for detecting adaptively introgressed segments of archaic hominin ancestry in a panel of human genomes.

Genome-wide methods for detecting selection
SMBE-PO-248
Title: Genetic risk variants for autoimmune diseases: adaptive variants or hitchhikers?
B. Yunusbayev [1,*]
[1]Institute of Genomics University of Tartu, Tartu, Estonia

Abstract: Genomic signatures of positive selection often overlap with genes related to immune response, and it is likely that adaptive phenotypes that previously conferred resistance to pathogens may contribute to the aetiology of autoimmune diseases. While this hypothesis currently often discussed, there is still little empirical data to support it. It is not clear whether the genetic variants that were driving positive selection are the same that contribute to autoimmune disease risk (model 1) or disease-variants were hitchhiking on positively selected haplotypes (model 2)? Here, we use 2500 high-coverage full genome sequences on a genetically homogenous European population (Estonians) to reconstruct demographic history and map putative adaptive variants. Accurate knowledge about the demographic history for the target population, allowed us to apply the machine learning approach that distinguishes between neutrally evolving loci and those simulated under soft and hard sweep scenarios. We then predicted most likely favoured mutations in these regions and gathered statistical evidence in favour of the two competing models by comparing our set of putatively adaptive variants with known risk variants for autoimmune diseases such as Chron's disease, Irritable Bowel disease, and Psoriasis.

Genome-wide methods for detecting selection

SMBE-PO-249

Large-scale detection of molecular adaptation in poison arrow frogs using Vespa-Slim

B. Constantinides [1 2,*], M. J. O'Connell [3 4], K. S. Ting [5], C. Creevey [5], A. Webb [6]

[1]Modernising Medical Microbiology, University of Oxford, Oxford, [2]Computational and Molecular Evolutionary Biology Group, University of Leeds, Leeds, [3]Computational and Molecular Evolutionary Biology Group, University of Nottingham, Nottingham, [4]Bioinformatics and Molecular Evolution Group, University of Leeds, Leeds, [5]School of Biological Sciences, Queen's University Belfast, Belfast, United Kingdom, [6]Bioinformatics and Molecular Evolution Group, Dublin City University, Dublin, Ireland

Abstract: Conducting large-scale selective pressure variation analyses of protein coding sequences involves numerous preparatory operations, from identifying and aligning gene families to reconstructing phylogenies, before performing and interpreting a series of statistical tests for selective pressure variation. We have designed a software pipeline capable of performing selective pressure analyses on a large scale and we demonstrate it's utility on novel sequencing data from eight species gathered in the Peruvian Amazon comprising 2,751 gene families. Poison arrow frogs exhibit remarkable adaptation to sequester alkaloid from their diet while at the same time presenting resistance to their toxic diet. Recent studies of a voltage-gated sodium channel family have reported a signatures of adaptive evolution on mutations conferring resistance to at least batrachotoxin (BTX) in some of the most poisonous species. However, the mechanisms involved in synthesising and transporting BTX to the skin are poorly understood and are likely to have undergone further key molecular adaptations. Analysing the 2,751 gene families extracted from novel sequence data from poison frogs, we set out to determine adaptive events contributing to the unique life history traits of these frogs. We present our findings together with our software pipeline 'Vespa-slim', a lightweight version of our toolkit for genome-scale detection of selective pressure variation.

Genome-wide methods for detecting selection

SMBE-PO-250

Modeling the effects of gene flow, divergent selection and background selection on genome scans in haplodiploids and sex-chromosomes

J. C. Frazão [1], E. E. Bendall [2], C. R. Linnen [2], V. C. Sousa [1,*]

[1]centre for ecology, evolution and environmental changes, University of Lisbon, Lisboa, Portugal, [2]Department of Biology, University of Kentucky, Lexington, KY, United States

Abstract: There is increasing evidence that gene flow between populations adapted to different environments is widespread across many taxa. Genomic scans of differentiation are now widely used to detect outlier regions, potentially under divergent selection. However, background selection can confound these analysis as it might also leads to peaks of differentiation. Differentiation is also expected to differ between diploid and haplodiploid species, but little is know about this effect. Because in haplodiploics and sex-chromosomes, recessive mutations are exposed to selection in the haploid sex, pervasive linked selection may produce exceptionally heterogeneous genomic differentiation. Here, we model the joint effects of gene flow, divergent selection and background selection using simulations. Our preliminary results suggest that haplodiploids have higher peaks of divergence and greater among site variance compared to diploids, irrespective of the dominance of beneficial mutations. For models with recessive and slightly deleterious mutations, we find that diploids are affected by associative overdominance, creating valleys of differentiation not seen in haplodiploids. To test predictions from simulations we used genome-wide SNP data from hybridizing haplodiploids (*Neodiprion lecontei* and *Neodiprion pinetum;* both pine sawflies of order Hymenoptera), and a meta-analysis with data from autosomes and sex-chromosomes. We find that haplodiploids and sex-chromosomes have higher mean and variance in F_{ST} compared to autosomes, consistent with a major impact of divergent selection in creating heterogeneous genomic patterns. This suggests that divergent selection is more efficient in haplodiploids in scenarios with gene flow.

Genome-wide methods for detecting selection
SMBE-PO-245
Empirical Evidence of GC-Bias Gene Conversion in Bacteria
S. Barreto [1 2,*], F. Bertolla [2], V. Daubin [1], L. Duret [1]
[1]LBBE, Lyon 1 University, [2]Microbial Ecology Laboratory, Lyon 1 University, Villeurbanne, France

Abstract:
The determinant of genomic base composition in organisms is a much-debated question. Some authors have considered base composition as a trait that is in itself subject to natural selection while others argued that it is the (sole) result of mutational equilibrium. Especially in Archaea and Bacteria, this trait displays a wide range of variability, which root causes remain to be convincingly elucidated.

In Eukaryotes and particularly in Mammals, it is now widely acknowledged that recombination is an important driver of base composition, through a mechanism known as GC-Biased Gene Conversion (gBGC), which leads to a local increase in GC-content in recombination hotspots, that blurs evidence of selection and can even go counter to it. Recently, this characteristic footprint has been detected in bacterial genomes as well. While it provides indirect evidence for the presence of a recombination-associated GC bias in Bacteria, direct experimental measures have yet to put this to the test.

We thus analyzed base conversion frequencies in nearly 5000 experimentally obtained recombination products spanning four phylogenetically disparate bacterial species, ranging from around 500 to 2000 products by species. We found a significant transmission bias of G and C bases in three of four species analyzed; the fourth species showed evidence of over-transmission of A and T bases. It constitutes to our knowledge the first empirical evidence of a GC-bias during gene conversion in Bacteria. The phylogenetic distribution of these three species suggests that gBGC could be widespread in Bacteria as well.

Genome-wide methods for detecting selection

SMBE-PO-258

Rise and fall of chemosensory genes sheds light on host-preference evolution in pollinating fig wasps

B.-W. Lo [1,*], H.-Y. Wang [2]

[1]Institute of Ecology and Evolutionary Biology, [2]Graduate Institute of Clinical Medicine, National Taiwan University, Taipei, Taiwan

Abstract: Pollinating fig wasps (Agaonidae) have one of the most reduced chemosensory genes in insects, which probably associated with host specialization in obligate mutualism. On the other hand, olfaction plays a crucial role in maintaining host specificity in the fig-fig wasp coevolution. To understand how reduced chemosensory genes maintain host-specificity during species divergence, we sequenced the genomes and transcriptomes of *Wiebesia pumilae* and its closely related species *W.* sp., both of which were recently codiverged with their hosts, *Ficus pumila var. pumila* and *Ficus pumila* var. *awkeotsang*, respectively, and compared them with the published genome of *Ceratosolen solmsi*. As expected, pollinating fig wasps showed severe gene reduction in three out of the five chemosensory gene families, including olfactory receptors (OR), gustatory receptors (GR), and odorant binding proteins (OBP). Nevertheless, we found lineage-specific OR gene duplication events in two chromosome regions. Evidence of positive selection were more frequently found on duplicated genes than on non-duplicated ones. Between species differentially expressed genes were enriched for chemosensory related gene ontologies. All the differentially expressed chemosensory genes were single copy in Agaonidae. Our results indicated that lineage-specific duplication of OR accompanied by selection is associated with broad-scale co-cladogenesis in Agaonidae. As for the fine-scale evolution, we found changes in expression profiles of chemosensory genes is associated with local adaptation during co-speciation. Our result provides new understanding in the possible underlying mechanism of host-preference evolution in pollinating fig wasps.

Genome-wide methods for detecting selection
SMBE-PO-262
Statistical inference of the origin of natural selection in an admixed population
N. Cooke [1,*], D. Bradley [2], S. Nakagome [1]
[1]Psychiatry, [2]Genetics, Trinity College Dublin, Dublin, Ireland

Abstract: The impact of natural selection on beneficial alleles can be observed in modern human genetic variation; however deciphering the origins of these alleles is made difficult by the vast complexity of human history. Here we describe a new statistical framework of Approximate Bayesian Computation (ABC) that can detect which ancestral lineage an allele undergoing selection first appeared. To demonstrate this framework we assume a model based on the history of modern Japanese, in which the present-day population derives ancestry in unequal proportions from two groups, the Jomon and the Yayoi, who admixed with each other 3,000 years ago, having previously diverged 30,000 years ago. By comparing the present-day genetic variation observed at a selected allele to data simulated within our model, we test if our approach can accurately predict which ancestral population it first emerged in. In this presentation, we demonstrate the power of our framework using a simulation study and the application of our approach to modern Japanese genomic data.

Genome-wide methods for detecting selection

SMBE-PO-257

Natural selection on mutations affecting human complex traits

Y. B. Simons [1,*], G. Sella [2 3 4]

[1]Genetics, Stanford University, Stanford, [2]Biological Sciences, [3]Systems Biology, [4]Program for Mathematical Genomics, Columbia University, New York, United States

Abstract: Many phenotypes of interest are genetically complex, in that heritable variation in the phenotype arises from numerous segregating loci distributed across the genome. While complex traits and their evolution have been studied for many decades, we still know little about key parameters that underlie their evolution. Here, we leverage human genome-wide association studies (GWAS) results to infer the distribution of selection effects and target size of mutations affecting human morphological, biomedical and life history complex traits, to our knowledge for the first time. Specifically, we build on our recent theoretical work to show that the population frequency of a trait-affecting variant imposes an upper bound on the strength of selection acting on it, while its effect size on the trait imposes a lower bound. We then construct a smoothed, maximum composite-likelihood estimator that relies on all the genome-wide significant (GWS) associations for a trait to infer the distribution of selection coefficients of the variants they tag. We use this distribution to estimate the rate of mutations affecting the trait and their distribution of selection coefficients. Simulations suggest that our method performs well with >100 GWS associations, providing precise estimates of selection coefficients in the range of $10^{-5}<s<10^{-2}$. We apply our method to UK Biobank GWAS, focusing on traits with non-negligible explained heritability ($h_g^2>0.05$), including height, BMI, male pattern boldness, age at menarche and systolic blood pressure. With parameter estimates in hand, we predict how the number of associations and explained heritability for these traits should increase with GWAS sample sizes.

Genome-wide methods for detecting selection

SMBE-PO-255

Evidence for widespread balancing selection in human populations

V. Soni*, M. Vos [1], A. Eyre-Walker [2]

[1]European Centre for Environment and Human Health, University of Exeter, Exeter, [2]EBE, University of Sussex, Brighton, United Kingdom

Abstract: The role that balancing selection plays in the maintenance of genetic variation remains unknown. In humans, the signature of balancing selection has been detected at various genomic localities, but the overall frequency of balancing selection has yet to be quantified. Recent studies have identified loci under long term balancing selection in primates by identifying polymorphisms shared between two species. These studies have focused on cases in which the two species are extremely unlikely to share a neutral polymorphism (i.e. they are sufficiently divergent that all polymorphism that was present in the ancestor of the two species will have gone to fixation in at least one of the species). This makes these tests weak because balancing selection must persist for a long time.We present a new simple method whereby neutral polymorphisms are used to inform us what to expect under neutrality by comparing the number of polymorphisms shared between two populations, at putatively functional sites with those at putatively neutral sites. Through simulations we have proven that our statistic has power to detect balancing selection. We have applied this statistic to human population genetic data taken from the 1000 genome project. We estimate that approximately 24% of all non-synonymous polymorphisms shared between African and non-African human populations are being maintained by balancing selection; this equates to approximately ~10,000 non-synonymous balanced polymorphisms.

Genome-wide methods for detecting selection

SMBE-PO-256

Ignore at your own peril: synonymous rate variation among sites in protein coding sequences.

S. Wisotsky [1 2,*], S. Muse [2], S. Kosakovsky Pond [1]

[1]Temple University, Philadelphia, [2]North Carolina State University, Raleigh, United States

Abstract: Statistical models are crude approximations of the biological processes they describe. While some of these approximations are essential for computational tractability and statistical performance, others remain in use largely due to convention and inertia. The cost and consequences of many assumptions have not been systematically explored. For instance, it has been assumed that the rate at which synonymous substitutions occur within genes does not vary from site to site. This assumption remains widespread despite accumulating evidence that transcriptional, translational, structural, genomic, and other processes render many synonymous changes non-neutral, and by extension, modulate site-specific evolutionary rates. In this study we systematically examined whether including synonymous rate variation (SRV) in evolutionary models is important for statistical (goodness of fit) and inferential (are biological conclusions affected) reasons. Parametric simulations convincingly demonstrate that codon-based analyses of natural selection are strongly affected by SRV: not accounting for SRV in models when it is present in the data, even at moderate levels, leads to runaway rates of false positives. We modified an existing method, BUSTED, that tests for gene-wide episodic positive selection to account for SRV and correct statistical misbehavior on simulated data, calling the resulted method BUSTED[S]. We then tested both methods on a database of mammalian gene alignments curated for studies of positive selection, Selectome, and found that nearly 50% of the genes detected to be under positive selection can be misclassified due to unmodeled SRV. Because the computational cost of including SRV is minimal (increasing the run time of the analysis by 3- to 5-fold) we suggest that those interested in detecting gene-wide evidence of positive selection should make the switch to SRV inclusive methods.

Genome-wide methods for detecting selection

SMBE-PO-268

Contemporary evolution in human: genome-wide signatures of ongoing purifying and balancing selection at nonsynonymous sites

X. Wei [1,*], Z. Pan [2], R. Nielsen [1]

[1]Integrative Biology and Statistics, UC Berkeley, Berkeley, United States, [2]Sun Yat-Sen University, Guangzhou, China

Abstract: The release of 500,000 genomes from the UK Biobank (UKB) provides unprecedented opportunities for studying ongoing selection using deviations from Hardy-Weinberg Equilibrium (HWE). Compared to the HWE expectation, there is a genome-wide pattern of increased heterozygosity of the British ancestry individuals in the UKB cohort. This observation could potentially be caused by genotyping errors. However, we find that low Minor Allele Frequency (MAF <5%) nonsynonymous SNPs have increased deviations from HWE compared to other SNPs of the same MAF. This observation cannot be explained by genotyping errors, but is likely due to purifying selection against recessive deleterious allele. Perhaps surprisingly, low MAF archaic SNPs from Neanderthal admixture show fewer deviations from HWE, a pattern that can be replicated in simulations incorporating selection and addition, by mutation, of new deleterious alleles, for many generations after the time of admixture. We also find that high MAF (30-50%) nonsynonymous SNPs have increased heterozygosity compared to other SNPs of the same MAF. These SNPs are enriched for genes previously identified to be under balancing selection and they also show evidence from survivorship curves of decreased mortality for heterozygous individuals. Population genetic simulations under realistic parameter settings can recapitulate these observations. Our study demonstrates the power of using deviations from HWE to detect selection in large cohorts and that ongoing selection in humans is common.

Genome-wide methods for detecting selection

SMBE-PO-266

Population genomics of the opportunistic plant pathogenic bacterium Ralstonia solanacearum in the UK: evolutionary analysis in space and time

M. Stoycheva [1,*]

[1]Biology, University of York, York , United Kingdom

Abstract: Bacterial strains belonging to *Ralstonia solanacearum* species complex are important plant pathogens being capable of infecting several host plants in more than 50 plant including important crops such as potatoes, tomatoes and bananas. As a result, this bacterium has been ranked as the second most important plant bacterial pathogen due to its scientific and economic importance. *Ralstonia solanacearum* can also survive in various natural environments including rivers, soils or the rhizosphere of secondary hosts from where it often spreads to agriculturally important environments. Comprehensive genomic analysis exploring the evolution and dispersal of this pathogen at global scale is still lacking. We will present whole genome population genomic analysis of 384 *R. solanacearum* samples from 66 countries, various hosts and dating back to the 1945. This will be combined with detailed fine grain spatial and temporal analysis of *R. solanacearum* outbreaks in the United Kingdom from annual river sampling surveys. We expect that the evolutionary divergence of this heterogenous species will be evident at both local and global spatial scales. This could be driven by selection pressures experienced in the environmental reservoirs versus crop host plants. Extensive analysis of the *R. solanacearum* pangenome is expected to reveal further genetic changes associated with agricultural trade routes and human-mediated transmission and control interventions. Together, these analyses will help us to understand plant pathogen evolution in space and time at both global and local scale and link them with persistence in different environments and primary and secondary plant hosts.

Genome-wide methods for detecting selection

SMBE-PO-237

Evaluating Tests for Positive Selection based on Codon Models with Multiple Nucleotide Changes

L. C. Mikula [1,*], C. Kosiol [1]

[1]Centre for Biological Diversity, University of St.Andrews, St.Andrews, United Kingdom

Abstract: Within the last decade, the acquisition of large amounts of genomic data has given rise to an abundance of methods aiming to disentangle evolutionary mechanisms. As the main force behind adaptive evolution, positive selection has garnered a lot of attention. The Yang-Nielsen branch-site test has become standard for detecting positive Darwinian selection on coding DNA sequences in a phylogenetic context. It contrasts a model for codon evolution that permits positive selection on an priori specified foreground lineage with one that assumes neutral/constrained evolution across all lineages. The model specification is based on the nonsynonymous-synonymous substitution rate ratio, a classic means of detecting the signature of positive selection. Nucleotide changes are assumed to occur singly and independently.

More recently it has been claimed that excess nonsynonymous changes due to fixation of neutrally evolving multiple nucleotide mutations can therefore lead to false positive results of thhe branch-site test. Newer approaches to codon model based neutrality tests have striven to address this issue: A mechanistic modification to the underlying codon model involves the introduction of a double-nucleotide to single-nucleotide substitution rate parameter. Alternatively, simplified empirical codon models estimate exchangeabiltiy parameters both individually for single nucleotide changes and grouped by multiplicity and effect on transcription for multiple nucleotide changes via maximum likelihood methods.

I aim to compare the sensitivity and power of these newer models via simulations based on great ape data using the programming packages PAML and HYPHY. Furthermore, I wish to examine the role of amino acid properties and their impact on results.

Genome-wide methods for detecting selection

SMBE-PO-271

Can signals of balancing selection be identified in parasite genomes despite the influence of demography?

S. Forrester [1,*], J. Carnielli [1], V. Costa [2], S. James [3], C. Costa [2], J. Mottram [1], D. Jeffares [1]

[1]York Biomedical Research Institute, University of York, York, United Kingdom, [2]Laboratório de Pesquisas em Leishmanioses, Universidade Federal do Piauí, Teresina, Brazil, [3]Genomics facility, University of York, York, United Kingdom

Abstract: Balancing selection (BS) is important in the evolution of parasite genomes, and often occurs in regions of the genome responsible for host: parasite interactions. Evidence of this has previously been observed from the protozoan Plasmodium genus, where membrane proteins were shown to have excess diversity (Nygaard *et al*., 2010; Volkman *et al.,*2002). We examined the evidence for BS in Leishmania *infantum*and *L. donovani*, intracellular protozoans responsible for visceralising leishmaniasis; a neglected infectious disease that is associated with poverty with foci in East Africa, the Indian subcontinent and Brazil (WHO, 2018). Using population genomics data from populations of *L. infantum*/*donovani*in Ethiopia, Brazil and the Indian subcontinent (ISC), we called SNPs and indels to allow for comparative analysis. Demography was vastly different in these three regions. Between the most diverse region, Ethiopia, and the least diverse, Brazil, there is a 100 fold range of nucleotide diversity (θ_π= 1.7 x10^{-3}in Ethiopia, θ_π= 1.8 x10^{-5}in Brazil), consistent with Ethiopia being the oldest population. The distribution of summary statistics (θ_π, MAF and Tajima's D) in each population indicates population structure and strong effects from demography. Admixture cross validation (CV) was performed to identify 2 populations in Brazil, 3 core populations and one hybrid in Ethiopia, and 6 in India; these correspond with PCA analysis. We also observe tenfold differences in SNP/INDEL ratios, consistent with strong influences of demography and population structure on deleterious alleles. We show that balancing selection is likely unrelated to gene function, or very strongly affected by demography as neither Tajima's D nor Mc Donald Kreitman test neutrality index ratios are correlated between sister species/populations. We will present the results of model-based analysis of balancing selection from these three populations of Leishmania (Cheng & DeGiorgio, 2018), as a more definitive test of consistent vs. demography-affected diversity patterns. Finally, we will present localisation of GFP-tagged candidate genes in macrophages.

Genome-wide methods for detecting selection
SMBE-PO-261
Protein analysis as a tool to expose signals of evolutionary adaptation
J. M. Moreno [1,*], M. M. Coelho [1]
[1]cE3c - Centre for Ecology, Evolution and Environmental Changes, Lisboa, Portugal

Abstract: Detecting signatures of selection at molecular level has been a challenge when studying adaptation of populations. Most widespread tests to detect these signatures rely on the ratio of non-synonymous to synonymous mutations in DNA sequences. However, these tests do not always present suitable statistical power to detect signatures of selection, while in other scenarios, signatures detected may not be functionally relevant. Here, we present a methodology combining sequence-based methods for detecting signatures of selection together with protein characterisation at their different structural levels of organisation in order to better expose signals of evolutionary adaptation. Protein characterisation involves analysis of several physicochemical parameters, prediction of functional features as post-translational modifications, structural analysis, and an upper-level investigation of protein-protein interactions. We analysed four genes from the circadian-related CRYPTOCHROME gene family in different populations of the freshwater fish species from the genus *Squalius*, distributed in Iberian rivers along a latitudinal cline of light and temperature. We found evidences of positive selection acting on these genes, but only integration of results from protein analysis with selection analysis allowed a proper characterisation of adaptive patterns of different populations according the environmental conditions. Hence, we present this approach as a major tool to further studies in molecular evolution to detect adaptive events.

Genome-wide methods for detecting selection

SMBE-PO-264

GWAS of phylogenetically independent fungal proteins

I. Riquelme Medina [1,*], J. O. McInerney [2], F. Whelan [2]

[1]Evolution & Genomic Sciences, University of Manchester, Manchester, [2]School of Life Sciences, University of Nottingham, Nottingham, United Kingdom

Abstract: Fungi are a diverse group of heterotrophic Eukaryotic organisms. Fungi live in a wide range of habitats, and have multiple different lifestyles, including as saprobes (surviving off of other decaying organisms), or as symbionts and/or parasites of plants (e.g. mycorrhiza), animals, and other fungi. These diverse lifestyles can be seen even within closely related fungal species; for example, the family *Onnia* includes species which live as saprobes, mycorrhiza, and plant parasites. Because of the large amount of variability within closely related fungal groups, these organisms have had to repeatedly adapt to their environment at multiple points in their phylogenetic history. In this work, we asked whether fungal species which are phylogenetically widespread but have similar lifestyles, adapt to these lifestyles using different or similar genetic strategies.

We approached this question by looking for genes which appear together across different species in the fungal phylogeny more often than we would expect by chance (co-occurrence), with the hypothesis that co-occurring genes would be dually required for expressing a particularly phenotype. From these gene sets, we have focussed on those that co-associate with particular fungal lifestyles. Further, we have also explored genes which avoid each other with the hypothesis that the presence of these two genes together would be detrimental for particular lifestyles. We built software to find these co-occurrence and -avoidance patterns while taking phylogenetic signal (lineage dependence) into account by scoring the significant gene-gene relationships according to whether they are independent of the phylogenetic relationship (for e.g. co-occurring genes found on several branches of the tree are considered more significant than those only on one branch).

By using this approach we have found hundreds of lineage independent co-occurring clusters of genes that we will further analyze to determine which clusters can be linked to the fungi lifestyles.

Genome-wide methods for detecting selection

SMBE-PO-265

Strong selective sweeps have helped shape patterns of genetic diversity in murid rodents

T. R. Booker [1,*], B. C. Jackson [2], R. J. Craig [3], B. Charlesworth [3], P. D. Keightley [3]

[1]Department of Forestry and Conservation Science, University of British Columbia, Vancouver, Canada, [2]Department of Genetics, University of Cambridge, Cambridge, [3]Institute of Evolutionary Biology, University of Edinburgh, Edinburgh, United Kingdom

Abstract: Over the last three decades, understanding the contribution of positive and negative selection to patterns of genetic variability in eukaryotic organisms has become a major goal of evolutionary genomics. New mutations with advantageous or deleterious fitness effects can induce selective sweeps and background selection, respectively, and both processes may contribute to variation in neutral genetic diversity at linked sites. Here, we examine the profiles of genetic variability around protein-coding and regulatory elements in the genomes of wild mice, with samples from the three principal *Mus musculus* sub-species as well as *Mus spretus.* We find that reductions in genetic diversity near functional elements, a pattern consistent with the effects of selection at linked sites, are similar across these taxa despite inferred differences in effective population size and demographic history. Using a model which combines the effects of selective sweeps and background selection, we estimate selection parameters for advantageous mutations that are compatible with the observed reductions in diversity and compare these among taxa. One conclusion of this study is that strong positive selection is required to explain variation in genetic diversity across the genomes of murid rodents.

Genome-wide methods for detecting selection

SMBE-PO-263

Quantifying codon usage in signal peptides: Gene expression and amino acid usage explain apparent selection for inefficient codons

A. L. Cope [1,*], R. L. Hettich [2], M. A. Gilchrist [3]

[1]Genome Science and Technology, University Of Tennessee, Knoxville And Oak Ridge National Laboratory, Knoxville,TN, [2]Chemical Sciences Division, Oak Ridge National Laboratory, Oak Ridge, TN, [3]Department of Ecology and Evolutionary Biology, University Of Tennessee, Knoxville, Knoxville, TN, United States

Abstract: Codon usage is shaped by both selection and mutational biases, with the relative strength of these forces largely determined by gene expression. ROC-SEMPPR, a mechanistic model rooted in population genetics, is able to differentiate between the selective and mutational biases shaping codon frequencies by accounting for the variation in gene expression across genes. Here, we show how ROC-SEMPPR can be used to test for selective differences on codon usage between different regions of genes. We test for differences in selection on codon usage in signal peptides (short N-terminal peptides serving as markers for protein secretion) of *E. coli*. Previous work found an increased bias for inefficient codons in signal peptides relative to the 5'-ends of cytoplasmic genes. This was interpreted as evidence for selection for translation inefficiency, the argument being translation inefficiency in the signal peptide region could improve the efficiency of protein secretion. Using ROC-SEMPPR, we find the strength and direction of selection on codon usage in signal peptides is consistent with the 5'-ends of cytoplasmic genes. We also use ROC-SEMPPR to test for a general weaker selection on codon usage at the 5'-ends of genes, regardless of the presence/absence of a signal peptide. Consistent with previous work, we find selection on codon usage is generally weaker at the 5'-ends, with the strength of selection gradually increasing along a gene. Furthermore, we show other common methods of codon usage analysis are heavily biased by both amino acid usage and gene expression on data simulated under the same parameters.

Genomic perspectives on plant and animal domestication

SMBE-PO-275

Unexpected patterns in the relative genetic diversity of autosomes and sex chromosomes in cattle

R. R Da Fonseca*, I. Ureña, S. Afonso, A. E. Pires, E. Jørsboe, L. Chikhi, C. Ginja

Abstract: In mammals the lower effective population size in chromosome X relative to the autosomes (one copy in males) is expected to result in the nucleotide diversity on sex chromosome X being around 75% of that on autosomes. However, in cattle, the effective size of the X is increased by the practice of keeping very few reproducing males per herd, so its nucleotide diversity might be shifted to higher values. We have analyzed the whole genomes from 128 individuals, including low-depth sequencing data from eight primitive European cattle breeds, whole genome data from three European commercial cattle, one African taurine and three zebu breeds, to uncover genomic patterns associated with the different breeding contexts.

The nucleotide diversity in sex chromosome X ranged from 35 to 60% of that of autosomes, which is much lower than expected. Furthermore, the impact of a bottleneck or population structure, potentially caused by breeding practices, should be higher on chromosome X and genetic drift would be expected to result in higher F_{ST} values for chromosome X relative to autosomes. However, comparisons within taurine and within indicine cattle show a much higher F_{ST} for autosomes than for chromosome X. This agrees with extensive male-biased gene flow within taurine and within indicine – since males have a single copy of chromosome X, introgression will be more efficient on the autosomes. It is known that female populations are more likely to be geographically constrained and human-driven crossbreeding may have been carried out mainly using males. This could also explain the difference in ancestry assignments for autosomes and chromosome X, with signatures of previously described indicine admixture in the African taurine autosomes, but not observed in chromosome X.

Genomic perspectives on plant and animal domestication
SMBE-PO-274
COPY NUMBER VARIATION IN DIVERSE RICE POPULATIONS
Z. N. Lye [1,*], M. D. Purugganan [1 2]
[1]Center for Genomics and Systems Biology, New York University, New York , United States, [2]Center for Genomics and Systems Biology , New York University Abu Dhabi, Saadiyat Island, Abu Dhabi, United Arab Emirates

Abstract: Rice is a key crop for world food security and is the staple food for over half of the world's population. Rice has also served as an excellent model for studying evolution and the process of domestication. In the past domestication has been primarily studied in terms of single nucleotide polymorphisms (SNPs). Copy number variation (CNV), is the gain, loss or duplication of genomic sequences between individuals of a population. CNVs have been largely neglected in studies of domestication and are a source of genetic variation during domestication, and during the diversification of domesticated taxa. Here we characterize CNVs in diverse race landraces representing all major rice sub-populations. We emphasize genic CNVs which can illuminate functional and evolutionary processes that vary among individuals and subspecies.

Genomic perspectives on plant and animal domestication

SMBE-PO-281

Time-dependent molecular evolution in ancient DNA

A. Lin [1,*], L. Frantz [2], S. Ho [3], G. Larson [1]

[1]University of Oxford, Oxford, [2]Queen Mary University of London, London, United Kingdom, [3]University of Sydney, Sydney, Australia

Abstract: Introduction

Estimating rates of molecular evolution is necessary in order to infer demographic dynamics and evolutionary timescales from genetic data that had been sampled at specific points in time. For phylogenetic analysis of time-structured data, ages of the ancient DNA sequences can be drawn from radiocarbon-dated samples to calibrate the molecular clock. The hypothesis of time-dependent molecular evolution states that the rate of observable evolution varies depending on the timeframe over which the rate is measured. Domestic animals and their wild progenitors are an excellent to proxy to test the hypothesis of time dependency given their relatively short generation time and their ubiquity in the archaeological record.

Objectives

The hypothesis of time-dependent molecular evolution will be tested by comparing the directly-dated mitochondrial genomes of multiple ancient archaeological samples across a balanced temporal distribution. By analysing these data using dated molecular clock analyses, changes in the molecular substitution rate across different time scales can be estimated.

Methods

Hundreds of samples from at least five vertebrate species used for analyses were either radiocarbon dated or are associated with archaeological sites with high confidence. From these directly-dated samples, high-coverage full mitochondrial genomes used for analysis. We used different phylogenetic methods in estimating the rates from time-structured data, including root-to-tip regression, scalable relaxed clock dating, approximate Maximum Likelihood inference, and Bayesian inference.

Results

Preliminary results from ancient suid and bison datasets suggest evidence of time-dependent molecular evolution within timeframes that span between the present day to 14,000 BP, and from the present day to 128,000 BP, respectively. There is a negative relationship between rate and time, where as the sample age increases, there is a corresponding decline in estimated substitution rate.

Conclusion

To our knowledge, this study will be the largest meta-analysis yet using ancient DNA in testing the hypothesis of time-dependent molecular evolution, and preliminary results suggest that the observed time-dependency is a biological phenomenon. In addition, this study may yield interesting observations on the genetic processes of domestication within and between the different animal populations of study, as well as inform and improve fossil calibration of current molecular clock models.

Genomic perspectives on plant and animal domestication

SMBE-PO-283

Genomic Perspective towards the Domestication of Guinea Fowl in Africa

M.-S. Peng [1 2 3,*], A. C. Adeola [1 2], Q.-K. Shen [1 2 3], S. Duan [4], Y.-W. Miao [5], Y. Du [4], M.-S. Wang [1 6 7], J. K. Lichoti [8], O. S. Charles [9], O. J. Sanke [10], P. M. Dawuda [11], A. O. Okeyoyin [12], J. Musina [13], P. Njoroge [13], B. Agwanda [13], M. G. Strillacci [14], E. Gorla [14], A. Bagnato [14], S. Kusza [15], H. A. Nanaei [16], R. Pedar [16], N. T. Abdulloevich [17], M. E. Afanasevna [17], K. B. Ibrohimovich [17], S.-F. Wu [1], X. Chen [18 19], W.-K. Yang [18 19], N. O. Otecko [1 2 3], R. W. Murphy [20], L. M. Nneji [1 2], A. Esmailizadeh [1 16], Y. Dong [21 22 23], S. C. Ommeh [13 24], Y.-P. Zhang [1 2 3 25 26]

[1]State Key Laboratory of Genetic Resources and Evolution, Kunming Institute of Zoology, CAS, Kunming, China, [2]Sino-Africa Joint Research Center, CAS, Nairobi, Kenya, [3]Kunming College of Life Science, UCAS, [4]Nowbio Biotechnology Company, [5]Faculty of Animal Science and Technology, Yunnan Agricultural University, Kunming, China, [6]Howard Hughes Medical Institute, UCSC, [7]Department of Ecology and Evolutionary Biology, University of California Santa Cruz, Santa Cruz, United States, [8]State Department of Livestock, Ministry of Agriculture Livestock Fisheries and Irrigation, Nairobi, Kenya, [9]Department of Veterinary Medicine, University of Ibadan, Ibadan, [10]Taraba State Ministry of Agriculture and Natural Resources, Jalingo, [11]Department of Veterinary Surgery and Theriogenology, University of Agriculture Makurdi, Makurdi, [12]National Park Service Headquarter, Abuja, Nigeria, [13]Department of Zoology, National Museums of Kenya, Nairobi, Kenya, [14]Department of Veterinary Medicine, University of Milan, Milan, Italy, [15]University of Debrecen, Debrecen, Hungary, [16]Department of Animal Science, Faculty of Agriculture, Shahid Bahonar University of Kerman, Kerman, Iran, Islamic Republic Of, [17]E.N. Pavlovsky Institute of Zoology and Parasitology, ASRT, Dushanbe, Tajikistan, [18]Research Center for Ecology and Environment of Central Asia, CAS, [19]Key Laboratory of Biogeography and Bioresource in Arid Land, Xinjiang Institute of Ecology and Geography, CAS, Urumqi, China, [20]Centre for Biodiversity and Conservation Biology, Royal Ontario Museum, Toronto, Canada, [21]College of Biological Big Data, [22]State Key Laboratory for Conservation and Utilization of Bio-Resources in Yunnan, [23]Key Laboratory for Agro-biodiversity and Pest Control of Ministry of Education, Yunnan Agricultural University, Kunming, China, [24]Jomo Kenyatta University of Agriculture and Technology, Nairobi, Kenya, [25]State Key Laboratory for Conservation and Utilization of Bio-resource in Yunnan, Yunnan University, [26]Center for Excellence in Animal Evolution and Genetics, CAS, Kunming, China

Abstract: The domestication of the helmeted guinea fowl (*Numida meleagris*) in Africa remains an enigma. Herein, we report the *de novo*assembly of the domesticated helmeted guinea fowl genome and new whole-genome sequences for 135 birds including domesticated animals, wild progenitors and closely related species. Comparative genomic analyses reveal signals of selection on genes related to pathogen resistance, showing adaptation to harsh environments in Africa. Inference on population history analysis indicates an origin in West Africa around 4,000–7,000 years ago. Scanning for selective sweeps detects a strong candidate gene *GRIK1*for neural change and behavior shift in domestication. The domestication scenarios are mirrored by the patterns of pearl millet (*Cenchrus americanus*) domestication in West Africa. In addition, we find *TYR*contributes to white plumage during the recent breeding in Italy. Our study provides the first case to investigate animal domestication in Africa based on genomic level.

Genomic perspectives on plant and animal domestication
SMBE-PO-284
Sequencing and analysis of Bronze Age Bos Taurus genomes
C. Rossi [1,*], D. G. Bradley [1], V. Mattiangeli [1], M. Mashkour [2]
[1]Smurfit Institute of Genetics, Trinity College Dublin, Dublin, Ireland, [2]Muséum National d'Histoire Naturelle Département Ecologie et Gestion de la Biodiversité, Paris, France

Abstract: The domestication of plants and animals is one of the most pivotal developments in human history, yet key events of the domestication process remain equivocal. Domestic animal archaeogenetics can now shed new light on human history of domestication by interrogating the shift in genomic identity of ancient domesticated animals through time. By virtue of next generation sequencing technologies, it is possible to generate DNA sequence of hundreds of thousands of loci, and even whole genomes, from archeological samples. With this genomic information, we can place ancient animals in a genetic context to give insights into their ancestry as well as the groups of early farmers who practised animal husbandry.

Thus far in my PhD, ancient DNA has been extracted from over two dozen petrous bones of ancient *Bos taurus* and screened for high endogenous DNA content. The sampled bones were recovered at Bronze Age sites in the Netherlands and Lebanon. Two ancient *Bos* samples have been selected for high coverage sequencing using Illumina HiSeq. These samples have been sequenced to an average of over 1X coverage using shotgun sequencing technology. After quality processing and alignment, their SNP data has been added to a curated matrix of modern cattle breeds (*Bos taurus*, *Bos indicus* and cross-breeds), from Europe, Africa, Asia and the Middle East. Using this matrix, the approximate genomic identity of the Dutch and Lebanese bovine samples have been visualised using Principal Components Analysis. The relative genetic influence of them have been further interrogated using D-statistics and outgroup *f*-3 statistics.

Genomic perspectives on plant and animal domestication

SMBE-PO-273

Dual distant hybridization components facilitated high altitude adaptation in Tibetan cattle

S. Wang [1,*], M.-S. Wang [1], Y. Zhang [2], R. Nielsen [3], D.-D. Wu [1], X.-D. Ding [2]

[1]State Key Laboratory of Genetic Resources and Evolution, Kunming Institute of Zoology, Chinese Academy of Sciences, Kunming, [2]National Engineering Laboratory for Animal Breeding, College of Animal Science and Technology, China Agricultural University, Bejing, China, [3]Department of Integrative Biology, University of California Berkeley, Berkeley, United States

Abstract: Adaptive introgression is recognized as an important mechanism of high-altitude adaptation. Recent studies have reported evidences suggesting that adaptive introgression play a significant role in high-altitude adaptation of Tibetan cattle. Our previous published research has found genetic components (for example, EGLN1) from yaks in Tibetan cattle, and these regions may help Tibetan cattle quickly adapt to the hypoxia environment on the Qinghai-Tibet plateau. To gain more insights into the complex component of Tibetan cattle, we analyzed 18 Apeijiaza Tibetan cattle and 19 Indian zebu whole genomes, together with other Chinese indigenous cattle and yaks. We found Apeijiaza population originates from admixture between bovine and zebu that occurred ~9 generations ago, in strong accordance with breeding records. In addition, there are ~17.8% zebu and ~0.1% yak ancestry in Apeijiaza. We analyzed the local ancestry across the genomes of Apeijiaza and identify some regions with high proportion of zebu ancestry that may be strongly selected. There regions with high zebu ancestry contained genes associated with skin cancer caused by ultraviolet radiation (for example, HRAS, BCL2, and TRIM32), that may be already selected in zebu originating from Indian with tropical monsoon climate. Furthermore, we identified a HIF-regulated tissue permeability factor (VASP) that has been introgressed from yak to Tibetan cattle, probably facilitating their adaptation to highland hypoxia environment. Our study uncovers the dual foreign genetic components in Tibetan cattle as an important source of high-altitude adaptation, and provides compelling example into introgression-mediated adaptation of extreme environments.

Genomic perspectives on plant and animal domestication

SMBE-PO-279

Recent host tracking of the fungal pathogen Cercospora beticola during domestication of sugar beet, Beta vulgaris

L. Potgieter [1 2,*], A. Feurtey [1 2], R. de Jonge [3], M. Varrelmann [4], M. McMullan [5], M. Bolton [6], E. Stukenbrock [1 2]

[1]Environmental Genomics, MPI for Evolutionary Biology, Plön, [2]Environmental Genomics, Christian Albrechts University, Kiel, Germany, [3] Plant-Microbe Interactions, Utrecht University, Utrecht, Netherlands, [4]Institut für Zuckerrübenforschung an der Universität Göttingen, Göttingen, Germany, [5]The Earlham Institute, Norwich, United Kingdom, [6]Northern Crop Science Laboratory, United States Department of Agriculture, Fargo, ND, United States

Abstract: *Cercospora beticola* is a fungal pathogen of sugar beet that causes the disease Cercospora Leaf Spot (CLS). Sugar beet is a relatively modern crop with a well documented domestication history in Central Europe. The ancestor of sugar beet is the wild beet species *Beta vulgaris* spp. *maritima* which can also be infected by *C. beticola*. Cultivated and wild beet species provide an excellent model system to study the effect of host domestication on pathogen evolution due to the recent domestication, and host range expansion of sugar beet. We collected isolates from *C. beticola* from wild and cultivated beet from several sites in Europe and the US. Sequencing the 37Mb haploid genomes of 150 isolates allowed us to identify population differentiation of pathogens on wild and cultivated host species suggesting divergent host specialization. Over all, the genomes of the *C. beticola* populations on the two distinct hosts harbour similar amounts of genetic variation, and are likely closely related. Nevertheless, the genomes of *C. beticola* also bear signatures of recurrent gene flow. We have identified regions that exhibit a higher extent of sequence divergence between *C. beticola* populations isolated from the wild and cultivated hosts. These regions include genes that may evolve under divergent selection during recent adaptation to the distinct hosts. Results from this study are beginning to shed light on the evolution of *C. beticola* associated with geographic domestication, beet breeding, and monoculture host populations.

Genomic perspectives on plant and animal domestication

SMBE-PO-277

Assessing the genetic contribution to behavioural traits in dogs

P. Wiener [1,*], J. Friedrich [1], P. Arvelius [2], E. Strandberg [3], R. Pong-Wong [1], A. Talenti [1], J. Hickey [1], M. Haskell [4]

[1]Roslin Institute, University of Edinburgh, Easter Bush, United Kingdom, [2]Swedish Armed Forces Dog Training Centre, Marsta, [3]Swedish University of Agricultural Sciences, Uppsala, Sweden, [4]Scotland's Rural College, Edinburgh, United Kingdom

Abstract: The processes of animal domestication and breed development in dogs involved numerous physiological changes, but in addition, led to important changes in behaviour. During the process of dog domestication from ancestral wolves, genetic changes in behaviour have allowed a close bond with humans. Further selection during breed development led to a wide range of behavioural characteristics. Thus there is great interest in identifying behaviour-associated genes in dogs. However, behavioural traits are complex phenotypes, which have been shown to be influenced by numerous genetic and environmental factors, complicating their analysis. We used principal components analysis to define behaviour traits for German Shepherd dogs (GSDs) based on dog owner responses to the Canine Behavioral Assessment and Research Questionnaire (C-BARQ) and additional questions about playfulness. While accounting for various environmental factors showing association with these traits, the genetic component was then investigated by heritability estimation and genomic analyses. Further analyses were performed to identify signatures of selection across the genome. Several behavioural traits exhibited moderate heritabilities, with the highest estimates identified for Playfulness and Non-social fear. We identified several genomic regions that showed genome-wide or chromosome-wide significant association with the analysed traits; for several traits, multiple regions were identified. Putative signatures of selection were identified on several chromosomes. Our results support the hypothesis that behaviour traits are influenced by multiple genes. However, the genomic analyses also pointed to a significant influence of specific genomic regions on several traits. Further analysis is required to investigate whether the putative signatures of selection are related to selection for behaviour.

Genomic perspectives on plant and animal domestication

SMBE-PO-280

Local ancestry and functional genomics of trypanotolerant and trypanosusceptible admixed African cattle breeds

G. P. McHugo [1,*], M. J. Dover [1], T. J. Hall [1], G. M. O'Gorman [2], E. W. Hill [1], D. E. MacHugh [1 3]

[1]School of Agriculture and Food Science, University College Dublin, Dublin, Ireland, [2]National Office of Animal Health Ltd., Enfield, United Kingdom, [3]Conway Institute of Biomolecular and Biomedical Research, University College Dublin, Dublin, Ireland

Abstract: African cattle represent a complex mosaic of *Bos taurus* (taurine) and *Bos indicus* (zebu) ancestry with most breeds containing varying levels of taurine-zebu admixture. The two types of cattle diverged at least 500,000 years ago and significant genomic differences have accumulated since that time. One important evolutionary adaptation in certain African taurine populations is a genetically determined tolerance to infection by protozoan trypanosome parasites (*Trypanosoma* spp.), which are transmitted by infected tsetse flies (*Glossina* spp.) and cause African animal trypanosomiasis (AAT) disease.

The annual financial burden of AAT is approximately $4.5 billion, and AAT is one of the largest constraints to livestock production in the areas of sub-Saharan Africa with significant tsetse densities. The West African taurine N'Dama breed is trypanotolerant; they have a capacity to control parasite loads and to limit disease pathology compared to trypanosusceptible zebu breeds. However, zebu or zebu-taurine hybrid animals are generally larger, produce higher milk yields and are therefore favoured by many farmers. Using local ancestry analysis of genome-wide high-density SNP data, we have examined hybrid West African cattle populations to study sub-chromosomal admixture. A sliding window approach was used and genes within 1 Mb up- and downstream from the top ancestry windows were used in gene set enrichment analyses. Results from this study demonstrate that the top physiological system development and function pathways for genes located within taurine local ancestry windows include immunobiology pathways, while the top zebu pathways are related to growth and development.

Genomic perspectives on plant and animal domestication
SMBE-PO-278
Post-bottleneck European polecat populations show high degrees of genome introgression.
G. J. Etherington [1,*], A. Ciezarek [1], W. Haerty [1], F. Di Palma [1]
[1]Organisms and Ecosystems, The Earlham Institute, Norwich, United Kingdom

Abstract: The domestic ferret (*Mustela putorius furo*) was domesticated from the European polecat (*M. p. putorius*) around 2500 years ago, probably to hunt rodents and small game. The European polecat is a widespread carnivore of much of the Western Palearctic. Once common across much of the United Kingdom, by the end of the 1800's persecution found it restricted to a small refugia in central Wales. At the same time, domestic ferret started to establish feral populations across the UK. Since then, legal protection has allowed the polecat to expand its range and is now found across much of its former territory. During this range expansion, European polecats came into contact with feral domestic ferret and hybridised.
Here we use whole genome sequencing to examine the degree of genome introgression in both the hybrid populations and British European polecats. We compared the genomes of European mainland polecats, British polecats, domestic ferrets and polecat-ferret hybrids to compare the degree of genome introgression within and between populations. We find that the degree of introgression varies between hybrids and that phenotypically wild British polecats, even those close to the original refugia, show genome introgression with domestic ferrets, whereas European polecats from mainland Europe do not.

Genomic perspectives on plant and animal domestication

SMBE-PO-282

Identifying sex-linked genes in Cannabis sativa provides biomarkers for early sexing of cannabis plants and new insights on plant sex chromosome evolution.

D. Prentout*, O. Razumova, H. Henri, B. Rhoné, C. Feng, J. Käfer, G. Karlov, G. Marais

Abstract: Growing *Cannabis sativa* for THC-production is becoming legal in an increasing number of countries and reducing production costs is an important goal. *C. sativa* is a dioecious plant with XX female and XY male individuals and only females are used for THC production. The cannabis Y chromosome has not been sequenced, which limits the development of methods for early sexing of cannabis plants. Using a novel method to identify sex-linked genes meant for organisms such as *C. sativa* in which sex chromosomes are poorly studied, we identified more than 500 sex-linked genes. From these, six Y-linked markers for early sexing of cannabis plants were validated using PCR on male and female plants from various hemp cultivars, cannabis cultivars which produces low THC rate. Using all of the sex-linked genes, we found that cannabis has strongly degenerated Y chromosome and represents one of the oldest plant sex chromosome system documented so far. Our work provides insights on sex chromosomes evolution in plants and also, almost ready-to-use markers for early sexing of cultivated cannabis plants.

Inside Africa: Uncovering patterns of human genetic diversity

SMBE-PO-287

Selection, metabolism and resistance to infectious diseases in Africa

L. Perreira*

Abstract: The continuous characterization of genome-wide diversity in population and case-cohort samples, allied to the development of new algorithms, are shedding light on host ancestry impact and selection events on various complex diseases, namely infectious diseases. Pathogens and metabolites have been identified as the main selection motors in human evolution and, in many instances, their role is closely intertwined. We will illustrate such an example in dengue fever disease, which was identified when we conducted joint ancestry and association tests in an admixed Cuban cohort characterized for 2.5 million SNPs. We identified African-ancestry protection against the hemorrhagic phenotype through two genes intervening in lipid metabolism. Lipids are essential for the virus to enter in human cells and replicate therein. Functional tests have confirmed the involvement of these genes in dengue disease, and open up new avenues for the development of therapies. Dengue virus is not per se a significant selective motor, as its associated mortality rate is low. Other related viruses, such as yellow fever virus, could have been the drivers of the local African adaptation against several related infectious diseases. This African protection continued to favour African-descendants in the new world environment, when dengue virus was there introduced in the 19th/20th century.

Inside Africa: Uncovering patterns of human genetic diversity
SMBE-PO-286
Structural variation analysis of diverse human populations identifies population-specific variants and sequences not present in the reference

M. Almarri [1,*], A. Bergström [1 2], J. Prado-Martinez [1], Y. Chen [1], S. McCarthy [1 3], C. Tyler-Smith [1], Y. Xue [1]
[1]Wellcome Sanger Institute, Hinxton, [2]Francis Crick Institute, London, [3]Department of Genetics, University of Cambridge, Cambridge, United Kingdom

Abstract: Whole genome sequencing projects have provided unprecedented insights into the evolutionary history of our species. While the majority of studies have focused on single nucleotide variants, structural variants, which include deletions, duplications and multiallelic copy number variants, contribute a greater diversity at the base level than any other class of variation. We have previously generated and analysed a high-coverage dataset of 911 samples from the Human Genome Diversity Project (HGDP-CEPH) panel, composed of 54 diverse populations. Here, we explore the diversity and structure of different classes of structural variants, performing population-specific analysis in addition to wider continental comparisons. We find that Oceanians harbour many almost fixed private variants, in contrast to other continental groups. For individual populations, we find multiple instances of high frequency variants that are not present in nearby populations. For example, we find a novel deletion in the maltase-glucoamylase gene *MGAM*, involved in digestion of starch, that is only present in the Karitiana population of South America. Using published archaic genomes, we find potentially introgressed variants from Neanderthals and Denisovans reaching appreciable frequencies in some populations. We discover a large range of copy numbers for multiallelic variants, and find cases of genes, such as *HPR* and *ORM2* that have mutated to high copy number on specific haplotypes. We additionally sequenced 26 samples from 13 populations using 10x Genomics linked-read technology, generating diploid *de novo* assemblies from which we validate and refine variant breakpoints and identify megabases of unique sequences not found in the human reference genome.

Inside Africa: Uncovering patterns of human genetic diversity

SMBE-PO-285

Making sense of GWAS: Understanding the genetic basis of human hair shape using mouse models

C. Basu Mallick [1,*], J. Ridell [1], X. Zheng [1 2], D. Ozdemir [1], M. Szpak [3], Q. Ayub [4], K. Adhikari [5], H. Brown [6], A. Ruiz-Linares [7], C. Mou [8], P. Hohenstein [9], M. Cheeseman [1], Y. Xue [3], I. J Jackson [1 10], C. Tyler-Smith [3], D. J Headon [1]

[1]Division of Developmental Biology, The Roslin Institute, University of Edinburgh, Edinburgh, United Kingdom, [2]College of Animal Science and Technology, Wuhan, China, [3]Wellcome Sanger Institute , Cambridge, United Kingdom, [4]Genomics Facility, School of Science, Monash University , Bandar Sunway, Malaysia, [5]Department of Genetics, Evolution and Environment , Department of Genetics, Evolution and Environment, and UCL Genetics Institute, University College London, London, [6]Division of Genetics and Genomics, The Roslin Institute, University of Edinburgh, Edinburgh, [7]Department of Genetics, Evolution and Environment, UCL Genetics Institute, University College London, London, United Kingdom, [8]College of Animal Science and Technology, Huazhong Agricultural University, Wuhan, China, [9]Department of Human Genetics, Leiden University Medical Center, Leiden, Netherlands, [10]MRC Human Genetics Unit, Institute of Genetics and Molecular Medicine, University of Edinburgh, Edinburgh, United Kingdom

Abstract:

Genome-wide association studies (GWAS) have revolutionized our understanding of diseases and biological adaptations over the last decade.The focus now is to push the boundaries further towards prioritizing these variants, performing functional follow-up to understand the biological processes from “sequence” to “consequence”.

Human hair is a distinct feature of one’s identity and differs in its shape, color, texture and growth patterns across populations. To date, 17 loci related to hair shape have been reported. Recent studies have shown that some of these variants are hallmarks of positive selection and modulators of signalling pathways thereby attesting their importance of in-depth investigation.

In this study, we focussed on an amino acid substitution (Q30R) in *PRSS53* reported to be associated with straight hair in humans. This variant is also positively selected in East Asians with derived allele frequency of 76%, however the adaptive phenotype remains elusive. This gene is also upregulated in psoriatic lesions. Hence, to ascertain the phenotypes related to the gene and the specific variant, we generated *Prss53*knockout and Q30R substitution mice. We found that *Prss53* knockout mice have curly whiskers and wavy coat. Assessment of Q30R mice showed effect on hair phenotype but also interestingly a skin phenotype. Our cell-based assay showed no interaction with EDAR pathway, thereby suggesting that both *EDAR* and *PRSS53* variants contributing to straight hair might be operating through independent pathways. Our study demonstrates the pleiotropic effect of the variant and the current findings will further assist in understanding the biomedical implications of the gene.

Inside Africa: Uncovering patterns of human genetic diversity
SMBE-PO-288

Haplogroup context is a key element to deciding if an mtDNA variant is a mutation or population polymorphism, complete study of mitochondrial tRNA's across multiple taxa

H. O'Keefe, R. Queen, P. Lord, J. Elson*

Abstract: Mitochondrial disorders are heterogeneous, showing variable presentation and penetrance. Over the last three decades, our ability to recognize mitochondrial patients and diagnose these mutations, linking genotype to phenotype, has greatly improved. However, it has become increasingly clear that these strides in diagnostics have not benefited all population groups. Recent studies have demonstrated that patients from genetically under-studied populations, in particular those of black African heritage, are less likely to receive a diagnosis of mtDNA disease. It has been suggested that haplogroup context might influence the presentation and penetrance of mtDNA disease; thus the spectrum of mutations that are associated with disease in different populations. However, to date there is only one well established example of such an effect: the increased penetrance of two Leber's hereditary optic neuropathy mutations on a haplogroup J background. This paper conducted the most extensive investigation to date into the importance of haplogroup context on the pathogenicity of mtDNA mutations. We searched for proven human point mutations across 726 multiple sequence alignments derived from 33 non-human species absent of disease. 58 pathogenic point mutations arise in the sequences of these species. We assessed the sequence context and found evidence of population variants that could modulate the phenotypic expression of these point mutations masking the pathogenic effects seen in humans. This supports the theory that sequence context is influential in the presentation of mtDNA disease and has implications for diagnostic practices. We have shown that our current understanding of the pathogenicity of mtDNA point mutations, primarily built on studies of individuals with haplogroups HVUKTJ, will not present a complete picture. This will have the effect of creating a diagnostic inequality, whereby individuals who do not belong to these lineages are less likely to receive a genetic diagnosis.

Inside Africa: Uncovering patterns of human genetic diversity
SMBE-PO-291
The impact of population variation in the analysis of microRNA target sites
M. Helmy [1,*], A. Hatlen, A. Marco
[1]School of Biological Sciences, University of Essex, Colchester, United Kingdom

Abstract: The impact of population variation in the analysis of microRNA target sites

Mohab Helmy, Andrea Hatlen and Antonio Marco

The impact of population variation in the analysis of regulatory interactions is an underdeveloped area. MicroRNA target recognition occurs via pairwise complementarity. Consequently, a number of computational prediction tools have been developed to identify potential target sites, that can be further validated experimentally. However, as microRNA target predictions are done mostly considering a reference genome sequence, target sites showing variation among populations are neglected. Here we study variation at microRNA target sites in human populations and quantify their impact in microRNA target prediction. We found that African populations carry a significant number of potential microRNA target sites that are not detectable in the current human reference genome sequence. Some of these targets are conserved in primates and only lost in Out-of-Africa populations. Indeed, we identified experimentally validated microRNA/transcript interactions that are not detected in standard microRNA target prediction programs, yet they have segregating target alleles abundant in non-European populations. In conclusion, here we show that ignoring population diversity may left out regulatory elements essential to understand disease and gene expression, particularly neglecting populations of African origin.

Inside Africa: Uncovering patterns of human genetic diversity
SMBE-PO-290
Disentangling ancient Eurasian admixture in East Africa
R. Hammarén [1,*], C. Schlebusch [1]
[1]Department of Organismal Biology, Uppsala University, Uppsala, Sweden

Abstract: Archeological findings point to long-standing, but sporadic interactions between the Middle East and Eastern Africa. The shared linguistic connection, the spread of the Semitic languages through Eastern Africa also support this history of interconectionality.
Genetic studies have instead suggested that a single major Eurasian back-migration took place in this region around 3000 years BP. Many East African populations today show large proportions of Eurasian admixture. A subset of these East African populations with Eurasian admixture migrated southwards. They were pastoralists and reached modern-day South Africa around 2000 years BP, and likely gave rise to the current day Khoekhoe people. The aim of this study is to investigate Eurasian admixture into Eastern Africa further by using modern population genetic approaches to discern the timing and size of admixture events. Furthermore, we will try to identify the Eurasian source population/s by extracting Eurasian genomic segments in East African populations and projecting this variation on a range of published genetic datasets from Africa and Eurasia.

Inside Africa: Uncovering patterns of human genetic diversity
SMBE-PO-289
The genetic history of Africa based on modern and ancient DNA
C. M. Schlebusch*

Abstract: In the last few decades, genetics played an increasingly important role in exploring human history. Genetic studies provided conclusive information that helped to answer challenging questions, such as the "Out of Africa" migration of modern humans. Moreover, genetics helped to establish Africa as the birthplace of anatomically modern humans. The history of human populations in Africa is complex and includes various demographic events that influenced patterns of genetic variation across the continent. Several studies based on mitochondrial DNA, Y-chromosomes, autosomal markers and whole genomes contributed to unraveling the genetic sub-structure of African populations. Through these studies, it became evident that deep African history is captured by connections among African hunter-gatherers, and that the deepest population divergence date to around 300,000 years before present. Furthermore, it was shown that agriculture had a large influence on the distribution of current-day Africans and that West African agriculturist populations populated the whole of sub-Saharan Africa, replacing and/or assimilating former groups. Other farming groups from Northeast Africa, admixed with Middle Eastern populations and also expanded southwards. These later population movements disrupted pre-existing population distributions and complicate inferences regarding deep human history. With the increased availability of full genomic data from diverse African populations we have more power to infer human demography. Furthermore, the first successful African ancient DNA genomes allow for direct temporal comparisons. With the promise of many more African modern and ancient genomes to come, the next few years will be exciting for investigating our species deep genetic history, rooted in Africa.

Insights from fitness landscapes into evolutionary pathways

SMBE-PO-295

Combinatorial genetic analysis of a regulatory network reveals the importance of higher order epistasis for gene deletion phenotypes

A. M. New [1,*], B. Lehner [2 3 4]

[1]Systems Biology Program, Centre for Genomic Regulation (CRG), The Barcelona Institute for Science and Technology, Barcelona, [2]Systems Biology Program, Centre for Genomic Regulation (CRG), The Barcelona Institute for Science and Technology, [3]Universitat Pompeu Fabra (UPF), [4]Institució Catalana de Recerca i Estudis Avançats (ICREA), Barcelona, Spain

Abstract: A key challenge in biology and evolution is to understand how mutations combine to alter phenotypes. Each genetic variant in a genome can have diverse effects, for example decreasing, increasing, inactivating, or changing the function of a protein or RNA. In contrast, systematic analyses of how mutations interact have typically used a single variant of each gene, most often a null allele. We therefore lack an understanding of how the full range of genetic variants that occur in individuals can interact. To address this shortcoming, we developed an approach to combine >5000 pairs of diverse mutations in a model regulatory network. The outcome of most mutation combinations could be accurately predicted by simple rules that capture the 'stereotypical' genetic interactions (epistasis) in the network. However, for individual genotypes, additional, unexpected pairwise and higher order genetic interactions can be important. These include 'harmonious' combinations of individually detrimental alleles that reconstitute alternative functional switches. Our results provide an overview of how the full spectra of possible mutations in genes interact and how these interactions can be predicted. Moreover, they illustrate how single, pairwise and higher order combinations of mutations can rapidly alter gene essentiality and rewire regulatory networks.

Insights from fitness landscapes into evolutionary pathways

SMBE-PO-305

Recurrent Sequence Evolution After Recurrent Gene Duplication

S. Von Der Dunk [1,*], B. Snel [1]

[1]Theoretical Biology & Bioinformatics, Utrecht University, Utrecht, Netherlands

Abstract: Examples abound of parallel evolution at the molecular (sequence) level as well as the functional level. As a consequence of the non-linear mapping between genotype and phenotype, recurrent evolution at one level is not always accompanied by recurrence at the other level (Natarajan, 2016).

So far, large-scale studies have focused on recurrent mutations between orthologous genes (Bazykin et al., 2007; Rokas & Carroll, 2008), i.e. comparing four orthologs comprising two separate evolutionary paths. These studies have revealed that high rates of parallel mutations result from purifying selection which limits the possible mutations that are permitted.

The evolutionary dynamics of orthologous genes are generally static. The fitness landscape rarely changes after a speciation event and as a result selection tends to maintain the ancestral gene function. This is not the case for paralogous genes. Duplicated genes are expected to accumulate functional changes due to relaxed negative selection or, in some cases, even positive selection acting on them, reflecting changes in the fitness landscape.

We extend on previous large-scale studies of recurrent sequence evolution by turning to paralogous genes for the first time. By linking concerted recurrent mutations that occurred after multiple independent duplication events, we identify those sequence changes that are potentially linked to actual functional changes and we reveal common evolutionary paths within gene families.

We detect extensive recurrence in various protein families, suggesting that certain aspects of evolution are predictable. Recurrent changes tend to be asymmetric affecting one paralog over the other and sometimes reflect independent loss or gain of subcellular locations. More broadly, this study contributes to our understanding of the link between genotype and phenotype on a dynamic evolutionary path.

Bazykin, Georgii A., et al. "Extensive parallelism in protein evolution." *Biology direct* 2.1 (2007): 20.

Natarajan, Chandrasekhar, et al. "Predictable convergence in hemoglobin function has unpredictable molecular underpinnings." *Science* 354.6310 (2016): 336-339.

Rokas, Antonis, and Sean B. Carroll. "Frequent and widespread parallel evolution of protein sequences." *Molecular biology and evolution* 25.9 (2008): 1943-1953.

Insights from fitness landscapes into evolutionary pathways
SMBE-PO-299
Understanding the genome-wide DFE of new mutations in Escherichia coli
M. Sane[*], D. Agashe[1]
[1]National Centre for Biological Sciences, Tata Institute of Fundamental Research, Bangalore, India

Abstract: Understanding the distribution of fitness effects (DFE) of new mutations is crucial as it can determine the dynamics of adaptation. Most experimental studies of the DFE have focused on highly beneficial mutations in single genes, ignoring a large fraction of mutational effects which can form standing genetic variation and contribute to adaptation. To obtain a more complete understanding of the DFE, we studied the genome-wide DFE of hundreds of single mutations in *Escherichia coli* populations evolved under relaxed selection in a mutation accumulation (MA) experiment, which allows an unbiased sampling of the entire underlying DFE. We find that the DFE changes substantially when DNA repair genes are lost from the genome. Gain or loss of DNA repair genes alters mutation spectra. Previous work in our lab showed that DNA repair genes are frequently gained/lost across the bacterial phylogeny, suggesting that gain/loss of DNA repair genes could significantly impact the evolutionary dynamics by altering the mutational fitness effects available to be sampled. Secondly, we find that environment and strain-by-environment interaction also impacts the DFE. Finally, we find that ancestral fitness is negatively correlated with both average beneficial fitness effect of new mutations and number of beneficial mutations. Together, our work furthers understanding of the factors that influence the genome-wide DFE of new mutations virtually unseen by selection. Our work is the first to test the impact of mutation spectrum on the DFE. Overall, our work has important implications for understanding the impact of new mutations on evolutionary dynamics of bacterial populations.

Insights from fitness landscapes into evolutionary pathways

SMBE-PO-294

SELVa: Simulator of Evolution with Landscape Variation

E. Nabieva [1,*], G. A. Bazykin [1]

[1]Center of Life Sciences, Skolkovo Institute of Science and Technology, Moscow, Russian Federation

Abstract: Organisms evolve to increase their fitness, a process that may be described as climbing the fitness landscape. Yet the landscape itself is changing over time: from the point of view of an organism, the change may occur in the environment, altering the fitness landscape; from the point of view of a single position in the genome, the landscape may additionally change due to epistatic interactions with other sites that are also evolving. To study the effect of fitness landscape changes, we present SELVa, the Simulator of Evolution with Landscape Variation. SELVa simulates sequence substitutions along a user-provided phylogenetic tree while varying the fitness landscapes that govern these substitutions. While some existing simulators of molecular evolution do accommodate fitness landscape changes at specific branches, SELVa is unique in its focus on and the toolkit for modeling landscape variation. SELVa allows the user to specify the regime of the landscape change (stochastic or deterministically set by the user according to a selection of rules), the choice of the initial and subsequent fitness landscapes (user-specified or sampled from one of several supported distributions), and whether or not the landscape is shared among parallel branches of the evolutionary tree. Written in Java and distributed as a platform-independent executable jar file, SELVa provides a flexible framework for studying the consequences of fitness landscape variations under various scenarios. SELVa is open-source and freely available at https://github.com/bazykinlab/SELVa.

Insights from fitness landscapes into evolutionary pathways

SMBE-PO-296

Human Enhancers With Complex Evolutionary Architectures Are More Active Than Enhancers With Simple Architectures

S. L. Fong [1,*], J. A. Capra [2]

[1]Human Genetics, [2]Biological Sciences, Vanderbilt University, Nashville, United States

Abstract: Enhancers are distal genomic regions that bind transcription factors (TFs) and regulate gene expression. Mutations and structural variation in enhancers give rise to new gene regulatory functions, but also associate with disease. While evolutionary conservation in protein coding genes is a powerful predictor of function, low conservation across enhancer landscapes makes enhancer variant interpretation a challenge. For example, only 1% of active human liver enhancers are conserved among mammals, indicating activity turns over quickly though these sequences are often shared across species.

Given the high frequency of enhancer mutations in complex disease and human biology, methods for interpreting these genetic variants are needed. We hypothesize the age of an enhancer sequence may hold information about its function. Enhancers are often composites of old “core” sequences and younger “derived” sequences, but how age architecture relates to function is not understood.We propose young, derived regions of enhancers increase enhancer function by introducing transcription factor binding sites (TFBSs) that potentiate activity.

We explored the relationship between enhancer evolutionary architecture and function by integrating evolutionary analysis of FANTOM enhancers across diverse cells and tissues, TFBS from ChIP-seq, and massively parallel reporter assays (MPRAs). Across tissues, enhancers are enriched in older genomic regions. We find over half of human enhancers arose in the ancestor of placental mammals, validating previous observations in liver tissue. We find 63% of enhancer sequences have “simple” architectures consisting of one age, while the rest are “complex,” consisting of sequences of multiple ages. These evolutionarily “simple” enhancers have lower activity than “complex” enhancers. Within complex enhancers, derived regions have higher MPRA activity than older core regions. This supports a model in which the presence of derived regions promotes strong enhancer activity. We hypothesized TFBS density might explain these functional differences. However, derived regions have lower TFBS density than simple enhancers and complex cores, suggesting that specific TFs in derived segments promote activity. We identified 90 TFs enriched for binding in simple enhancers, complex enhancer cores, or derived regions.

We propose human enhancers have two types of age architectures, those composed of composites of multiple sequence ages, and those in sequences with a single age. These two enhancer architectures differ in activity, TFBS density, and the TFs they bind. Complex and simple architectures likely harbor uneven distributions of disease variants. Thus, considering enhancer age architecture may aid the interpretation of human regulatory regions, functional targets, determinants of selection, and disease severity associated with enhancer variants.

Insights from fitness landscapes into evolutionary pathways
SMBE-PO-300
Mutation bias in empirical genotype-phenotype landscapes
A. V. Cano*, J. L. Payne [1]
[1]ETH Zurich, Zurich, Switzerland

Abstract: Mutation is a biased stochastic process, with some kinds of mutations occurring more frequently than others. Previous work has used synthetic genotype-phenotype landscapes to study how such mutation bias affects adaptive evolution. Here, we consider 746 empirical genotype-phenotype landscapes, each of which describes the strength with which a transcription factor binds its target DNA sequences, to study the influence of mutation bias on the adaptive evolution of increased binding strength. By using empirical genotype-phenotype landscapes, no assumptions are needed about landscape topography or about the DNA sequences that each landscape contains. The latter is particularly important, because the set of sequences that each landscape contains determines the kinds of mutations that can occur along a mutational path to an adaptive peak. That is, each landscape has an intrinsic mutation bias. Our results suggest that the interplay of this intrinsic bias with the bias in the mutation process influences landscape navigability, population diversity, and the predictability of evolution.

Insights from fitness landscapes into evolutionary pathways
SMBE-PO-303
Predicting hybrid fitness with Fisher's geometric model
B. De Sanctis [1,*], J. Welch [2]
[1]Department of Applied Mathematics and Theoretical Physics, [2]Department of Genetics, University of Cambridge, Cambridge, United Kingdom

Abstract: Hybridisation brings together novel combinations of alleles. The results can be fitness benefits (heterosis), severe fitness decline (hybrid breakdown), or some combination of the two. Fitness landscape models have been used to make predictions about these variable outcomes, but the models are often parameter rich, or restricted to particular classes of hybridisation (e.g., between closely-related inbred lines). A simpler and general approach uses Fisher's geometrical model, which maps genotypes to fitness using a model of optimizing selection on multyiple quantitative traits (Barton 2001; Martin 2014; Fraisse et al. 2016). Recent work (Simon et al. 2018) has made progress by treating the effects of allelic combinations as a Brownian Bridge in *n*-dimensional trait space. However, this approximation has not been derived in a rigorous way. Here, we derive the complete pdf of hybrid fitness as a function of the effects of fixed mutations, differentiating the parental lines, under Fisher's geometrical model. This will (i) address the robustness of the Brownian Bridge approximation (Simon et al. 2018), including the case of variable dominance and (ii) explore the effects of the fixation process on the distribution of effects.

References

Barton N. H. (2001). The role of hybridization in evolution. Mol Ecol 10(3):551-68.

Fraïsse C, Gunnarsson PA, Roze D, Bierne N, Welch JJ (2016). The genetics of speciation: insights from Fisher's geometric model. Evolution 70: 1450-1464.

Martin G. (2014). Fisher's geometrical model emerges as a property of complex integrated phenotypic networks. Genetics 197:237-255.

Simon A, Bierne N, Welch JJ (2018). Coadapted genomes and selection on hybrids: Fisher's geometric model explains a variety of empirical patterns. Evolution Letters 2:472-498.

Insights from fitness landscapes into evolutionary pathways

SMBE-PO-302

Distribution of amino acids on a phylogenetic tree as a mirror of single-position fitness landscape

G. Klink [1,*], G. Bazykin [1 2]

[1]molecular evolution, Institute for Information Transmission Problems RAS, [2]Skolkovo Institute of Science and Technology, Moscow, Russian Federation

Abstract: Estimation of single-position fitness landscapes and their variability using sequence and phylogenetic data is a very promising yet challenging task. Recently we developed a method that detects amino acids with variable fitness between two species by analysing substitution patterns across a phylogenetic tree. Using this method, we searched for amino acids with variable fitness in gp160 protein of HIV-1, and our results were in agreement with mutational scanning experiments. However, using simulations of evolution, we found out that our method works well only for intermediate changes in fitness. Thus we developed a new phylogenetic method that estimates, for each protein position, a probability that its fitness landscape is variable across a phylogenetic tree, using this tree and information about amino acids in different species in considered site. As it does not need numerous substitutions to occur, it can work for wider range of fitness shift. Moreover, it does not depend on ancestral reconstruction accuracy. For each amino acid that occurred in a site elsewhere on a phylogeny, the method finds three features, namely density, prevalence and incidence. After that, using neuronal network that was trained on simulated evolution of protein sites with different fitness landscapes it finds the most likely scenario to receive a joint position of amino acids that are observed in a site in the space of three features. Now we study opportunities of our new method using a phylogenetic tree of gp160 protein from A and B subtypes of HIV-1.

Insights from fitness landscapes into evolutionary pathways

SMBE-PO-301

Exaptive Subfunctionalization: Can Selective Pressure for the Primary Activity Lift the Secondary Activity to a Selectable Level?

B. van Loo [1,*], M. Heberlein [1], P. Mair [2], T. Kaminski [2], B. Eenink [1], C. Dilkaute [3], M. Schüürmann [3], A. Zinchenko [2], J. Jose [3], F. Hollfelder [2], E. Bornberg-Bauer [1]

[1]Institute for Evolution and Biodiversity, University of Münster, Münster, Germany, [2]Department of Biochemistry, University of Cambridge, Cambridge, United Kingdom, [3]Department of Pharmaceutical and Medicinal Chemistry, University of Münster, Münster, Germany

Abstract: Adaptation toward new functions often occurs via duplication of an existing protein-coding gene and subsequent accumulation of adaptive mutations in one of the gene copies. *De novo* emergence of a completely novel function in a protein is highly unlikely. Therefore the 'new' function is expected to be present as a secondary activity, albeit typically at low levels, in the 'old' protein prior to gene duplication. The latter means 'old' and 'new' activity must be able to coexist in the same protein.

Recent specificity profiles recorded for several existing sulfatases show that the ratio between the primary sulfatase activity and secondary phosphoesterase activity is constant across increasing levels of primary sulfatase activity [1]. The latter suggests that adaptation toward increased sulfatase activity could be accompanied by an equal increase in phosphoesterase activity. In this exaptive subfunctionalization (ESF) model an increased requirement for the primary activity required actually provides the driving force for improvement of the 'new' activity.

We are currently testing the validity of the ESF model by performing laboratory evolution of a moderately proficient sulfatase toward higher sulfatase activity. We have successfully established a lysis-free assay system to screen for increased sulfatase activity using micro-droplet based high-throughput screening technology [2], resulting in mutants with up to 28-fold increased sulfatase activity. Further rounds of adaptive evolution and activity testing for the secondary phosphoesterase activity are currently underway.

[1] van Loo *et al* (2019) *J. Am. Chem. Soc.* **141**, 370-84

[2] van Loo *et al* (2019) in prep, *bioRxiv* **479162**, doi: https://doi.org/10.1101/479162.

MEGE: Microbial Eukaryotic Genomic Evolution

SMBE-PO-313

Taxonomically-rich phylogenomic approaches to study the eukaryotic tree of life

M. A. Cerón-Romero [1 2,*], X. Maurer-Alcalá [3], J.-D. Grattepanche [2], Y. Yan [2], M. Fonseca [4], L. Katz [2 5]

[1]Biology, University Of Massachusetts, Amherst, [2]Biology, Smith College, Northampton, United States, [3]Institute of Cell Biology, University of Bern, Bern, Switzerland, [4]CIIMAR, University of Porto, Porto, Portugal, [5]Organismic and Evolutionary Biology, University Of Massachusetts, Amherst, United States

Abstract: Though substantial improvements in phylogenomic approaches and DNA/RNA sequencing over the last decades have enabled resolution of ancient relationships in the eukaryotic tree of life (EToL), the root remains elusive. The difficulty of placing the root lies in the fact that eukaryotes are ~1.7–2.1 billion years old and their genomes are dynamic (i.e. subject to both LGT and EGT, and evolving at highly variable rates of evolution). Genome rearrangements such as gene duplications and losses can blur the phylogenetic signal of the vertical inheritance, understanding the tempo of these rearrangements is therefore critical for understanding the earliest events in the history of eukaryotes. Likewise, accurate phylogenetic inference is important for understanding the tempo of gene duplication and losses across EToL. To address these issues, we developed two tools: PhyloToL, a flexible and modular phylogenomic pipeline that allows both taxon-rich phylogenetic inference across the tree of life and analysis of protein family evolution; and PhyloChromoMap, a tool for mapping the phylogeny of every gene along chromosomes. We use these tools to tackle two problems 1) estimating the root of EToL and 2) describing the tempo of gene duplications and losses across EToL. Our preliminary analyses reveal a root that contradicts the current 'popular' views of either a unikont-bikont or Excavata root. Also, our preliminary analyses data show evidence for a high rate of ancient gene duplication across EToL followed by many parallel losses. Together, our phylogenomic tools allow researchers to ask diverse questions about lineages that diverged ≥100 million years ago.

MEGE: Microbial Eukaryotic Genomic Evolution

SMBE-PO-309

Evolutionary trends in epitopes and low complexity regions of multiple Plasmodium species

S. N. Medley [1,*], A. Beaudet [1], H. Piontkivska [2], F. U. Battistuzzi [1 3]

[1]Dept. of Biological Sciences, Oakland University, Rochester, [2]Dept. of Biology, Kent State University, Kent, [3]Center for Data Science and Big Data Analytics, Oakland University, Rochester, United States

Abstract: The genomes of unicellular eukaryotes in the Plasmodium genus are of high evolutionary interest because of their pathogenic lifestyle. Some features of these genomes, such as epitopes, are relatively well known because of their medical applications, but not necessarily well understood from an evolutionary perspective. Similarly, low complexity regions (LCRs) are known to occur in high frequency in Plasmodia, but their evolutionary history and function are poorly understood. Interestingly, LCRs and epitopes share some characteristics in sequence structure, suggesting that the evolutionary mechanisms and functions of these regions could be similar. To investigate this possibility, we have conducted a genome level analysis of LCRs and epitopes across multiple species of Plasmodium. Our primary goal was to explore the variability of these regions over time in an effort to identify evolutionary mechanisms unique to and shared between LCRs and epitopes. We found different trends of sequence conservation with approximately 20% of epitopes being more conserved than their background genes while less than 2% of LCRs are conserved. Although the conservation of LCRs is lower than epitopes, we found a core number of LCRs with similar conservation to epitopes, indicating a possible functional role for these regions. Even based on structure, LCRs and epitopes differ with LCRs showing a preference for disordered sites (>80%) and epitopes having a higher proportion of alpha helix sites. Overall, these results show a dynamic evolutionary nature of these regions within Plasmodium species that is likely to have implications for their functionality over time.

MEGE: Microbial Eukaryotic Genomic Evolution

SMBE-PO-311

Massive, unprecedented intein content in two Anaeramoeba genomes reveals new aspects of intein mobility in eukaryotes.

L. Gallot-Lavallée [1,*], J. Jerlström-Hultqvist [1], C. W. Stairs [2], I. Cepicka [3], A. J. Roger [1], J. M. Archibald [1]

[1]Department of Biochemistry & Molecular Biology, Centre for Comparative Genomics & Evolutionary Bioinformatics, Dalhousie university, Halifax, Canada, [2]Institute for Cell and Molecular Biology, Uppsala, Sweden, [3]Department of Zoology, Cepicka lab, Prague, Czech Republic

Abstract: Inteins are self-splicing, selfish mobile protein elements with an enigmatic origin and evolution. Inteins are found in bacteria, archaea, eukaryotes and even viruses. However, how these selfish elements spread and the factors contributing to their persistence is poorly understood, in particular in eukaryotes where they are scarce. Here we show that the genomes of the anaerobic protists *Anaeramoeba ignava* and *A. flamelloides* have 113 and 51 inteins, respectively, in stark contrast to 4 found in the most intein-rich eukaryotic genome described previously. The *Anaeramoeba* inteins belong to 2 classes and reside in a wide range of proteins, some also invaded in eukaryotes, in diverse prokaryotes or viruses. Other *Anaeramoeba* inteins are in entirely new genomic locations. Using sequence similarity-based networks and phylogenomic methods, we show that some of the *A. ignava and A. flamelloides* inteins can be traced back to their common ancestor, while others appear to have likely been acquired from viruses. Some of the *Anaeramoeba* inteins have moved intragenomically, either between ancient paralogs, or into unrelated proteins with common motifs. Virus-derived inteins are found in diverse proteins, supporting the idea that large dsDNA viruses of eukaryotes have contributed to the spread of inteins with relaxed target site specificities. Taken together, our large and novel intein dataset extends the spectrum of eukaryotic intein-containing proteins and provides insight into eukaryotic intein dynamics and evolution.

MEGE: Microbial Eukaryotic Genomic Evolution

SMBE-PO-315

Genomic variation across strains of the toxic bloom-forming haptophyte Prymnesium parvum reveals possible toxicity determinants

A. Pendleton [1,*], R. Auber [1], G. Southard [2], W. Driscoll [3], J. Wisecaver [1]

[1]Purdue University, West Lafayette, [2]Texas Parks and Wildlife, Austin, [3]Pennsylvania State University Harrisburg, Harrisburg, United States

Abstract: The golden alga, *Prymnesium parvum*, is a mixotrophic microbial eukaryote capable of being both a photosynthesizing autotroph while also acting as a voracious aquatic predator that utilizes toxins, called prymnesins, to kill prey. During winter months, massive increases in population densities (*i.e.* blooms) of *P. parvum* can lead to dangerous levels of toxins in freshwater systems and large-scale fish kills. Though many details of the toxin's metabolic pathway remain unclear in this species, lab assays can help distinguish intercellular toxicity from predatory ability within a strain. Two strains of particular interest, 12A1 and 12B1, were isolated from the same liter of water in a Texas bloom, yet they display large phenotypic variation; 12B1 consistently displays low toxicity in assays, whereas its sympatric strain, 12A1, is highly toxic. Mirroring this phenotypic difference, large genetic variation has been revealed through sequencing. Initial ploidy estimates with short read data predicts that all high-toxicity strains, including 12A1, are diploid whereas 12B1 appears haploid. Further, *de novo* genome assemblies with Oxford Nanopore long reads for all lab-assayed strains, has yielded genomes that are twice as large in toxic strains compared to the non-toxic 12B1, substantiating possible ploidy differences. Assessments of single copy orthologs with BUSCO indicate that 12B1 has a gene duplication rate of 1.7%, while 12A1 presents a 37% rate. Varying patterns such as these are suggestive of possible whole-genome duplications or hybridization events within the lineage. We endeavor to sequence dozens more strains that are globally and temporally varied in order to better survey the genomic variation, estimate levels of population heterozygosity, help detail life history stages, and reveal genetic determinants of toxicity within this captivating haptophyte species.

MEGE: Microbial Eukaryotic Genomic Evolution

SMBE-PO-325

Evidence of extensive intraspecific noncoding reshuffling in a 169-kb mitochondrial genome of a basidiomycetous fungus

H.-H. Lee [1], H.-M. Ke [1], C.-Y. I. Lin [1], T. J. Lee [1], C.-L. Chung [2], I. J. Tsai [1,*]

[1]Biodiversity Research Center, Academia Sinica, [2]Department of Plant Pathology and Microbiology, National Taiwan University, Taipei City, Taiwan

Abstract: Comparative genomics of fungal mitochondrial genomes (mitogenomes) have revealed a remarkable pattern of rearrangement between and within major phyla owing to horizontal gene transfer (HGT) and recombination. The role of recombination was exemplified at a finer evolutionary time scale in basidiomycetes group of fungi as they display a diversity of mitochondrial DNA (mtDNA) inheritance patterns. Here, we assembled mitogenomes of six species from the Hymenochaetales order of basidiomycetes and examined 59 mitogenomes from two genetic lineages of *Pyrrhoderma noxium*. Gene order is largely colinear while intergene regions are major determinants of mitogenome size variation. Substantial sequence divergence was found in shared introns consistent with high HGT frequency observed in yeasts, but we also identified a rare case where an intron was retained in five species since speciation. In contrast to the hyperdiversity observed in nuclear genomes of *P. noxium,* mitogenomes' intraspecific polymorphisms at protein coding sequences are extremely low. Phylogeny based on introns revealed turnover as well as exchange of introns between two lineages. Strikingly, some strains harbor a mosaic origin of introns from both lineages. Analysis of intergenic sequence indicated substantial differences between and within lineages, and an expansion may be ongoing as a result of exchange between distal intergenes. These findings suggest that the evolution in mtDNAs is usually lineage specific but chimeric mitotypes are frequently observed, thus capturing the possible evolutionary processes shaping mitogenomes in a basidiomycete. The large mitogenome sizes reported in various basidiomycetes appear to be a result of interspecific reshuffling of intergenes.

MEGE: Microbial Eukaryotic Genomic Evolution

SMBE-PO-308

Analyses of genome structures in testate amoebae (Arcellinida) using single-cell 'omics

K. Munyenyembe [1,*], A. K. Weiner [1], L. A. Katz [1 2]

[1]Biological Sciences, Smith College, Northampton MA, [2]Program in Organismic and Evolutionary Biology, UMass – Amherst, Amherst MA, United States

Abstract: Testate (shell-building) amoebae of the order Arcellinida (Amoebozoa) are microbial eukaryotes that are highly abundant in freshwater ecosystems such as bogs, fens and lakes. Within the microbial community inhabiting these ecosystems, testate amoebae represent top predators. Because of their abundance, high sensitivity to abiotic environmental factors and the preservation of their tests in the fossil record, testate amoebae serve as excellent bioindicators to past and present climate change. Although Arcellinida are well studied from a morphological perspective, our knowledge on their genome evolution and genome structures remains very limited, which is mostly because they cannot be cultivated in the laboratory. To address these challenges, we have successfully adapted single-cell protocols for single-cell genomics and transcriptomics. We have by now characterized transcriptomes and genomes from multiple individuals (~50 transcriptomes and ~15 partial genomes) of Arcellinida. We are now working to assemble and analyze them using our custom-made pipeline PhyloToL. We are also using bioinformatic approaches to make inferences on patterns of molecular evolution in this lineage, and to map transcripts to the genome so that we can explore the structure and distribution of genes. In addition, we are calculating GC contents of coding and non-coding regions to gain insights on the distribution of genes in the genome and patterns of codon usage. Single-cell 'omics combined with bioinformatics are powerful tools for studying Arcellinida genome evolution and gaining insights on genome structures of this important group of protists.

MEGE: Microbial Eukaryotic Genomic Evolution

SMBE-PO-326

Mosaic aneuploidy and a complex life cycle shape the Leishmania genome

S. U. Franssen [1,*], C. Durrant [1], O. Stark [2], B. Moser [2], T. Downing [3], M. Sanders [1], M. Berriman [1], T. Wirth [4], G. Schönian [2], J. A. Cotton [1]

[1]Wellcome Sanger Institute, Hinxton, Cambridgeshire, United Kingdom, [2]Charité Universitätsmedizin, Berlin, Germany, [3]Dublin City University, Dublin, Ireland, [4]Muséum National d'Histoire Naturelle, Paris, France

Abstract: The unicellular protozoan parasites of the genus *Leishmania* cause the neglected tropical disease leishmaniasis, which is particularly prevalent in developing countries. *Leishmania* parasites are transmitted by sandflies and replicate inside macrophages of human or other mammalian hosts. Several unusual genetic features shape the *Leishmania* genome: First, despite they mainly reproduce clonally, hybrids within and between species exist. Second, *Leishmania* species show mosaic aneuploidy, where individual chromosomes experience copy number variation even within a clonal strain. Third, genes are expressed in polycistronic clusters and only post-transcriptionally regulated, suggesting copy number variations to have a particularly strong impact on gene regulation and adaptive evolution.

We investigate the genome-wide diversity of 151 globally distributed natural isolates from the *L. donovani* complex (comprising *L. donovani* and *L. infantum*) predominantly causing fatal visceral leishmaniasis. Phylogenetic reconstruction separates *L. donovani* samples into five diverse subgroups roughly coinciding with geographical origin. In contrast, little diversity is observed for globally distributed *L. infantum* samples. We find evidence for past hybridisation in some samples with high heterozygosity. However, the majority of the isolates are almost entirely homozygous with less than 7 SNPs/Mb. In accordance with a high turnover rate of aneuploidy, patterns of chromosomal copy numbers across samples are inconsistent with the genome-wide phylogeny. Using individual clades as replicates, we observed chromosome-specific copy number and somy variability. Chromosomes with elevated variability typically show reduced heterozygosity, which provides the first experimental evidence for the reduction of heterozygosity due to aneuploidy turnover. Moreover, sub-chromosome scale duplications and deletions are frequent but variable between chromosomes and affect the majority of all genes. Our results suggest profound impacts of the interplay between facultative sexuality, extensive sub-chromosome scale copy number variation and highly variable aneuploidy in *Leishmania* that make the population genetics of this human pathogen highly unusual.

MEGE: Microbial Eukaryotic Genomic Evolution
SMBE-PO-310
Analyses of fungal pangenomes
C. G. P. McCarthy [1,*], D. A. Fitzpatrick [1]
[1]Dept. of Biology, Maynooth University, Maynooth, Ireland

Abstract: Pangenomes evolve in prokaryotes as a result of rapid evolution and promiscuous horizontal transfer of genetic material. Although eukaryotes are generally under more restrictive evolutionary constraints than prokaryotes (e.g. lower levels of HGT), pangenomic structure has also been identified in plant, algal and fungal species. Using a previously-published methodology based on sequence homology and conserved microsynteny in addition to bespoke pipelines, we have investigated the pangenomes of four model fungal species: *Saccharomyces cerevisiae*, *Candida albicans*, *Cryptococcus neoformans* var. *grubii* and *Aspergillus fumigatus*. Between 80-90% of gene models per strain in each of these species are "core" genes that are highly-conserved across all strains of that species, many of which are involved in housekeeping and conserved survival processes. In many of these species the remaining “accessory” gene models are clustered within chromosomal extremities and may play a role in phenotypic diversity within their species. Various functional analyses of fungal pangenomes shows that core and accessory eukaryotic species genomes encompass a variety of phenotypes and suggest that gene duplication events play a larger role in eukaryote pangenome evolution than HGT. We are currently refining our analytic pipelines for general use and we are also researching the pangenomes of other model fungal organisms.

MEGE: Microbial Eukaryotic Genomic Evolution
SMBE-PO-307
Phylogenomic reconstruction of the phylum Ciliophora using single-cell 'omics

Y. Wang [1,*], Y. Yan [2], F. Gao [1], W. Song [1], L. A. Katz [2]
[1]Institute of Evolution & Marine Biodiversity, Ocean University of China, Qingdao, China, [2]Department of Biological Sciences, Smith College, Northampton, United States

Abstract: Ciliates are a large and diverse group of microbial eukaryotes that are defined by the presence of two distinct types of nuclei – germline and soma – in every cell. Ciliates play important roles in many biological and ecological studies, and cultivable genera such as *Tetrahymena* and *Paramecium* are models in many cell biology studies. However, the phylogenetic relationships within ciliates have been constructed only with single marker genes, such as SSU rDNA, LSU rDNA, mtSSU rDNA and/or a few protein coding genes (i.e.actin and α-tubulin). In this study, we isolated 130 ciliates representing 10 classes, and then generated and sequenced single-cell genomes and transcriptomes. We are using the resulting data to reconstruct the phylogeny of ciliates through our PhyloTOL pipeline, with the intent of gaining a better understanding of ciliate evolution based on 'omics data from uncultivable lineages. In addition, we are searching for evidence for the presence of gene-sized chromosomes (i.e. short contigs with two telomeres, which have been found in the ciliate classes, Spirotrichea, Armophorea and Phyllopharyngea) in additional lineages, and we are characterizing genetic code evolution in ciliates (i.e. reassign standard stop codons to amino acids) at a finer scale. Together, these analyses will provide a greater resolution to genome evolution in this ~0.8-1 billion year old clade.

MEGE: Microbial Eukaryotic Genomic Evolution

SMBE-PO-316

Widespread tRNA Deamination In The Last Common Ancestor Of The Opisthokonts

M. Carr [1,*]

[1]Biological & Geographical Sciences, University of Huddersfield, Huddersfield, United Kingdom

Abstract: Patterns of codon usage bias vary considerably across life, and metazoans in particular, with both mutation pressure and natural selection contributing to bias. The highly variable nature of their codon usage prevents the reconstruction of ancestral traits using extant metazoan species. Traits conserved across the unicellular lineages related to Metazoa, however, do point to ancestral characteristics in the earliest members of the metazoan lineage (premetazoans). Presented here are patterns of codon usage bias in two choanoflagellates, a filasterean and a nuclearioid, which are all unicellular opisthokonts, and highlight the high level of conservation across one billion years of evolution.

The conserved traits can be considered ancestral to metazoans and fungi. In particular, it is shown that codon usage bias is directed towards a set of translationally optimal GC-ending codons. Selection appears to operate through translational efficiency and translational accuracy, with even the most weakly biased genes showing the signature of selection. The major tRNA genes for twofold degenerate amino acids match optimal codons, however this is not the case for three-fold to six-fold amino acids. For these amino acids optimal codons show cytosine at the synonymous position, whereas tRNA anticodons have adenosine at the wobble site. It can be seen that tRNA molecules undergo deamination of adenosine at the wobble site to inosine, facilitating complementary binding to cytosine in optimal codons. Highly biased genes in the opisthokont protists, preferentially use deaminated tRNA molecules in preference to unmodified molecules in order to optimise protein synthesis.

MEGE: Microbial Eukaryotic Genomic Evolution

SMBE-PO-312

Heterothallism as the ancestral thallism state in Phyllosticta genus

D. A. L. Petters-Vandresen [1,*], M. A. Machado [2], C. Glienke [1]

[1]Genetics, Federal University of Paraná, Curitiba, [2]Centro de Citricultura Sylvio Moreira, Cordeirópolis, Brazil

Abstract: *Phyllosticta* genus comprises many plant pathogens, endophytes and saprophytes, and they differ in their biology and ecology by many aspects, such as the system of sexual reproduction, which plays an important role in the plant diseases caused by this genus. Some species, such as *P. capitalensis, P. abieticola and P. musarum,* are known to be homothallic due to the ascospores production in pure cultures. Other species, such as *P. citricarpa* and *P. citribraziliensis,* were suggested as heterothallic based in genomic analysis of the mating-type locus. However, for most of the *Phyllosticta* species, the mating strategies are unknown, as there is no information available on ascospores production, crossing studies or genomic data. In order to investigate the evolution of mating strategies in *Phyllosticta* genus, we performed an ancestral character reconstruction analysis in Mesquite using a multi-locus phylogenetic tree (ITS, *tef1*, *act*, *gapdh*) of *Phyllosticta* and Botryosphaeriaceae species to map the characters. Species were defined as heterothallic, homothallic or unknown based in genomic data and taxonomic characterizations available in the literature. Our results suggest that heterothallism is the ancestral state in *Phyllosticta* genus and the most-likely mating-strategy for the *Phyllosticta* species in which the thallism state is unknown. In addition, there were at least five shifts from heterothallism to homothallism in *Phyllosticta* genus. These results contribute to the understanding of the different mating strategies of *Phyllosticta* species under an evolutionary approach, in order to promote better comprehension of the mechanisms involved in sexual reproduction and its role in different plant diseases.

MEGE: Microbial Eukaryotic Genomic Evolution

SMBE-PO-319

Aphelid phylogenomics illuminates the early evolution of Fungi

D. Moreira [1,*], G. Torruella [1], X. Grau-Bové [2], S. Karpov [3], J. Burns [4], A. Sebé-Pedrós [5], E. Völcker [6], P. López-García [1]

[1]CNRS - Universite Paris-Sud, Orsay, France, [2]Liverpool School of Tropical Medicine, Liverpool, United Kingdom, [3]Russian Academy of Sciences, Saint Petersburg, Russian Federation, [4]American Museum of Natural History, New York, United States, [5]Centre for Genomic Regulation, Barcelona, Spain, [6]Penard Lab, Cape Town, South Africa

Abstract: Opisthokonts are one of the major eukaryotic super-groups and encompasses two large lineages, the Holomycota and the Holozoa, which contain well known multicellular organisms, respectively Fungi and Metazoa, together with a vast cohort of unicellular relatives. Phylogenetic and functional analyses have shown that many genes typically involved in metazoan multicellularity were already present in the unicellular relatives of animals. On the holomycotan branch, the situation is complex because the boundaries defining Fungi are fuzzy and, at the same time, environmental 18S rRNA metabarcoding combined with more classical taxonomic studies have revealed a wide diversity of previously unrecognized lineages branching deeply within classical Fungi (i.e., osmotrophic lineages, including chytrids) and prior to their divergence (aphelids, rozellids, microsporidia, and nucleariids). Aphelids are little-known phagotrophic parasites of algae whose life cycle and morphology resemble those of the parasitic rozellids (Cryptomycota, Rozellomycota). Although their affiliation to the Holomycota was clear, their precise phylogenetic position based on single-gene analyses (RNA polymerase and rRNA genes) remained unresolved. To settle this question, we generated full life-cycle transcriptome data for the aphelid species *Paraphelidium tribonemae*. In-depth multi-gene phylogenomic analyses using several protein datasets place this aphelid as the closest relative of fungi to the exclusion of rozellids and Microsporidia. In contrast with the reduced genome of the rozellid *Rozella allomycis*, we infer a rich proteome for our aphelid species, more similar to those of free-living species, with a metabolism similar to fungi, including cellulases likely involved in algal cell-wall penetration and enzymes involved in chitin biosynthesis. Our results sugpport that fungi evolved from complex aphelid-like ancestors that lost phagotrophy and became osmotrophic.

MEGE: Microbial Eukaryotic Genomic Evolution

SMBE-PO-321

Dual nuclear architecture and phenotypic assortment increase evolvability following sex

J. Tarkington [1,*], R. A. Zufall [1]

[1]Biology and Biochemistry, University of Houston, Houston, United States

Abstract: The variety of ways that biological information gets packaged and passed from generation to generation inevitably impacts the production and maintenance of genetic variation. Organisms with genetic systems that produce just the right amount of variation at just the right time are likely to outcompete those that do not. Ciliates possess two nuclei, a diploid germline nucleus and a transcriptionally active polyploid somatic nucleus that gets reset every sexual generation. The ciliate *Tetrahymena thermophila*possess an additional unusual genetic feature, called phenotypic (allelic) assortment, which increases the amount of genotypic variation during the vegetative growth following sex until all allelic variation is segregated and all cells are homozygous in the somatic nucleus. This process happens randomly at every locus producing an enormous amount of combinatorial variation from a single newly formed heterozygous sexual progeny. Fisher's fundamental theorem predicts that this increased genetic variation should in turn increase evolvability. To test this hypothesis, I compared the rate of adaptation in *T. thermophila*populations that underwent phenotypic assortment to those that did not. Populations that underwent phenotypic assortment adapted more quickly than those that did not suggesting that the additional genetic variation generated by phenotypic assortment increases evolvability following sex. Previous explanations for the success of this unusual genetic system have suggested it is a way to sequester and deal with transposable elements but our results suggest that the dual nuclear architecture of ciliates may be successful because it maximizes adaptation in the short-term without any of the long-term consequences when the environment changes.

MEGE: Microbial Eukaryotic Genomic Evolution

SMBE-PO-322

Phentoypic diversity through complex ancestries in yeasts

D. Peris [1 2 3,*], E. P. Baker [2 3], X. Li [4], M. C. Kuang [2], J. Kominek [2 3], Q. K. Langdon [5], M. Adams [6], J. A. Koshalek [6], A. B. Hulfachor [2], D. A. Opulente [2 3], Q.-M. Wang [7], F.-Y. Bai [7], C. Landry [8], J. P. Sampaio [9], P. Gonçalves [9], D. Libkind [10], D. Hall [11], K. E. Hyma [12], J. Fay [12], M. I. Sardi [2 3], A. P. Gasch [3 13], U. Bond [14], C. T. Hittinger [3 13]

[1]Food Biotechnology, IATA-CSIC, Paterna, Spain, [2]Laboratory of Genetics, [3]DOE Great Lakes Bioenergy Research Center, University of Wisconsin-Madison, Madison, [4]Division of Biology and Biomedical Sciences, Washington University, St. Louis, [5]Laboratory of Genetics, University of Wisconsin-Madison, Paterna, [6]Biotechnology center, University of Wisconsin-Madison, Madison, United States, [7]State Key Laboratory of Mycology, Institute of Microbiology, Chinese Academy of Sciences, Beijing, China, [8]Département de Biologie, Institut de Biologie Intégrative et des Systèmes (IBIS), Université Laval, Quebec, Canada, [9]Departamento de Ciências da Vida, Faculdade de Ciências e Tecnologia, Universidade Nova de Lisboa, Caparica, Portugal, [10]Laboratorio de Microbiología Aplicada, Biotecnología y Bioinformática, Instituto Andino Patagónico de Tecnologías Biológicas y Geoambientales, IPATEC (CONICET-UNComahue), Bariloche, Argentina, [11]Biochemistry, Lawrence University, Madison, [12]Department of Genetics, Center for Genome Sciences and Systems Biology, Washington University, St. Louis, [13]Department of Genetics, University of Wisconsin-Madison, Madison, United States, [14]School of Genetics and Microbiology, Trinity College Dublin, Dublin, Ireland

Abstract: The genus *Saccharomyces* has become one of the most important model genera to understand evolution. The scientific community has put a lot of effort into understanding the evolution and population structure of some of the S*accharomyces* species. Indeed, recently more than 1,000 *Saccharomyces cerevisiae*, 162 S*accharomyces paradoxus*, 46 S*accharomyces uvarum*, and 25 S*accharomyces eubayanus* strains have been sequenced. However, population structure remains fairly underexplored in S*accharomyces mikatae,* S*accharomyces kudriavzevii,* S*accharomyces arboricola*, and the recently discovered S*accharomyces jureii*. Here, we studied the mitochondrial inheritance of 2,000 *Saccharomyces* strains, we sequenced a representative of each available *Saccharomyces* lineage at high quality, and we included additional *Saccharomyces* sequences to study the population structure and gene flow within and between lineages. Together, with representatives of recently sequenced *Saccharomyces* species, we were also able to detect and quantify several reticulate events between species at the nuclear, mitochondrial, and the 2-micron plasmid genomes. Reticulate events had an impact in the phenotypes of domesticated strains for the production of alcoholic beverages, and they also have been important in the lost or maintenance of phenotypic traits differentiating introgressed strains from those of the same lineage.

MEGE: Microbial Eukaryotic Genomic Evolution

SMBE-PO-320

Stable exogenous gene expression in Acanthamoeba

A. B. Moreno [1,*], E. Wistrand-Yuen [1], D. Gonçalves-Carneiro [2], L. Guy [1]

[1]Uppsala University, Uppsala, Sweden, [2]The Rockefeller University, New York, United States

Abstract: *Acanthamoeba* spp. are free-living amoebae ubiquitous in a variety of soil, water and air environments.
Several species of *Acanthamoeba* have been identified as opportunistic human pathogens, causing severe diseases such as *Acanthamoeba* keratitis and granulomatous amoebic encephalitis. They also serve as hosts and reservoirs for many intracellular human pathogens, including *Legionella pneumophila, Coxiella burnetii, Listeria monocytogenes* and *Chlamydia*.
Due to the similar phagocytosis processes in amoebae and macrophages, it has been proposed that amoebae act as an evolutionary crib for intracellular bacteria to infect human cells.
Amoebae are thus commonly used as models for bacteria-host studies. *Dictyostelium discoideum* is an example of these model amoeba, however it is not the natural host of most accidental human pathogens, and there are some restrictions in using it, such as a limited temperature range. The genomic complexity of *Acanthamoeba* has proven a major difficulty for the study of these species and its use as a model organism.
Transfection of *Acanthamoeba* has been previously reported, albeit with low efficiency, but the foreign genes have all been replicated episomally. Here we describe a method for stable expression of genes in *Acanthamoeba,* where the exogenous genes are inserted in the chromosome, codon optimised, and are under *Acanthamoeba* promoters.
The ability to stably introduce or express genes in *Acanthamoeba* will allow for a wider range of genetic analyses, and possibly lead to a better understanding of these pathogens, their mechanisms of virulence, and the evolution of host-pathogen interactions.

MEGE: Microbial Eukaryotic Genomic Evolution

SMBE-PO-327

Rooting the eukaryotic radiation with new models of genome evolution.

C. Petitjean [1,*], D. Bass [2], G. J. Szollosi [3], T. A. Williams [1]

[1]Biology, University of Bristol, Bristol, [2]Life Sciences, Natural History Museum London, London, United Kingdom, [3]Evolutionary Genomics Research Group, MTA-ELTE Lendület, Budapest, Hungary

Abstract: The deep relationships among the main lineages of eukaryotes, and in particular the root of the eukaryotic tree, remain debated. Two main rooting hypotheses are actively discussed. The first, the Unikont/Bikont hypothesis (UB), places the root between the Unikonts, including Metazoa, Fungi, Amoebozoa and some related protist lineages, and the bikonts, i.e. all the other lineages, including Archaeplastida and a large diversity of unicellular eukaryotes. The second hypothesis, Neozoan/Excavate (NE), proposes the root to be between the excavates, a very diverse group of protists, and all the other eukaryotes. Each of these hypotheses has major implications for the nature of the last eukaryotic common ancestor (LECA), as its complexity level, gene content or the evolution of the main features of each lineage. In order to solve this fundamental evolutionary question, we are exploring the use of concatenation, multispecies coalescent, and recently-developed approaches to gene tree-species tree reconciliation that allow species trees to be rooted without an outgroup. Our analyses make use of a broadly-sampled dataset of 98 complete genomes and largely-complete transcriptomes of Eukaryotes, including new lineages. We present ongoing work on the topology and root of the eukaryotic tree and the metabolic capabilities of the last eukaryotic common ancestor.

MEGE: Microbial Eukaryotic Genomic Evolution

SMBE-PO-323

Understanding genome rearrangement in ciliates from an evolutionary perspective

Y. Feng [1,*], L. Beh [2], L. Landweber [3]

[1]Biological Sciences, Columbia University, New York, [2]Princeton University, Princeton, [3]Columbia University, New York, United States

Abstract: Ciliates, single-celled eukaryotes with separate somatic and germline genomes, experience genome rearrangement every time they initiate their sexual life cycle. Some lineages delete germline-limited DNA and ligate the flanking sequences in their original order, while others also descramble genes by inversion or translocation. An interesting question that arises is how this process of genome remodeling evolved. Previously, our lab studied genome rearrangement in the ciliate *Oxytricha trifallax*. Here, we focus on *Euplotes*, a distantly related spirotrichous lineage, to understand this process from an evolutionary perspective. We sequenced and assembled the somatic genome of *Euplotes woodruffi*, and then investigated this question from two directions.

We first developed a computational pipeline based on split-read mapping, which infers rearrangement features on somatic chromosomes in the absence of a germline genome assembly. We particularly focused on inferring the microhomologous repeat sequences present at rearrangement junctions. We identified novel scrambled genes from these data and validated several by PCR. This provides a glimpse into *Euplotes*' DNA rearrangements. We are also assembling the germline genome of *E. woodruffi* with the benefit of long reads from third-generation sequencing. With both genomes in hand, we can infer a complete rearrangement map. In addition, comparative genomic analysis of *Oxytricha* and *Euplotes* reveals genome-wide differences in rearrangement complexity, as well as rearrangement patterns for each ortholog, providing us with insight into the evolution of genome architecture and chromosome rearrangements in ciliates.

Microbial Evolution in Complex Environments

SMBE-PO-336

Deciphering a novel packaging system in Gram-negative PICIs

N. K. Alqurainy [1,*], A. F. Salom [1], L. Miguel [1], J. R. Penades [1]

[1]institute of infection immunity & inflammation , University of Glasgow, Glasgow, United Kingdom

Abstract: Phage inducible chromosomal islands (PICIs) are clinically relevant mobile genetic elements widespread in bacteria. PICIs are a subset of phage satellites implicated in virulence and host adaptation, hence, it is vital to understand how they are transferred. Most of the previously characterised PICIs can be packaged by two main strategies. Some PICIs encodes a homolog of the phage terminase small subunit ($TerS_S$) and packages their concatemeric DNA into phage virion particles using headful packaging mechanism. The $TerS_s$ recognise and bind to the PICI-specific (*pac)* sequence in the complex with phage terminase large subunit to initiates headful packaging. Other, non-encoding ($TerS_s$), PICIs carry the sequential helper phage cos site in the PICI genome, which will be then recognised by the helper phage terminase ($TerS_P$) complex leading to highly transfer frequencies. In this study, we have reported for the first time a new cohesive family of PICI in Gram-negative bacteria encoding for unprecedented packaging mechanism. As these PICI encode for capsid morphogenesis genes to form a mature small capsid, we hypothesis that, the PICI hijacking preformed phage tail proteins resulting in a high-frequency transfer. Using the PICI in Escherichia coli strain EDL933 as models, we show that the PICI only parasite on phage tail proteins to mobilise and transfer. Overall, this study highlighting new sophisticated strategy utilised by these PICIs for their efficient mobility in nature.

Microbial Evolution in Complex Environments

SMBE-PO-331

Deianiraea vastatrix, the first extracellular Rickettsiales, provides novel insight into the evolution of the order

D. Sassera [1,*]

[1]Department of Biology and Biotechology, Universita' degli Studi di Pavia, Pavia, Italy

Abstract: *Rickettsiales* (class *Alphaproteobacteria*) are a diverse bacterial order that encompasses important human and animal pathogens, manipulators of host reproduction, and mutualists. In this wide range of capabilities and behaviours, strict intracellularity has always been found to be the common feature.

The discovery of a novel *Rickettsiales* associated with *Paramecium*, *Deianiraea vastatrix*, challenges this paradigm. This bacterium displays a host-associated but always extracellular lifestyle, including the ability to replicate outside host cells. Genomic analyses reveal the presence of genes putatively involved in extracellular interaction with *Paramecium*, and amino acids biosynthesis capabilities that are higher than those of all other *Rickettsiales* combined. Phylogenetic and phylogenomic reconstructions show that *D. vastatrix* is the first member of a novel family clade within *Rickettsiales*, the *Deianiraeaceae*, which has an estimated diversity comparable to the other three *Rickettsiales* families, and does not appear to be phylogenetically basal to the other families of the order.

Considering these findings and re-evaluating the different means of interaction of *Rickettsiales* bacteria with eukaryotic cells, it is possible to formulate an hypothesis for the evolution in *Rickettsiales* that is alternative to the classical scenario implying an early evolution of intracellularity in the ancestor of the order. According to the novel 'intracellularity-late' scenario, the last *Rickettsiales* common ancestor would have been an extracellular, metabolically and structurally versatile bacterium, with strong capabilities to interact with other cells. Obligate intracellularity would have evolved later, exploiting these capabilities, in parallel and independently, in different sub-lineages.

The present finding could help to understand the origin of *Rickettsiales* lineages and to clarify the evolution of the current diverse interactions with their host cells. Furthermore, it could impact on the open debate on the lifestyle of the last common ancestor of mitochondria within *Alphaproteobacteria*.

Microbial Evolution in Complex Environments

SMBE-PO-333

Single Nucleotide Mapping of the Locally Accessible Trait Space in Yeast Reveals Pareto Fronts that Constrain Initial Adaptation

Y. Li [1], D. Petrov [1], G. Sherlock [2,*]

[1]Biology, [2]Genetics, Stanford University, Stanford, United States

Abstract: Tradeoffs constrain the improvement of performance of multiple traits simultaneously. Such tradeoffs define a Pareto optimality front that represents a set of optimal individuals that cannot be improved in any one trait without reducing performance in another. Surprisingly, experimental evolution often yields genotypes with improved performance in all measured traits, suggesting a lack of Pareto optimality, perhaps indicating an absence of tradeoffs at least in the short-term. Moreover, even when the improvement in one trait is found to be associated with the loss of performance in another, it is hard to establish that such a negative correlation is not due to the specific sampled mutations, with other unsampled adaptive mutations not showing such negative trait correlations. Here we use dense sampling of first step adaptive mutations in *S. cerevisiae* evolving in glucose-limited media to ask whether such adaptive mutations do result in tradeoffs. We evolved barcoded yeast populations under various conditions that select for improved performance in different parts of the yeast growth cycle. Similar to previous studies we find that the founding clone is not on a Pareto optimal front and its performance can indeed be improved simultaneously in all measured traits. However, we also find that early evolution includes adaptive mutations of sufficient magnitude to reach the Pareto optimality line, which thus defines the short-term evolutionary tradeoffs. Specifically, by isolating ~500 adaptive clones and quantifying their performances in each part of the growth cycle, we defined Pareto optimality fronts between fermentation and respiration, and between respiration and stationary phases. Our analysis suggests that no single mutation in the founding yeast genome can circumvent the detected tradeoffs. Further, we sequenced hundreds of adaptive clones, identified the molecular basis underlying the identified trade-offs and revealed novel targets of adaptation.

Microbial Evolution in Complex Environments

SMBE-PO-329

Understanding the Connection Between Phenotype and Genotype in Enterobacteriaceae

J. Alkhouri [1,*], F. Santiago [1], S. Sindi [1], M. Barlow [1]

[1]School of Natural Sciences - Quantitative and Systems Biology, University of California, Merced, Merced, United States

Abstract: It is the central dogma of modern biology that genotype determines phenotype. That relationship however, is not yet well understood. Antibiotic resistance is a good model system for understanding the relationship between genotype and phenotype. The correlation between the presence of resistance genes and the expression of resistance phenotype is much less than we expect. This leads us to hypothesize that antibiotic resistance phenotypes result from complex interactions between multiple resistance genes and regulatory mutations. We're studying resistance to ceftolozane/tazobactam because this is a new drug combination and there is great variation in resistance phenotypes in a collection of hospital isolates that we have been collecting since 2013 from Dignity Health, Mercy Medical Center in Merced, CA.

We assessed susceptibility of 109 *E. Coli* clinical isolates to ceftolozane/tazobactam with Kirby-Bauer disk diffusions. We sequenced (Illumina high throughput genomic sequences), and assembled their genomes, then identified all known resistance genes based on the Comprehensive Antibiotic Resistance Database.

The results showed that seven (5%) of the 109 isolates were resistant to ceftolozane/tazobactam based on CLSI break-points. The most common b-lactamases in these isolates were TEM-4 (71.43%), OXA-1 (57.14%), CTX-M (57.14%), and ampC (57.14%). TEM-4 seems likely to contribute to resistance to ceftolozane/tazobactam, and there are also many non b-lactamase candidate genes that may contribute to this phenotype.

While we do not yet know the gene combinations that result in ceftolozane/tazobactam resistance, we can say with certainty that a single gene is not responsible for this resistance phenotype, thus confirming our hypothesis.

Microbial Evolution in Complex Environments

SMBE-PO-337

Microbiota as a selective force driving Pseudomonas aeruginosa adaptation to the cystic fibrosis lung

M. J. Bottery [1,*], V. Friman [1], J. Pitchford [1]

[1]Biology, University of York, York, United Kingdom

Abstract: Cystic Fibrosis (CF) is an inherited recessive genetic disorder effecting 10,400 people in the UK. CF causes thick and sticky mucus to build up within the lung and leads to frequent and severe lung infections. The opportunistic pathogen *Pseudomonas aeruginosa* is the leading cause of morbidity and mortality in individuals with CF. Initially *P. aeruginosa* infections can be controlled and even eradicated by intensive antibiotic treatment, however *P. aeruginosa* infections inevitably become chronic and unclearable through antibiotic treatment. Due to advances in genomic sequencing we are beginning to understand the genetic changes that occur in *P. aeruginosa* during the transition from acute to chronic infection. Although some adaptive changes can be associated with strong selective pressures, such as antibiotic use, many of the selective forces which may affect adaptation, such as interactions with the surrounding microbiota, remain unexplored. My research explores how interactions between *P. aeruginosa* and commonly co-occurring pathogens *Staphylococcus aureus* and *Stenotrophomonas maltophilia* alters the fitness and evolutionary trajectories of *P. aeruginosa* during the adaptation to the complex environment of the CF lung. Through co-evolution of environmental *P. aeruginosa* with clinically isolated *S. aureus* and *S. maltophilia* in complex conditions that closely replicate the CF lung, we are able to track the real time evolution and diversification of *P. aeruginosa* in response to the CF microbiota.

Microbial Evolution in Complex Environments

SMBE-PO-328

Linking selection for GC content with DNA repair in prokaryotic genomes.

P. L. F. Johnson [1,*], J. L. Weissman [1], W. F. Fagan [1]

[1]Biology, University Of Maryland, College Park, United States

Abstract: Genomic GC content varies widely among microbes for reasons unknown. While mutation bias partially explains this variation, prokaryotes near-universally have a higher GC content than predicted solely by this bias. Debate surrounds the relative importance of the remaining explanations of selection versus biased gene conversion favoring GC alleles. Some environments (e.g. soils) are associated with a high genomic GC content of their inhabitants, which implies that this content may be a selective adaptation to particular habitats. Here, we report a novel association between the presence of the non-homologous end joining DNA double-strand break repair pathway and GC content; this observation suggests that high GC content may be an adaptation to facilitate repair of double strand breaks when homologous recombination is not possible. We discuss potential mechanisms accounting for the observed association, and provide preliminary evidence that sites experiencing higher rates of double-strand breaks are under selection for increased GC content relative to the genomic background.

Microbial Evolution in Complex Environments

SMBE-PO-335

Bayesian phylogenomic feature selection for predicting prokaryotic optimal growth temperature

E. R. R. Moody [1,*], J. Fearn [2], D. Pisani [1 3], C. Campbell [2 4], T. A. Williams [1]

[1]Biological Sciences, [2]Engineering Mathematics, [3]Earth Sciences, [4]Intelligent Systems Lab, University of Bristol, Bristol, United Kingdom

Abstract: Prokaryotic life is incredibly diverse and inhabits a wide range of different environments from arctic ice-sheets to underwater volcanoes. What is it about the genomes of these organisms that allow them to thrive in such a broad, and occasionally extreme, range of conditions? Here, we introduce a phylogenetic relevance vector machine (PhyRVM), which combines traditional phylogenetic regression with automatic relevance determination. We show that PhyRVM accurately predicts the optimal growth temperature of bacterial and archaeal species to within 4℃. Fitting the model to ~5000 prokaryotic genomes indicates that amino acid compositions, as well as certain dinucleotide combinations, are the most important features for determining optimal growth temperature. We also predict optimal growth temperatures for uncultivated species as well as consider the implications for inferring extinct ancestral species temperature preferences. Our results also demonstrate systematic biases in experimentally reported optimal growth temperatures. PhyRVM might also be of use in other settings where observational data are used to predict biological parameters of interest.

Microbial Evolution in Complex Environments

SMBE-PO-334

Fast adaptation of bacteriophage T7 in spatial analogue of serial dilution experiment

D. Fusco [1], C. Maberly [2], C. Rodrigues [2], O. Hallatschek [3], W. Moebius [2,*]

[1]University of Cambridge, Cambridge, [2]University of Exeter, Exeter, United Kingdom, [3]University of California Berkeley, Berkeley, United States

Abstract: The vast majority of evolution experiments in the lab focus on well-mixed or spatially discrete environments while much of microbial life occurs in spatially continuous environments. What is the resulting evolutionary pressure and evolutionary dynamics in the latter - and which aspects do we miss when concentrating on the former?
To address this question, we performed the spatial analogue of a serial dilution experimental evolution experiment with bacteriophage T7. Originally clonal populations of phage T7 were inoculated on lawns of *E. coli* BW25113 and plaque growth was monitored overnight. Subsequently, phage was picked from a region at the plaques' boundaries and reinoculated on fresh lawns once a day over the course of about two weeks. The speed of plaque growth increased remarkably. Measurements of the evolved populations in liquid reveal that this increase in front speed is associated with a pronounced change in life history trades, in particular a decreased burst size, making evolved phage more ‘rapacious’ than the corresponding ancestors.

Mitochondrial-Nuclear Interactions

SMBE-PO-338

Mito-nuclear interaction in the OXPHOS pathway genes during the Euphasmatodea radiation.

G. Forni[*], F. Plazzi [1], M. Iannnello [1], G. Piccinini [1], A. Cussigh [1], A. Luchetti [1], B. Mantovani [1]

[1]BIGEA, University of Bologna, Bologna, Italy

Abstract: Most of the stick insects diversity lies within the suborder Euphasmatodea - with over 3000 described species - while the sister-clade suborder Timematodea comprises only 22 species. The extant diversity of Euphasmatodea results from a continental-scale radiation with several components hypothesised to be involved. Aside from the abiotic factors, which may have triggered the radiation process, an extensive redesign of the OXPHOS pathway could have acted as a key adaptation. We used transcriptomic sequencing data to extract 13 mitochondrial and over fifty nuclear protein coding genes of 10 Timema spp. and 19 Euphasmatodea spp. and the sequences were then analyzed in a fossil-caibrated phylogenetic framework. For several mitochondrial gene, we observed higher dN/dS values in the branch leading to the radiation of Euphasmatodea compared to the one leading to Timematodea, which point out to an episodic positive selection. Considering the climactic and geological context, these observations suggest that Euphasmatodea underwent an adaptation in response to high temperature and low oxygen concentration which characterized the period of the clade origin, subsequently triggering the radiation. As in other metazoans a co-evolution of mitochondrial and nuclear OXPHOS genes has been observed, we investigated whether the nuclear components of the cellular respiratory pathway changed accordingly to the mitochondrial counterpart or if they did not share a common evolutionary trajectory.

Mitochondrial-Nuclear Interactions

SMBE-PO-339

Seeking for mtDNA structural determinants of organisms longevity

V. Shamanskiy [1,*], V. Timonina [1], P. Konstantin [2 3], K. Gunbin [4]

[1]Center For Mitochondrial Functional Genomics, School of Life Science, Immanuel Kant Baltic Federal University, Kaliningrad, Russian Federation, [2]Center for Integrative Genomics, University of Lausanne, University of Lausanne, [3]Swiss Institute of Bioinformatics, Lausanne, Switzerland, [4]Institute of Cytology and Genetics SB RAS, Novosibirsk, Russian Federation

Abstract: Till now the mtDNA determinants of longevity, i.e. germline mtDNA variants which cause aging and thus correlate with longevity, are poorly known. Long-lived mammals have increased GC content [doi: 10.1089/rej.2008.0676], decreased amount of direct (both perfect and imperfect) and inverted repeats [dois: 10.1016/j.tig.2004.03.003, 10.1093/bioinformatics/btx729, 10.1016/j.mad.2006.07.008]. Still there is a lack of understanding of the mechanisms of these correlations: which type of non-B-DNA structures better correlate with longevity – cruciform or triplex or slip-mispaired structures (generated by inverted, mirror, and direct repeats correspondingly), or G-quadruplex structures (generated by nucleotide context enriched with guanine)? Why? To answer these questions for each species with sequenced complete mitochondrial genome (~4500, extracted from GenBank) we 1) need to find all types of nucleotide repeats based on unified mtDNA-awared algorithm and 2) need to discriminate the most probable non-B-DNA structures associated with such repeats and other various nucleotide contexts.

We organized our work in three steps:

First, we called the most frequent non-B-DNA structures occurred in mtDNA: Z-DNA by SIST software; DNA and RNA-DNA triplexes by Triplex R/Bioconductor package and Triplexator software; quadruplexes using pqsfinder R/Bioconductor package. In order to analyze full spectra of possible DNA conformations we analyzed mtDNA based on dinucleotide properties data using DiProDB database, and various software from EMBOSS package. After the identification of DNA structures and patterns we correlate the frequency of its occurrence with animal lifespan based on data presented in [doi: 10.3897/natureconservation.5.5734] using the PIC, the corGrafen, and the corPagel functions of ape R package. This analysis demonstrated that CpG island frequency, quadruplex frequency, DNA bend, and DNA melting temperature, as well as various A-depleted nucleotide patterns (TGTC; TCGG; GTCG, GTGG etc.) significantly positively correlated with generation length. On the opposite side, the DNA mean free energy, DNA mean enthalpy, and various A-enriched nucleotide patterns (AGCA; GCAG; CGAG; GAGG) negatively correlated with generation length.

Second, we made our web-database ImtRDB (http://bioinfodbs.kantiana.ru/ImtRDB/) where we stored and analyzed mtDNA repeats annotated by our algorithm in all Chordata species with sequenced complete mitochondrial genome. This database is focused on interspersed repeats of four basic classes with different level of degeneration (perfect and non perfect). In order to call these repeats we implemented simple dot-matrix-based algorithm, which fits two important mitochondrial DNA properties: circularity and an excess of short repeats.

Third, the link between DNA structures / patterns correlated with species lifespan and repeat evolutionary conservation makes it possible to uncover the molecular bases of longevity determinants origination and evolutionary fixation. Our results demonstrated that these regions enriched with mtDNA transcriptional and translational functional regions, therefore the regulation of mtDNA transcription and translation are the one of main keys in determining the organismal longevity.

Mitochondrial-Nuclear Interactions

SMBE-PO-341

Loss and import: tracing the evolution of mitochondrial tRNA gene loss and functional replacement in Silene

J. Warren [1,*], D. Sloan [1]

[1]Biology, Colorado State University, Fort Collins, United States

Abstract: The mitochondrial genome is integral to some of the most fundamental biological processes including energy conversion and the origin of species, yet mitochondrial gene content can vary dramatically between taxa. This research aims to elucidate the role of nuclear gene duplication, targeting, and adaptation events has on mitochondrial gene content through exploring the functional replacement of mitochondrial tRNA genes. There exist extreme differences in mitochondrial tRNA gene content across eukaryotes, with some species having far fewer tRNA genes than are necessary for protein synthesis. Maintaining mitochondrial translation thus requires the functional replacement and import of nuclear-encoded tRNA genes. Here, I present tRNA-seq and transcriptomic data from multiple closely related flowering species (angiosperms, genus *Silene*) that have dramatic differences in mitochondrial tRNA gene content. These data suggest that some of the nuclear-encoded tRNA genes replacing those which have been recently lost from the mitochondrial genome are distinct from the tRNA genes used for cytosolic translation, implying that duplicated nuclear tRNA genes have gained mitochondrial import and facilitated the replacement and loss of the mitochondrial tRNA counterparts. The replacement of mitochondrial tRNAs with anciently divergent nuclear tRNA species raises numerous unanswered questions about the identity and evolution of the enzymes that must process and edit these newly imported tRNAs. Because of *Silene'*s recent and ongoing mitochondrial tRNA gene loss, the system presents an opportunity to investigate what coevolutionary mechanisms facilitate mitochondrial genome variation and the widespread occurrence of functional replacement of mitochondrial tRNAs.

Mitochondrial-Nuclear Interactions

SMBE-PO-343

Investigate the potential of nuclear and mitochondrial ribosomal ratio as an index of animal growth rate

W.-L. Kong [1 2,*], R. Machida [3], Y.-Y. Lin [3], T. Miki [4], W. Makino [5], J. Urabe [5], S.-H. Gu [6]

[1]Department of Life Science, National Taiwan Normal University, [2]Biodiversity Program, Taiwan International Graduate Program, Academia Sinica and National Taiwan Normal University, [3]Biodiversity Reseach Center, Academia Sinica, Taipei, Taiwan, [4]Department of Environmental Solution Technology, Faculty of Science and Technology, Ryukoku University, Shiga, [5]Division of Ecology and Evolutionary Biology, GRaduate School of Life Sciences,, Tohoku University, Miyagi, Japan, [6]Department of Zoology, National Museum of Natural Science, Taichung, Taiwan

Abstract: Quantification of growth is one of the important steps to understand the flow of elements and energy in food webs although in-situ estimation of the growth rate in general is very difficult. The present study proposes the nuclear and mitochondrial ribosomal ratio as a novel index of animal growth rate. The ribosome is a biomolecule composed of ribosomal-RNAs and -proteins, through which various proteins are synthesized using information encoded in mRNAs. Two types of ribosomes can be found in metazoan cells: mitochondrial ribosome and nuclear-encoded cytosolic ribosome. Functional roles of the mitochondrial ribosome (production of the proteins required to generate adenosine triphosphate, which provides energy to the cell) and cytosolic ribosome (production of diverse proteins required for various biological reactions, including cell growth and division) are different. Furthermore, ribosomes are macromolecules that require large amounts of environmentally limiting elements, such as phosphorus and nitrogen. Therefore, it is hypothesized that the allocation of those elements between the two ribosomes changes with nutrient availability. The present study conducted laboratory growth experiments of Daphnia magna using various food concentration and temperature treatments. As a result, positive correlation was found between nuclear-encoded cytosolic to mitochondrial ribosomal (Nuc/Mito ribosomal) ratio and somatic growth rate. A significant positive correlation was found between food concentration treatments and the ratio, but not between temperature treatments and the ratio. Our results demonstrate that the ratio between nuclear-encoded cytosolic and mitochondrial ribosome is an effective growth rate estimator.

Mitochondrial-Nuclear Interactions

SMBE-PO-344

Evolutionary history of nuclear proteins affects their mitochondrial targeting

E. Yirmiya [1], E. Hazkani-Covo [1,*]

[1]Department of Natural and Life Sciences, The Open University of Israel, Ra'anana, Israel

Abstract: Evolutionary history of nuclear proteins affects their mitochondrial targeting

Erez Yirmiya and Einat Hazkani-Covo

Department of Natural and Life Sciences, The Open University of Israel, Ra'anana, Israel

Mitochondrion was once a free-living proteobacterium. Following endosymbiosis of proteobacteria into an archon host, genes were transferred from the mitochondrial ancestor to the host genome, a process known as endosymbiosis gene transfer (EGT). Today most proteins needed for the mitochondrial function are encoded on the nuclear genome and are then targeted into the mitochondria. We tested how the evolutionary history of eukaryotic proteins affect the targeting to the mitochondria using organelle localization data from Gene Ontology as well as the protein phylogeny. Specifically, we studied how protein origin: archaeon, eubacteria or eukaryotic, as well as how the current targeting of other family members are associated with protein charge and mitochondrial targeting signals (MTS).

Our results show that proteins from eubacteria origin tend to be targeted more to the mitochondria, and have higher charge and MTS than proteins from archaeon or eukaryotic origins. Additionally, proteins that are part of phylogenetic trees of eubacterial origin tend to keep mitochondrial targeting throughout phylogeny by keeping their high charge and MTS. In contrast, protein families of archaeon or eukaryotic origin are more likely to gain or lose mitochondrial targeting throughout phylogeny, while their mitochondrial proteins have lower charge and MTS. This finding expands our understanding of how proteins of eubacterial origin are enriched in the mitochondrial proteome.

Mitochondrial proteins that are additionally targeted to one or more locations are known to have lower signal peptide compared to exclusive mitochondrial proteins. However, our results suggest that this effect is only true on trees that include non-mitochondrial proteins but is waived when all proteins are targeted to the mitochondria.

Our results indicate that the evolutionary origin of proteins as well as the current targeting history of their family members are important in mitochondrial targeting. It is possible that recent ancestors from a eubacterial origin, which have higher charge and MTS than recent ancestors of archaeon or eukaryotic origin, are less likely to osculate their charge and MTS throughout phylogeny. The difference in charge and MTS in mixed trees between exclusive and dual-targeted mitochondrial proteins may indicate that duality in those trees is a step from or to the mitochondria. In contrast, in eubacterial trees that are exclusively mitochondrial, duality is not an intermediate step of subcellular re-localization. Therefore, exclusive and dual targeted mitochondrial proteins on eubacterial trees have similar charge and MTS.

Mitochondrial-Nuclear Interactions

SMBE-PO-342

Plumage redness and mitochondrial function: mate choice for mitonuclear compatibility?

G. E. Hill [1,*], W. R. Hood [1], R. Grinter [2], C. Greening [3], A. N. Kavazis [4], Y. Zhang [5]

[1]Biological Sciences, Auburn University, Auburn, United States, [2]Biological Sciences, Monash University, [3]Biological Sciences, Monash, Clayton, Australia, [4]Auburn University, Auburn, [5]Biology, Memphis, Memphis, United States

Abstract: The mitonuclear compatibility hypothesis of sexual selection proposes that the primary purpose of mate choice is to match compatible sets of mitochondrial and nuclear OXPHOS genes to enable aerobic respiration. This hypothesis predicts that condition-dependent ornamental traits serve as honest signals of mitochondrial function. We tested the hypothesis that the process of carotenoid oxidation, and hence carotenoid coloration, is functionally linked to mitochondrial respiration. Most birds with red feathers convert yellow dietary carotenoids to red carotenoids in an oxidation process requiring the cytochrome P450 enzyme CYP2J19. We observed high levels of red ketolated carotenoids associated with the hepatic mitochondria of molting wild house finches (*Haemorhous mexicanus*), and upon fractionation, we found the highest concentration of ketolated carotenoids in the inner mitochondrial membrane. We further found that the redness of growing feathers was positively related to liver mitochondrial function. Structural modeling of CYP2J19 further supports the hypothesis that ketolation is functionally linked to cellular respiration. These observations suggest that feather coloration serves a signal of core functionality through inexorably links to cellular respiration in the mitochondria and establishes the plausibility of choice for mitonuclear compatibility.

Mitochondrial-Nuclear Interactions

SMBE-PO-348

Differential gene Expression and Co-Expression Network Analysis in the Manila Clam Ruditapes philippinarum Reveal Candidate Genes Involved in Mitochondrial Biology, Sex Determination, and Shell Biogenesis.

R. Xu [1,*], M. Iannello [1], L. Milani [1], F. Ghiselli [1]

[1]Department of Biological, Geological, and Environmental Sciences, University of Bologna, Bologna, Italy

Abstract: The Manila clam *Ruditapes philippinarum* is an important species in biology and aquaculture. It is characterized by the only known evolutionarily stable exception to the strict maternal inheritance of mitochondria, that is the doubly uniparental inheritance (DUI) by which two mitochondrial lineages are present: one transmitted through eggs (F-type) and the other through sperm (M-type). Because of such segregation, the conspecific F/M mtDNA sequence divergence reaches 50%, making this system a precious model for studying mitochondrial inheritance, heteroplasmy, and mito-nuclear coevolution. Moreover, *R. philippinarum* is one of the major cultured species in the world (4,228,000 tonnes in 2016). Despite that, basic genomics information about this species is scarce. In this study, we characterized the transcriptome of mature gonad, mantle, and adductor muscle of 15 females and 15 males (90 samples in total), focusing on sex- and tissue-specific transcription of genes involved in mitochondrial biology, sex determination/differentiation, and shell biogenesis. We found a large fraction of sex-biased transcripts in the gonad, whereas the adductor showed fewer differences, and no difference was found in mantle. We also performed a co-expression network analysis which revealed a total of 6,933 genes parsed into 11 modules, 3 of which were identified as sex-related modules and one module as specific for mantle. We identified candidate sex-specific genes including members of *Wnt*, *Sox,* and *Dmrt* family that might be responsible for the sexual differentiation and also tissue-specific genes such as *Hox* family genes and *IMSP5* in mantle that might be involved in shell biogenesis.

Mitochondrial-Nuclear Interactions

SMBE-PO-346

A new layer of genetic regulation in the mitochondrial genome: small mitochondrial RNAs

A. Pozzi [1,*], D. K. Dowling [1]

[1]School to Biological Sciences, Monash University, Melbourne, Australia

Abstract: Several studies have linked mitochondrial genetic variation to phenotypic modifications; albeit the identity of the mitochondrial polymorphisms involved remains elusive. The search for these polymorphisms led to the discovery of "highly-transcribed mitochondrial small non-coding RNAs encoded" (*smithRNAs*). These RNAs have been identified in several species already, and in most cases are encoded within other mtDNA-encoded genes, such as tRNAs. However, small RNAs are usually produced when an mRNA is degraded, thus sparkling controversy as to whether smithRNAs are functional RNAs or remains of degraded mRNAs.

To investigate the functional role of the smithRNAs, we leveraged data from six model organisms: *H.sapiens*, *M.musculus*, *G.gallus*, *D.rerio*, *D.melanogaster*, and *C.elegans*. Our dataset includes small RNA sequencing libraries from 37 studies. We aligned small RNAs from libraries of each of species to their mitochondrial genomes, and extracted transcriptional signatures of these RNAs.

We identified many species-specific smithRNAs. While most smithRNAs are usually conserved among tissues, they vary among species. In Chordata, the smithRNAs are concentrated within tRNAs, but in invertebrates are spread across all genes. We then analyzed immunoprecipitation-sequencing data to investigate links between smithRNAs and RNA interference. Among all species, we identify smithRNAs able to bind proteins involved in retrotransposon defense (Ago3,Aub, Miwi2) and gene regulation (Ago2, Alg5). Analyses of experiments utilizing knock-outs of key proteins involved in small RNA processing (Dcr, Ago2, Dgcr8, Xpo5, Miwi, Piwil1,) suggest the biogenesis of these smithRNAs is independent of these proteins.

Our work provides new insights into the link between mitochondrial genotype and phenotype.

Mitochondrial-Nuclear Interactions

SMBE-PO-349

Mito-nuclear Coevolution: a Perspective from Bivalve Oxidative Phosphorylation Genes

G. Piccinini [1,*], M. Iannello [1], G. Puccio [2], F. Plazzi [1], F. Ghiselli [1]

[1]Department of Biological, Geological and Environmental Sciences, University of Bologna, Bologna, [2]Department of Agricultural, Food, and Forest Sciences, University of Palermo, Palermo, Italy

Abstract: In Metazoa, 4 out of 5 complexes involved in oxidative phosphorylation (OXPHOS) are formed by subunits encoded by either mitochondrial (mtDNA) or nuclear (nuDNA) genomes, therefore mito-nuclear coevolution is expected. Previous works have shown co-adaptation of subunits encoded by different genomes and observed higher evolutionary rates of the nuclear components. This pattern—named "nuclear compensation hypothesis"—was proposed as being adaptive, resulting in the nuclear subunits compensating for the faster-evolving mitochondrial genes. In this study we analysed the sequence evolution (dN/dS) of 80 OXPHOS genes in 31 species of bivalve molluscs, a taxon showing extraordinary mtDNA variability and a distinctive mitochondrial biology. Overall, our data show clear signals of coevolution, since the nuclear subunits of complexes formed by genes encoded by both genomes (Complex I, III, IV, and V) experience higher evolutionary rates than those formed by nuDNA-encoded subunits only (Complex II). Interestingly, Maximum Likelihood trees obtained with either mtDNA-encoded or nuDNA-encoded OXPHOS genes have concordant topologies, despite previous phylogenomic works on bivalves showed that randomly-chosen nuclear markers yield a different topology in respect to mitochondrial ones. However, we did not find evidence of nuclear compensation when comparing the two genomes: mitochondrial genes showed higher dN/dS, contrarily to what observed in previous works. Moreover, evolutionary patterns of mitochondrial subunits were not coherently associated to their nuclear counterparts, and no site-specific signals of compensatory positive evolution were detected. Our analysis shows peculiar deviations from the coevolutionary patterns reported in other Metazoa, and we propose reconsiderations on the nuclear compensation hypothesis.

Mitochondrial-Nuclear Interactions

SMBE-PO-350

Within- and between-individual mtDNA heterogeneity in a naturally heteroplasmic species

M. Iannello [1,*], H. Asgharian [2], S. Bettinazzi [3], S. Breton [3], F. Ghiselli [1], L. Milani [1]

[1]Department of Biological, Geological, and Environmental Sciences, University of Bologna, Bologna, Italy, [2]University of California San Francisco, San Francisco, United States, [3]Department of Biological Sciences, University of Montreal, Montreal, Canada

Abstract: Mitochondrial DNA (mtDNA) has a fundamental role in evolution, energy production, cell biology, and disease. Since each cell of a multicellular eukaryote can contain up to tens of thousands of mitochondria—and each mitochondrion harbors multiple copies of mtDNA—a single individual carries a large and heterogeneous population of mtDNAs, a condition known as heteroplasmy. Heteroplasmy was once believed to be rare and/or linked to disease, but it is now clear that is a common condition. Genetic variability is the engine of evolution and studying mtDNA heterogeneity is fundamental to understand mitochondrial biology.

In this work, we investigated the within- and between-individual variability of mtDNA in *Ruditapes philippinarum*, a bivalve mollusc showing an unusual mechanism of mitochondrial transmission, the doubly uniparental inheritance (DUI). In the DUI system, two separate lineages of mitochondria exist: one is inherited through females (F-type) and the other through males (M-type). DUI has been reported in ~100 bivalve species and the conspecific mtDNA sequence divergence between the two lineages can reach 43%, corresponding to a divergence time of ~250 Myr. This makes DUI the only known stable exception to the maternal inheritance of mitochondria. Heteroplasmy is common in DUI animals and given the high sequence divergence between the two lineages this system can be an excellent model for studying mitochondrial heteroplasmy and mito-nuclear coevolution. We performed a mitochondrial enrichment in different tissues of female and male samples and used high-throughput sequencing to investigate the mtDNA heterogeneity among tissues, between sexes and across individuals.

Mitochondrial-Nuclear Interactions

SMBE-PO-345

PATERNALLY-INHERITED mtDNA VARIANTS AND SPERM PERFORMANCE

S. Bettinazzi [1,*], S. Nadarajah, A. Dalpé [1], L. Milani [2], P. Blier [3], S. Breton [1]

[1]Department of biological sciences, University of Montreal, Montreal, Canada, [2]Dipartimento di Scienze Biologiche, Geologiche e Ambientali, University of Bologna, Bologna, Italy, [3]Département de biologie, Université du Québec à Rimouski, Rimouski, Canada

Abstract: The selective neutrality of mitochondrial DNA (mtDNA) is nowadays strongly undermined. Studies on reproductive fitness indicate that specific mtDNA haplotypes strongly affect sperm motility and fertility, while having feeble effect on females. Because of the strict maternal inheritance (SMI) of mitochondria, mtDNA variants that are deleterious for male fitness, but with positive or even neutral effect on females can be retained in a population, a phenomenon known as “Mother’s curse”. A stable exception is the doubly uniparental inheritance (DUI) of mitochondria in bivalves. This system entails two mtDNA lineages that evolves independently and are transmitted separately through oocytes and sperm. This makes DUI an exclusive model to evaluate the result of direct selection on sperm mitochondria and its contribution to the reproductive fitness of males. In this study, we tested how maternal and paternal mtDNA variants impact bivalve sperm performance and bioenergetics in multiple DUI and SMI species of bivalves. Variation was also tested following inhibition of the main metabolic pathways, as well as after the introduction of oocyte-derived chemoattractants. Our results highlighted i) a divergence in sperm performance between DUI and SMI species, ii) a different proportion of aerobic and anaerobic energy production, although the aerobic metabolism contributes the most, iii) a metabolic shift in DUI sperm after the detection of egg chemical cues, towards a more combined strategy of energy production. Results are discussed in the light of the adaptive value of mtDNA variation, direct selection on paternal mitochondria, male-energetic adaptation and its evolutionary implications.

Molecular basis of neural circuit and behavioral evolution
SMBE-PO-351
The making of a receptor: an evolutionary scenario for 5-HT3 receptor in Chordata

G. Reales [1,*], V. R. Paixão-Côrtes [2], M. C. Bortolini [1]
[1]Departamento de Genética, Universidade Federal do Rio Grande do Sul, Porto Alegre, [2]Instituto de Biologia, Universidade Federal da Bahia, Salvador, Brazil

Abstract: Serotonin (5-HT) is one of the oldest and most important neurotransmitters, involved in sensorial perception, sexual behavior, cognition and memory, among many other essential functions. The serotonin type 3 receptor (5-HT3) is the only serotonergic pentameric ligand-gated ion channel, and upon activation it elicits a rapid, excitatory response in the central and peripheral nervous systems of vertebrates. 5-HT3 receptors may be composed of different subunits, encoded by five different genes in humans (*HTR3A-E*). However, we lack information about the processes that shaped 5-HT3 receptor evolution in other vertebrates. To elucidate the evolutionary history of 5-HT3 receptors in Chordata, we scanned 211 chordate genomes for 5-HT3 family members and inferred the main evolutionary events occurred during the gene family evolution, using synteny analysis, molecular modelling and phylogenetics. Our results reveal a complex evolutionary history. We found *HTR3A* in all species surveyed, coherently with its essentiality to form 5-HT3 receptors, while other genes underwent numerous events of gene loss, pseudogenization, and/or duplications during its evolution, highlighting the necessity for an updated nomenclature. We also describe a new putative family member, absent in mammals but present in birds, reptiles, and fishes. Finally, we propose two alternative scenarios for the origin and evolution of the 5-HT3 receptor family, one assuming independent duplications across lineages, and other assuming early duplications followed by subsequent losses.

Molecular basis of neural circuit and behavioral evolution
SMBE-PO-352
The deep homology of brains and apical organs.
R. Feuda*

Abstract: The sea urchin apical organ is considered the larval brain and derives from the embryo's apical domain. However, information on the regulatory program used to specify this domain during development is scant.
In this work, we analyzed the spatial expression of over 30 of regulatory genes, involved in the specification of the apical domain and proneural cells within it. Our results indicate that the apical domain starts as a single domain at blastula stage and is rapidly patterned into at least seven domains (including few pro-neural cells) within 18h of development. All transcription factors analyzed are expressed in specific regions of the apical domain and at least 15 of them are also expressed in pro neuronal cells. We then focused on identifying regulatory interactions underlining the specification of the apical proneural cells. Using morpholino knockdown we show that*foxq2*, *hbn*, and *soxC* are required to specify proneural cells.
The evolutionary relationships of the larval apical organ and neurogenic structures in other animals are unclear. To assess whether the regulatory toolkit for the development of the apical organ is similar to other nervous structures in bilateria, we identified the orthologues for each of the 33 regulatory genes in other model systems (e.g. worm, fly and mouse). We found that the vast majority of transcription factors expressed in the sea urchin apical domain (>85%) are expressed also in the nervous system of other bilateria. This finding indicates that these regulatory genes might constitute a shared neurogenic regulatory toolkit in bilateria.

Molecular basis of neural circuit and behavioral evolution

SMBE-PO-357

The chromatin accessibility and cis-regulation on the brain of a migratory bird.

J. S. Lugo Ramos [1,*], G. Durieux [1], M. Liedvogel [1]

[1]Behavioural Genomics, Max Planck Institute for evolutionary biology, Plön, Germany

Abstract: The animal behaviour of flocks, herds and individuals moving between two different geographical areas during specific days of the year, create the fascinating and mesmerizing scenario of migratory seasons. For birds, migration depends on multiple adaptations optimized for flight, timing and navigational skills. Results from crossbreeding experiments, suggest that the heritability and genetics of those migratory traits rely on a few genes with large effects, but the underlying genetic architecture remains to be characterised. Population genomics and candidate gene approaches have so far been inconclusive about the identity of genes or molecular mechanisms related to bird migration, owing to limitations of using purely sequence-based approaches. Here, we use a common garden experiment contrasting populations of European blackcaps *(Sylvia atricapilla)* with different orientation and migratory patterns during migration and out of the migratory season. We used ATAC-seq to compare chromatin accessibility in three brain areas expected to be relevant to seasonal hormonal regulation and navigational traits: i) the suprachiasmatic nucleus (SCN), a brain region controlling circadian and circannual regulatory processes, ii) hippocampus a brain region involved in spatial navigation and memory, and iii) Cluster N a brain region that is involved in magnetic compass orientation. Our preliminary results show that chromatin accessible regions are located in non-coding regions such as promoter sequences and intergenic regions with potential cis-regulatory sequences. We find clearest differences in chromatin accessibility between birds during migratory and off-season controls, rather than differences between brain regions. Our experimental design allows to couple these results with RNA-seq and genomic population analysis to obtain a broader picture of the molecular elements playing a role in migration.

Molecular basis of neural circuit and behavioral evolution

SMBE-PO-355

The regulatory and coding sequences of Trnp1 co-evolve with cortical folding in mammals

Z. Kliesmete [1,*], L. Wange [1], B. Vieth [2], W. Enard [2], I. Hellmann [2]

[1]Anthropology and Human Genomics, Ludwig Maximilian University of Munich, München, [2]Anthropology and Human Genomics, Ludwig Maximilian University of Munich, Munich, Germany

Abstract: Mammals show extensive variation in relative brain size and gyrification, including recurrent independent increases and decreases of these traits. Thus, a comparative approach integrating phenotypic evolution with sequence evolution to identify significant correlations is a promising strategy to identify genetic elements important for the cortical evolution of humans and other species. Here, we investigate the contribution of both regulatory and coding sequences of Trnp1 to brain size and folding in mammals. Trnp1 is the first protein identified to control tangential and radial cortical expansion by regulating neuronal progenitor proliferation and cell fate (Stahl et al, 2013).

We detected a strong positive correlation between the protein sequence evolution (dN/dS ratios) and the encephalization coefficient (EQ) as well as the gyrification index (GI), across a phylogeny of 33 mammalian species. Moreover, using a site model, we find evidence for positive selection in sites located within intrinsically disordered regions of the protein.

Furthermore, we use a massively parallel reporter assay to quantify the enhancer activity of >3000 sequences representing seven regulatory regions of Trnp1 in humans and their orthologues in 75 mammals. We find that the activity of one of the assayed regulatory regions also correlates with gyrification and encephalization using phylogenetic generalized least squares (PGLS).

In summary, we link both protein sequence and regulatory sequence evolution of Trnp1 to the evolution of brain size and folding, and show that at least the protein evolution was driven by Darwinian selection.

Molecular basis of neural circuit and behavioral evolution

SMBE-PO-354

Genomic architecture underlying the evolution of a novel form of social organisation.

R. Pracana, E. Stolle [1], Y. Wurm [*]

[1]Queen Mary U London, London, United Kingdom

Abstract: Variation in social behavior is common yet our knowledge of the mechanisms underpinning its evolution is limited. A rare exception is in the fire ant *Solenopsis invicta*: Alternative variants of a supergene region, carried by a pair of "social chromosomes", SB and Sb, determine a key social trait.

Colonies including only SB/SB individuals have a single queen, the ancestral form of social organisation. If many SB/Sb heterozygotes are present, the colony instead has a derived form of social organisation with up to dozens of reproductive queens.

We characterise the evolution of this system, which we date to 500,000 years ago using comparative genomics, long molecule sequencing, comparative transcriptomics and population sequencing. Our results shed light on the changes in genomic architecture, gene content and gene network structure underpinning how a new form of social organisation comes to be.

Molecular basis of neural circuit and behavioral evolution

SMBE-PO-356

Age- and daytime affect octopamine and dopamine receptor expression in the mushroom bodies of honeybee foragers

T. Peng [1], C. Grüter[*]

[1]University of Mainz, Mainz, Germany

Abstract: Honeybees use the waggle dance to communicate about the location of profitable food sources. Experienced foragers (older) follow fewer dances and rely more on memory to locate a food sources compared to younger foragers. Octopamine (OA) and dopamine (DA) signalling in the brain might play important roles in mediating the decision to follow dances (use social information) or return to known food locations (use private information), as they are involved in reward perception. In accordance with this, a recent study has found that systemic treatment with OA and DA affects the use of social and private information in honeybee foragers. Thus, age related changes in dance following might be linked to age related changes in OA and DA signalling. Here, we explore whether the expression of OA and DA receptor genes (DopR1, DopR2, OctR1 & Octβ2R) depends on forager age and the time of day. To this end, we introduced newly emerged, marked bees into observation hives and captured them when they were either ~3 weeks (young foragers) or ~5 weeks (old foragers) old. Bees were caught either early in the morning (~9 a.m.), around noon (~12 a.m.), in the afternoon (~4 p.m.) or at night (~10 p.m.). We used qPCR to quantify gene expression in the mushroom bodies, a brain area known to be important for information processing and integration. We found that expression levels of all four receptors were significantly higher in older foragers than in younger foragers. Furthermore, DopR1 showed significant down-regulation in the morning, whereas OctR1 showed significant up-regulation in the morning. The expression of the other genes did not change during the day. These results are consistent with the hypothesis that age and/or experience related changes in waggle dance following are linked to age and/or experience related changes in OA and DA signalling.

Mutation Rate Evolution

SMBE-PO-382

Compensating Frameshifts Caused by Multiple Insertions and Deletions Are Common in RNA Virus Proteins

D. Park [1], C. J. Goh [1], J. S. Lee [1], Y. Hahn [1,*]

[1]Department of Life Science, Chung-Ang University, Seoul, Korea, Republic Of

Abstract: RNA viruses are the most common molecular pathogens infecting most of life forms, including humans, animals, and plants. The remarkable evolutionary ability of RNA viruses is due to the high mutation rate of viral genomes. Protein sequence evolution is mainly driven by nucleotide sequence substitutions. It is generally accepted that insertions and deletions (indels) are not common because these events usually cause frameshift mutations and production of truncated defective proteins. However, when a second indel event occurs and compensates for the first indel in a coding sequence, its open reading frame will be restored and may produce functional proteins. We systematically analyzed RNA viral protein coding sequences and identified a large number of compensating frameshift cases. The compensating frameshift in viral coding sequences results in a large-scale amino acid sequence changes although their nucleotide sequences are almost identical. In consequence, the amino acid substitution rates can be overestimated when viral proteins with or without compensating frameshift are compared, which may lead to incorrect inference of the evolutionary history of RNA viruses. Compensating frameshifts accelerate amino acid sequence changes in RNA virus proteins, which may facilitate their evasion from the host defense systems and adaptation to new hosts. Consequently, the compensating frameshift is one of mechanisms of RNA virus evolution.

Mutation Rate Evolution

SMBE-PO-362

MOLECULAR PHYLOGENY AND DATE OF GEMINATED SPECIES PAIRS WITHIN THE FISH GENERA Dormitator AND Gobiomorus

S. Galvan Quesada [1,*], O. Dominguez Dominguez [1], O. Chassin Noria [1], H. Cano Camacho [2], J. C. Montero Castro [1], I. Doadrio Villarejo [3]

[1]Facultad de Biología, [2]Centro Multidisciplinario de Estudios en Biotecnoligía, Universidad Michoacana de San Nicolas de Hidalgo, Morelia, Mexico, [3]Biodiversidad y Biologia Evolutiva, Museo Nacional de Ciencias Naturales, CSIC, Madrid, Spain

Abstract: *Dormitator* and *Gobiomorus* fish species represent the amphidromous fauna of the tropical and subtropical coastal environments. Both genera share a wide range of geographical distribution in the Neotropics, and each genus include a pair of putative geminate species across the Panama Isthmus. The closing of the Central American Seaway represented a dramatic event separating aquatic organisms on each side of the Isthmus. However, comparisons of divergence times among geminate species pairs do not support a single and simultaneous divergence time for all taxa, suggesting that species might have responded differently to the Isthmus geological evolution and its new habitat development. We constructed a time-scaled molecular phylogeny of *Dormitator* and *Gobiomorus* using mitochondrial and nuclear DNA sequence data to infer and date the cladogenetic events and to relate them to the biogeographical history of Central America. Our results revealed the presence of geminate lineages with a mean divergence time in both genera of ~1.5 Mya, representing the most recent date reported for geminated lineages. The time intervals of this divergence (0.3 a 3.2 Ma) are similar to reported dates for geminated species pairs of coastal preferences, representing the last events of connection between both American oceanic slopes before the final closure of the Isthmus. We also identified additional geographically delimited lineages within the already recognized species, than appear to be shaped inside every slope by isolation processes that are largely congruent with geologic, climatic and oceanographic changes occurred as consequence of the Panama Isthmus establishment in the middle of the continent

Mutation Rate Evolution

SMBE-PO-364

Direct estimation of mutation rate using single generation in three species of ants

P. Cohen[*], S. Inbar[1], E. Privman[1]

[1]Department of Evolutionary and Environmental Biology, University of Haifa, Haifa, Israel

Abstract: Mutations drive evolution, yet only for a few species has the mutation rate been empirically measured. With advanced sequencing technologies it is now possible to directly identify mutations that occurred between a parent and an offspring and thereby infer the species mutation rate. We used genomic sequencing to estimate the mutation rate in three species of ants. The haplodiploid sex determination in ants provide convenient samples – sons that are the product of arrhenotoky. The genomes of 52 - 95 haploid brothers from each species were sampled using reduced representation genomic sequencing, allowing the identification of many thousands of single nucleotide polymorphism (SNPs) genomic markers. As males develop from unfertilized eggs, we could use the sequences of the sons to recreate their mother's genotype. Accounting for sequencing and mapping errors, we identify de novo mutations in the brothers, and estimated the mutation rate of each species. The use of haploid samples requires much lower sequencing depth than diploids, allowing for large sample sizes for a limited budget. Our findings, which cover three different families in as many species, provide valuable information that can be used in theoretical and applied evolutionary and population genetic studies.

Mutation Rate Evolution

SMBE-PO-365

Genome-wide mutation profile of Halobacterium salinarum

S. Kucukyildirim [1,*], O. Ozdemirel [1], A. M. Ummet [1], D. Ulusal [1]

[1]Biology, Hacettepe University, Ankara, Turkey

Abstract: Background: As the ultimate source of genetic variation, mutations have a central place in biology and not only drive adaptive processes but also contribute to genetic disorders, and in some cases, extinction. Thus, understanding the mutation rate is critical in determining rates of molecular evolution, estimating effective population sizes, understanding the impact of mutations on organismal fitness, and evaluating the power of drift, selection, and recombination in shaping genomes. Investigating organisms adapted to life in extreme habitats may further our understanding of the mechanisms of genetic stability. Today, by applying high-throughput sequencing technology to mutation-accumulation experiments, it is possible to generate the most direct and unbiased estimate of the genome-wide spontaneous mutation rate and spectrum of an organism thus allowing us to distinguish the intrinsic and extrinsic forces driving the mutation process. Previously, genome-wide spontaneous mutation rates have been estimated for a variety of prokaryotes and eukaryotes. However, in order to understand how DNA replication and repair cooperate and ultimately determine the genome-wide mutation rate across the tree of life, it is necessary to expand these experimental assays to the unexplored Archaea domain. In this study, we proposed to determine a genome-wide view of the spontaneous mutations in the halophilic archeon *Halobacterium salinarum*by using the most direct and unbiased method, combining mutation-accumulation experiments and whole-genome sequencing. *H. salinarum*is obligate halophile requiring high salt conditions to grow, because it is aerobic and mesophilic, *H. salinarum*can be grown in conditions much like those used for other model organisms such as the bacterium *Escherichia coli*and the model yeasts *Saccharomyces cerevisiae*and *Schizosaccharomyces pombe*. This work enables us to explore evolutionary forces and molecular mechanisms that may have shaped the mutation rate and spectrum of *H. salinarum*.

Results and Discussion: Our methods measure genome-wide rate, spectrum and distribution of spontaneous mutations. Whole-genome sequencing of 80 mutation accumulation lines of *H. salinarum*after an average of ~1000 cell divisions yielded a base-substitution mutation rate of 0.0015 per genome per generation, which is surprisingly similar to the consensus value of mesophilic organisms with DNA genomes. Transitions were found more frequently than transversions, and contrary to all other characterized G/C-rich prokaryotes, *H. salinarum*A/T mutation bias was observed. Whole genome sequencing of mutation accumulation lines provides the comprehensive insights of mutations and reveals what factors of molecular and genomic architecture affect the mutational process. Our study provides a more complete view of how several mechanisms of mutation, mutation repair, and bias act simultaneously to produce the raw material for evolution.

Mutation Rate Evolution

SMBE-PO-361

Imperfection of molecular clock in bacteria is due to episodes of increased mutation rate in mutator phenotypes

S. Garushyants [1 2,*], G. Bazykin [1 2], M. Gelfand [1 2]

[1]IITP RAS, [2]Skoltech, Moscow, Russian Federation

Abstract: The uniformity of the rate of sequence evolution, known as the molecular clock, helps molecular dating, phylogenetic reconstruction and has served the ground for the neutral theory. The molecular clock, however, is overdispersed, and its fidelity differs between lineages. The reasons for the change in substitution rate include change in generation times, effective population size and environmental conditions. Still, much of the difference between lineages remains unexplained. Disruption of replication or repair machinery in experimental bacterial populations can result in a mutator phenotype which accumulates mutations orders of magnitude more rapidly than the wildtype population, but whether such genotypes contribute to interlineage divergence in evolution in nature is unknown. Besides increased mutation rate, mutator phenotypes are associated with biased frequencies of individual types of mutations. Here, we design an approach to make use of the differences in mutational patterns to estimate, for an evolving lineage, the fraction of time it has spent as mutator. We estimate this fraction for the evolution of lineages in E. coli. We show that larger fractions of time spent as mutator correspond to longer phylogenetic lineages. Differences in mutation rates caused by disruption of replication machinery is explain most differences between lineages, and of the overdispersion of molecular clock. Although theory predicts that mutator phenotypes are deleterious in a constant environment, the high fraction of evolution spent by lineages in the “mutator” mode found here suggests that bacteria spend much of their life under fluctuating selection, having to adapt to a constantly changing environment.

Mutation Rate Evolution

SMBE-PO-360

Loss of GroEL mutational buffering affects fitness and mutational rate under highly bottlenecked population dynamics

B. Sabater-Munoz [1 2,*], R. Montagud-Martinez [1], M. A. Fares [1 3], C. Toft [1 4]

[1]Spanish National Research Council (CSIC) / Institute for Plant Molecular and Cellular Biology (IBMCP), Valencia, Spain, [2]Smurfit Institute of Genetics, Trinity College of Dublin, Dublin, Ireland, [3]Genetics, Smurfit Institute of Genetics, Trinity College of Dublin, Dublin, Ireland, [4]Spanish National Research Council (CSIC) / Institute for Integrative and Systems Biology (I2SysBio), Paterna, Spain

Abstract: Chaperones are involved in the folding of nascent client proteins, in the prevention of unfolded polypeptides aggregation and in the rescue of unfolded ones due to environmental stresses. The pioneering works of Rutheford and Lindquist, and those of Fares' team, and Touriki and Tawfik, have highlighted the importance of chaperones in buffering mutational effects by allowing for the adaptive evolution of its client proteins. These adaptive leaps might explain how ancient symbiosis still persist even under a strong genetic drift regime. But, how would an adapted consortium deal with the loss of this key system? When the organism lacks this rubustness system (as happens in many *Mycoplasma* species), proteome evolution becomes independent of protein folding. But what happens when the organism proteome is depending on its chaperone folding capabilities, and this system fails?

Experimental evolution of *Escherichia coli*under high-expression rate of GroEL is only possible when the system is subjected to strong genetic drift, as overexpression is significantly costly. Despite this limitation, the loss of GroEL overexpression increase the extinction rate, observing an equilibrium between GroEL level and fitness. By challenging *E. coli* to daily single-cell bottlenecks under high GroEL overexpression, we found that after a certain number of generations, a number of compensatory mutations arose in the system allowing to decrease the GroEL level while not effecting the fitness. How these two parameters, structural stability and functional innovation interact, still deserves further research.

Mutation Rate Evolution

SMBE-PO-383

The Good, the Bad and the Hairy: somatic mutational load in human tumors is better mirrored by Great Apes than by human genome-wide mutation densities

J. M. Heredia-Genestar [1,*], D. Juan [1], T. Marquès-Bonet [1 2 3], A. Navarro [1 2 3 4]

[1]Institute of Evolutionary Biology (CSIC-UPF), [2]CNAG-CRG, Centre for Genomic Regulation (CRG), [3]Institució Catalana de Recerca i Estudis Avançats (ICREA), [4]National Institute for Bioinformatics (INB), Barcelona, Spain

Abstract: Nucleotide variability varies across genomes. Using both, human-primate divergence and number of sites segregating in human populations, many studies have shown how mutation density is distributed in the human genome. Mutation distribution in human germlines correlates with several genomic features such as GC-content, CpG density, recombination rate, or replication time. The distribution of somatic mutations in tumors has also been widely studied. In human cancers, it does correlate poorly with human germline variation. In contrast, its distribution patterns are strongly associated with chromatin organization in the tumor's cell-of-origin.

We studied the genome-wide distribution of mutation density in the germline of chimpanzee and gorilla, measured as segregating sites per Mbp. Mutation density in both Great Apes has a moderate correlation with human variation. Strikingly, mutation distribution across the genomes of both Great Apes presents a much stronger correlation with mutation density in tumors than the surprisingly weaker human-tumor correlation.

Comparison of regional mutation densities between humans and Great Apes reveals that, in regions with different relative mutation densities between humans and Apes, tumors have densities like those of Great Apes. This effect is partially mediated by the higher diversity of Great Apes, although it has an inherently human-exclusive component. Interestingly, regions more densely mutated in humans than in Apes present higher GC-content, replicate earlier, and are enriched in open chromatin, promoters, and enhancers. Overall, our results suggest that human mutation densities are driven by human-specific factors and thus question their suitability as proxies of the mutational background in human tumors.

Mutation Rate Evolution

SMBE-PO-390

Dissecting the substitution rate variation within mitochondrial and nuclear genomes among Branchiopoda (Crustacea) lineages.

A. Luchetti [1,*], G. Forni [1], C. Savojardo [2], A. M. Skaist [3], P. L. Martelli [2], R. Casadio [2], S. J. Wheelan [3], B. Mantovani [1]

[1]Biological, Geological and Environmental Sciences, [2]Pharmacy and Biotechnology, University of Bologna, Bologna, Italy, [3]Oncology, Sidney Kimmel Comprehensive Cancer Center, The Johns Hopkins University School of Medicine, Baltimore, United States

Abstract: Variations of substitution rate are caused by a number of factors, linked either to life-history traits or to selective pressures or even to differential mutation rate. The crustacean class Branchiopoda dates back to the Cambrian and includes the orders Anostraca (fairy shrimps), Spinicaudata (clam shrimps), Cladocera (water fleas) and Notostraca (tadpole shrimps), the latter known for their long-term morphological stasis. The comparison of the newly sequenced notostracan genomes and available branchiopod genomes, at both the mitochondrial and the nuclear level, showed a lower substitution rate in tadpole shrimps. From the dN/dS analysis mitochondrial genes did not appear under selective pressures. However, they showed a shift toward a GC-richer nucleotide composition during the evolution of Anostraca, Cladocera and Spinicaudata that could explain the observed variation in substitution rate. On the other hand, the analysis of 433 orthologous gene clusters still showed significantly lower substitution rate in Notostraca, although a gene-by-gene analysis revealed a more complex situation: some genes (N=66) appeared evolving faster in Cladocera and Spinicaudata, and some others (N=28) in Notostraca, while the majority (N= 339) basically showed an equal rate among lineages. dN/dS analysis did not show correlation of possible selective pressures with changes in rates and, at variance of mitochondrial genome, the same hold for the GC content analysis. Overall, present data suggest that substitution rates in Branchiopoda could be driven by a shift in nucleotides composition in the mitochondrial genome but not in the nuclear one, while selective pressures do not seem to not play a role. Variation in life-history traits could be, therefore, more effective in shaping the substitution rate of nuclear genes.

Mutation Rate Evolution

SMBE-PO-386

High mutation rate evolution during adaptation to high salinity or sub-inhibitory concentrations of gentamicin.

C. Rose [1], M. Finnegan, F. Gatchitch, M.-P. Dubois, M. Callens, L. Pradier, S. Bedhomme[*]

[1]Centre d'Ecologie Fonctionnelle et Evolutive, CNRS, Montpellier, France

Abstract: The drastic increase in mutation rate due to hyper-mutator genotype has been shown to be favoured when a population is adapting to stressful conditions. This is because they allow a faster provision of beneficial mutations and hitchhike with the adaptive mutations they have triggered. Here, we use an *Escherichia coli* strain with a high probability of becoming hypermutator to investigate whether the nature of the stress and its intensity have an effect on the selection of hypermutator genotypes. To do so, we experimentally evolved *E. coli* populations across a gradient of salinity and of subinhibitory concentration of gentamicin. We show that the nature of the stress and the associated target size for adaptive mutation strongly influences the probability for a population to become hyper-mutator. We then decipher the similarities and differences in the molecular bases of adaption for hyper-mutator and wild-type genotypes.

Mutation Rate Evolution

SMBE-PO-380

A first direct estimate of the sex-specific mutation rate in olive baboons and implications for the evolution of mutation rates

F. Wu[1,*], A. Strand[2], P. Andolfatto[2], J. Wall[3 4], P. Moorjani[5 6 7], M. Przeworski[1 2 7]

[1]Department of Systems Biology, [2]Department of Biological Sciences, Columbia University, New York, [3]Department of Epidemiology & Biostatistics, [4]Institute for Human Genetics, University of California, San Francisco, San Francisco, [5]Department of Molecular and Cell Biology, [6]Center for Computational Biology, University of California, Berkeley, Berkeley, United States, [7]co-supervised this work, -, -

Abstract: In humans, most germline mutations are inherited from the father. This male bias is widely interpreted as supporting a model in which mutations arise from DNA replication errors during spermatogenesis. Under this model, the lower male bias in mutation should be substantially lower in a closely related species with similar rates of spermatagonial stem cell divisions but shorter generation times. To test this hypothesis, we resequenced two 3-4 generation nuclear families (totaling 29 individuals) of olive baboons (*Papio anubis*), whose average reproductive ages are approximately 11 years in both sexes. We inferred sex-specific mutation rates by analyzing the data in parallel to three 3-generation human pedigrees, incorporating estimated false negative and false positive rates. By this approach, we estimate the mutation rate per generation in baboons to be approximately half that of humans, when their sex-averaged generation time is approximately a third. Strikingly, despite the much shorter generation times of baboons, the degree of male bias in mutations is similar to what has been estimated in humans and what we obtained. These findings are inconsistent with a model in which mutations are due to replication errors and changes in generation times drive the evolution of mutation rates. Our observations thus add to growing evidence for the importance of non-replicative mutation mechanisms and/or the mutagenic nature of early embryogenesis. Beyond these findings, the matched analysis of human and baboon data serves as a blueprint for broader comparisons of mutation rates across the tree of life.

Mutation Rate Evolution

SMBE-PO-367

Discovery of selfish mutations expanding in the male germline

R. Salazar [1], B. Arbeithuber [1 2], M. Ivankovic [1], G. Povysil [1 3], M. Heinzl [1], J. Proell [4], I. Tiemann-Boege [1,*]

[1]Johannes Kepler University Linz, Linz, Austria, [2]Penn State University, Pennsylvania, [3]Columbia University Irving Medical Center, New York, United States, [4]Faculty of Medicine, Johannes Kepler University Linz, Linz, Austria

Abstract: De novo mutations (DNM) are an important player in heritable diseases and evolution, yet little is known about the different mutagenic processes in our germline. Of interest are a few highly recurrent DNM associated with congenital disorders that have been described as selfish mutations expanding in the male germline. We have adapted an ultra-sensitive sequencing (USS) technology based on duplex-barcoding to distinguish both DNA strands, thus increasing the sensitivity to one mutation in 10^7 sequenced bases. This technology was applied to study the expansion of selfish mutations in the *FGFR3* gene in the male germline. A ~3000bp coding region of the *FGFR3* gene including the transmembrane and the tyrosine kinase domains was sequenced with USS in two DNA sperm pools from an old and young donor group. DNM with frequencies of 10^{-3} – 10^{-5} were identified in 41 exonic positions, several of which have been described in congenital disorders and/or cancer. Also, various substitutions are viable and have been reported in population data (gnomAD), albeit at lower frequencies. Mutations found in both pools show higher frequencies in older donors. Our data have important implications in understanding the transmission of new germline mutations with paternal age that have yet unknown consequences in the offspring.

Mutation Rate Evolution
SMBE-PO-363
Characterising and exploiting changes in mutation through application of non-stationary substitution models
G. Huttley [1,*]
[1]Research School of Biology, Australian National University, Canberra, Australia

Abstract: The evolvability of mutagenesis can manifest as changes to both the rate and spectrum of substitutions. The impact of such changes is to induce a non-stationary divergence process that will manifest at both the nucleic acid and amino acid levels. The pervasiveness of non-stationary processes in natural systems has remained an open question due the scarcity of suitable formal models for interrogating this property.

We have addressed this shortcoming by developing a general framework for the specification and interrogation of non-stationary Markov processes. Our approach simplifies the construction of arbitrary non-stationary processes and can be readily applied to numerous state spaces: nucleotide, dinucleotide, trinucleotide, codon and amino acids.

Applying non-stationary models to genomic data is critical. Applying standard time-reversible (hence stationary) models to natural data that evolved under a non-stationary process results in systematic biases: time-reversible models systematically overestimate genetic distance; and, underestimate the ratio of nonsynonymous to synonymous substitutions. These biases can profoundly mislead conclusions regarding the dynamics of molecular evolution.

Beyond illuminating the shortcomings of time-reversible models, the non-stationary class of models has exciting potential for understanding the origins of changes to mutation and for providing entirely new perspectives on genetic divergence. We demonstrate the former via analysis of perturbed genomic regions. We demonstrate the latter by characterising a Markov process that confers the ability to infer the direction of evolutionary time without requiring an outgroup.

PyCogent3 now makes this new class of models readily available, providing the full capabilities for exploring non-stationary models for genome data analytics. Written in Python 3, PyCogent3 takes advantage of the enhanced visualisation capabilities of the Jupyter notebook environment. It is also multi-processor aware, capable of running in parallel on a single machine or across thousands of machines.

Mutation Rate Evolution

SMBE-PO-366

The Rate of Molecular Evolution When Mutation May Not Be Weak

A. J. de Koning [1,*], B. DeSanctis [2 3], I. Krukov [3]

[1]Biochemistry and Molecular Biology, University of Calgary, Calgary, Canada, [2]University of Cambridge, Cambridge, United Kingdom, [3]University of Calgary, Calgary, Canada

Abstract: A great many results in molecular evolution and population genetics assume that mutation is weak, such that mutations arise in the population and go to their fates one at a time. While this assumption is often reasonable, growing evidence suggests that in some large and hyperdiverse populations, and for particularly fast mutation types, mutation may not always be weak. In this talk, I show that by exploiting recent advances in computational population genetics approaches for the rapid, direct analysis of Wright-Fisher type models, the rate of evolution (together with moments of the probability distributions for the times between fixations) can be directly and rapidly computed in realistic population sizes without approximation or additional assumptions beyond those expressed by the population genetics model. Using this approach, I will show that with increasing population mutation rates: 1) the rate of neutral evolution decelerates and is not equal to the mutation rate in general; 2) with bidirectional mutation, there is a mutation-rate that maximizes the rate of adaptive molecular evolution (when selection is constant in time); 3) that molecular evolution likely proceeds by successive fixations even when the population mutation rate is quite large (as long as the timescale of variation in selection is longer than that for fixation); and 4) that we can account for the effects of standing variation by directly computing the rate of substitution for out of equilibrium populations in models where selection changes in time. I will conclude by showing that these findings may have substantial impacts on inference of selection coefficients from across-species sequence comparisons (e.g., with mutation-selection models) and on d_N/d_S approaches. The difference between our main results and those from weak-mutation theory can be explained with a straightforward generalization of the ideas of Kimura and others.

Mutation Rate Evolution

SMBE-PO-371

Analysis of germline de novo mutation rates on exons and introns

M. Rodriguez [1,*], S. Casillas [2], D. Weghorn [1], A. Barbadilla [2]

[1]Center for Genomic Regulation, Barcelona, [2]Institute of Biotechnology and Biomedicine, Cerdanyola del Valles, Spain

Abstract: A main assumption of molecular population genetics is that genomic mutation rate does not depend on sequence function. Challenging this assumption, a recent study has found a reduction in the mutation rate in exons compared to introns in somatic cells due to an enhanced exonic mismatch repair system activity. If this reduction happens also in the germline, it can compromise studies of population genomics, including the detection of the footprint of selection by using introns as proxies of neutrality.

We compiled and analyzed published germline *de novo* mutation data to test if the exonic mutation rate is also reduced in germ cells. We found no reduction in the mutation rate in exons compared to introns in the germline genome, in contrast to what has been previously described in somatic cells. Therefore, there is no evidence of an enhanced mismatch repair system activity in exons with respect to adjacent introns in germline cells. We demonstrate that this result is largely independent of the mutation extended sequence context beyond trinucleotides. Finally, we present a statistical approach to estimate genome-wide variability of mutation rates across introns and exons caused by external mutation rate covariates.

Our findings are consistent with known processes of genomic lesion formation and repair, but may point to a dichotomous nature of germline and soma with respect to mutational and DNA repair processes.

Mutation Rate Evolution

SMBE-PO-378

Most human RNA viruses show extraordinarily stringent selective constraints on protein evolution

J.-J. Lin [1,*], M. J. Bhattacharjee [1], C.-P. Yu [1], Y. Y. Tseng [2], W.-H. Li [1 3]

[1]Biodiversity Research Center, Academia Sinica, Taipei, Taiwan, [2]Center for Molecular Medicine and Genetics, Wayne State University, Detroit, [3]Department of Ecology and Evolution, University of Chicago, Chicago, United States

Abstract: In RNA virus evolution, natural selection occurs mostly as negative selection. However, how much the intensity of negative selection varies among species and what factors affect the variation are not well explored. Here, we studied the ratios of nonsynonymous to synonymous substitution rates (dN/dS) in protein-coding genes of human RNA and DNA viruses and mammals. Among the 21 RNA viruses studied, 18 showed a genome-average dN/dS between 0.01 and 0.10, indicating that over 90% of the mutations are eliminated by negative selection. Only HIV-1 showed a dN/dS (0.31) higher than that (0.21) in mammals. In contrast, among the eight DNA viruses studied, four showed a ratio >0.21 and only two showed a ratio <0.10. We found that both positive selection and population size play significant roles in the dN/dS variation among species. Moreover, we found a correlation between dN and dS, which is likely correlated with mutation rate. This study increases our understanding of the mechanisms of RNA virus evolution.

Mutation Rate Evolution

SMBE-PO-369

Natural Bedrock radioactivity influences the rate and spectrum of mutation

N. Saclier [1,*], P. Chardon [2], F. Malard [1], L. Konecny [1], D. Eme [1], A. Bellec [1 3], L. Duret [4], T. Lefébure [1], C. J. Douady [1]

[1]LEHNA-UMR5023, University Claude Bernard Lyon 1, Villeurbanne, [2]LPC-UMR6533, Université Clermont Auvergne, Clermont-Ferrand, [3]EVS-UMR5600, Université Jean Moulin Lyon 3, Lyon, [4]LBBE-UMR5558, CNRS - Université Claude Bernard Lyon 1, Villeurbanne, France

Abstract: All organisms on earth are exposed to low doses of natural bedrock radioactivity. The mutagenic impact of chronic exposure to very low doses of bedrock radioactivity is difficult to study *in vivo*. In the absence of reliable data this effect has been extrapolated from clinical evidence following artificially high doses of radioactivity caused by nuclear tests or accidents. Here, we circumvent this roadblock using the unique ecological properties of subterranean isopods exposed to different levels of natural bedrock radioactivity. We show that endemic subterranean species living in areas with high bedrock radioactivity have a higher mutation rate (+20%) than those in areas with low radioactivity. We also found that species in high bedrock radioactive environment have a higher probability of mutation from G to T and its complement. This type of mutation is a hallmark of oxidative stress, suggesting that bedrock radioactivity generates free radicals that influence the rate and spectrum of mutation. In these isopods, we show an impact of naturally-occurring bedrock radioactivity variation on the mutation rate equivalent to that seen when a drastic change in generation time occurs. As these isopods eat sediment, the influence of bedrock radioactivity may be particularly high in these species compared to organisms with different diets. However, while a 20% mutation rate increase may have a significant impact on a species evolution, a direct health effect is unlikely. Indeed, for example in humans, other factors such as the paternal age have similar if not higher mutational effect without direct health consequences.

Mutation Rate Evolution

SMBE-PO-370

Slow-evolving sites negatively impact reconstruction of deep Tree of Life phylogenies

L. T. Rangel [1,*], G. Fournier [1]

[1]Department of Earth, Atmospheric & Planetary Sciences, Massachusetts Institute of Technology, Cambridge, United States

Abstract: Slow-fast analysis has been broadly used in phylogenetic reconstruction. The underlying assumption of this method is that fast-evolving sites do not retain accurate phylogenetic signal due to site saturation via multiple substitution events. Therefore, removing these sites improves the signal-to-noise ratio in phylogenetic analyses, with the remaining slower-evolving sites preserving a more reliable record of deep evolutionary relationships. However, slow-evolving sites are less likely to have experienced substitutions along shorter branches, and therefore could be less likely to retain evolutionary information about many bipartitions. Here we show that slow-fast analysis can potentially negatively impact the accuracy of phylogenetic reconstruction in both real and simulated aligned sequence datasets. Simulated alignments generated under a predefined phylogeny, modeled after Tree of Life ribosomal protein datasets, consistently show that slow-evolving sites are less likely to recover true bipartitions than even the fastest-evolving sites. Furthermore, site rate is positively correlated with accurately recovering shorter branched bipartitions. We further tested our hypothesis using the concatenated ribosomal protein dataset published by Hug et al. (2016). We show that phylogenetic signal present among both the slowest and fastest evolving sites is significantly less compatible to the overall signal than within other sites. Furthermore, for this dataset, trimming fast sites, slow sites, or both has distinct levels of impact on phylogenetic reconstruction under different evolutionary models. This is perhaps most evident in the resulting placements of Eukarya and Asgard groups, which are especially sensitive to the implementation of different trimming schemes.

Mutation Rate Evolution

SMBE-PO-384

Pedigree-based estimation of germline mutation rate of Rhesus macaque (Macaca mulatta)

L. Bergeron [1,*], S. Besenbacher [2], J. Bakker [3], J. Zheng [4], M. H. Schierup [5], T. Gilbert [6], G. Zhang [1]

[1]Departement of Biology, University of Copenhagen, Copenhagen, [2]Department of Molecular Medicine, Aarhus University, Aarhus, Denmark, [3]Animal Science Department, Biomedical Primate Research Centre, Rijswijk, Netherlands, [4]China National Genebank, BGI, Shenzhen, China, [5]Bioinformatics Research Centre, Aarhus University, Aarhus, [6]Centre for GeoGenetics, Natural History Museum of Denmark, Copenhagen, Denmark

Abstract: Germline mutations are the source of every evolutionary adaptation. Thus, determining the rate at which *de novo* mutations occur in different species is essential for understanding evolution and mapping the genetic basis of biological traits. Next Generation Sequencing has allowed pedigree-based methods to precisely estimate per generation mutation rates by comparing whole genome sequences of parents and offspring. This has been widely applied in studies in human trios (Maretty et. al, 2017; Jónsson et. al, 2017) and is only beginning to be used in other primate species. However, since these studies have used different sequencing strategies and analyses, it is difficult to understand whether the observed variation in mutation rate among species is due to the different methodologies or their factual biological disparity. In our study, we establish a pipeline for estimation of the mutation rate with high depth sequencing. We applied the pipeline to 19 trios of Rhesus macaque (*Macaca mulatta*) and generated whole genome sequencing at a depth of 80X per individual. This high depth allows us to identify *de novo* mutations based on strict filters and to estimate per generation rate of mutation on single nucleotide and structural variations. Our results show a positive effect of the paternal age on the mutation rate, similar to humans. To further understand the patterns of mutation rate variation on a macroevolutionary scale, we also collected 173 trios from 43 mammalian, 18 bird, 10 fish and 7 reptile species. These unprecedent genomic data will largely expand our knowledge on mutation rate evolution.

Mutation Rate Evolution

SMBE-PO-358

Genome resequencing of the complete order Crocodylia to investigate patterns of evolution

K. Brittain [1,*], A. Osmanski [2], E. Jones [1], A. Suh [3], D. Ray [2], J. Gongora [1]

[1]Sydney School of Veterinary Sciences, The University of Sydney, Sydney, Australia, [2]Biological Sciences, Texas Tech University, Texas, United States, [3]Department of Ecology and Genetics, Uppsala University, Uppsala, Sweden

Abstract: The order Crocodylia consists of approximately 23 extant species in three families; Alligatoridae (alligators and caimans), Crocodylidae (crocodiles) and Gavialidae (gharial). These species inhabit a range of tropical and subtropical environments across the Americas, Africa, and Australasia. Previous research by the International Crocodilian Genomes Working Group (ICGWG) has sequenced the genomes of representatives from each of these families - the American alligator, Australian saltwater crocodile, and Indian gharial. That project revealed ancestral patterns of genome evolution among archosaurs and established these taxa as an emerging model given their robust immune systems, low cancer incidence, low heterozygosity, and their unusually slow rate of genomic evolution. Despite these unique adaptations and the conservation status of many crocodilians, the genomic basis of these adaptations remains poorly understood. To address this problem, a collaborative sequencing effort from the ICGWG is generating genome drafts of all remaining species of the order to improve our knowledge of their evolutionary history and unique biology. De novo and reference-based assemblies have been generated using a combination of sequencing methods.. We aim to study whether the low evolutionary change found in our early genome work is extended at the family and species level, as well as what variations may have occurred at a geographical level within crocodiles and alligators, with an interest to understanding adaptation to varying environments. To date, all but three species have been assembled. Preliminary codon evolution analysis has identified positively selected genes for further study, and preliminary immunological analysis has suggested some variations in the immune gene characterisation in crocodilians compared to birds and mammals. Together, these genome assemblies represent a molecular toolkit required to increase our understanding of the unique evolutionary history of this enigmatic yet charismatic order.

Mutation Rate Evolution

SMBE-PO-391

Migration promotes mutator alleles in subdivided populations

Y. Raynes [1,*], P. Sniegowski [2], D. Weinreich [1]

[1]Brown University, Providence, [2]University of Pennsylvania, Philadelphia, United States

Abstract: Mutator alleles that elevate the genomic mutation rate may invade non-recombining populations by hitchhiking with beneficial mutations. Mutators have been repeatedly observed to take over adapting laboratory populations and have been found at high frequencies in both microbial pathogen and cancer populations in nature. However, we have recently shown that mutator mutations are only favored by selection in sufficiently large populations and transition to being disfavored as population size decreases. In contrast, fitness-affecting beneficial and deleterious mutations never experience such sign inversion.

This population size-dependent sign inversion in selective effect suggests that population structure may also be an important determinant of mutation rate evolution. While large populations may favor mutators, subdividing such populations into sufficiently small subpopulations (demes) might effectively inhibit them. On the other hand, migration between small demes that otherwise inhibit hitchhiking may promote mutator fixation in the whole metapopulation. Here, we use stochastic, agent-based simulations and evolution experiments with the yeast *Saccharomyces cerevisiae* to explore the role of population subdivision and migration in mutation rate evolution.

We show that mutators can, indeed, be favored by selection in subdivided metapopulations composed of small demes connected by sufficient migration. More surprisingly, mutators fare even better in metapopulations with intermediate and rare migration, demonstrating that population subdivision plays a previously unsuspected role in promoting mutator success. We propose that the mechanism underlying the evolutionary advantage of mutator alleles in metapopulations with only minimal migration, is, intriguingly, related to that observed in studies of the evolution of recombination and cooperation in structured metapopulations.

Mutation Rate Evolution

SMBE-PO-389

Direct measurement of the de novo mutation rate in a non-model primate using linked-read sequencing reveals a high mutation rate and a lack of mutations in CpG sites

C. R. Campbell [1,*], M. dos Reis [2], J. W. Poelstra [1], K. E. Hunnicutt [3], G. P. Tiley [1], P. A. Larsen [4], A. D. Yoder [1]

[1]Department of Biology, Duke University, Durham, United States, [2]School of Biological and Chemical Sciences, Queen Mary University of London, London, United Kingdom, [3]University of Denver, Denver, [4]Veterinary and Biomedical Sciences, University of Minnesota, St. Paul, United States

Abstract: Spontaneous germline mutations are the raw material on which evolution acts and knowledge of their frequency and genomic distribution is crucial to understanding how evolution operates at both long and short timescales. The rate and spectrum of de novo mutations have so far been directly characterized only for a limited set of organisms, yet it is critical to investigate a wide range of species to examine the generality of patterns that have been identified so far. Using high-coverage linked-read sequencing of a family pedigree (n=8) of grey mouse lemurs (*Microcebus murinus*), we estimate the mutation rate at 1.64×10^{-8} mutations per basepair per generation, an estimate that is higher than for most previously characterized mammals. Our result underscores the lack of a negative relationship between effective population size and mutation rate in primates as would be predicted by the drift barrier hypothesis, a pattern that we show does continue to be visible at broader phylogenetic scales. Unexpectedly, we found little sex bias in the parent-of-origin of mutations, a transition-transversion ratio near 1, and only a modest overrepresentation of mutations at CpG-sites. Estimates of the mutation rate critically affect estimates of divergence time, recombination rates, ancestral population sizes, and other fundamental evolutionary processes and parameters. These results both contribute to the understanding of lemur evolution and encourage the continued surveillance of mutation rates across the tree of life.

Mutation Rate Evolution

SMBE-PO-387

Insights into human mutation rate evolution from sequencing of trios of other great apes

S. Besenbacher [1], C. Hvilsom [2], M. R. Belles [3], M. Coll [4], M. H. Schierup [5,*]

[1]Molecular Medicine, Aarhus University, 8000 Aarhus, [2]Genetics, Copenhagen Zoo, Copenhagen, [3]Aarhus Univerisity, [4]Aarhus University, 8000 Aarhus , Denmark, [5]Bioinformatics Research Centre, Aarhus University, 8000 Aarhus, Denmark

Abstract: Several large studies have estimated the mutation rate in humans using whole genome sequencing of hundreds of parent-offspring trios. These studies have consistently estimated a yearly mutation rate of approximately $0.43x10^{-9}$. This rate is, however, markedly lower than prior estimates of ~$1x10^{-9}$ per year from phylogenetic comparisons of the great apes. We used a probabilistic approach to estimate mutation rates in other great apes using Illumina deep sequencing (~35X) of 1 chimpanzee extended trio (father, mother, child and grandchild), 6 previously sequenced chimpanzee trios, 2 gorilla extended trios and 1 orangutan trio. Assuming that the relationship with maternal and paternal age is similar to humans we estimated a higher mutation rate in chimpanzee, gorilla and orangutan compared to humans by a factor of 1.5+/- 0.10, 1.5+/- 0.23 and 1.42+/- 0.22, respectively. These large differences in inferred rates contrast with the fact that the overall great apes' phylogeny almost adheres to a molecular clock with e.g. the chimpanzee branch only being 2% longer than the human branch and the gorilla branch 6% longer than the human branch. We interpret this as evidence that the apparent recent slowdown on the human branch since chimpanzee divergence by ~1/3 has occurred recently. We will discuss possible reasons for this apparent dramatic slow-down in human mutation rate using new data from a larger number of human trios and from a larger number of great apes trios. This includes direct new estimations of the relative rates of mutations in human and Neanderthals and estimation of the proportion of replication-dependent and replication independent de novo mutations. Furthermore, we quantify the expected contribution from changes to the age of puberty and the male and female generation times.

Mutation Rate Evolution

SMBE-PO-373

Characterizing recurrent site de novo mutations in humans

M. Noyes [1,*], T. N. Turner [1], K. Harris [1], E. E. Eichler [1 2]

[1]Genome Sciences, University of Washington School of Medicine, [2]Howard Hughes Medical Institute, University of Washington, Seattle, United States

Abstract: Every individual is born with approximately 75 *de novo* mutations (DNMs), new mutations that arose in their parents' germlines. Evolutionary models predict that no two DNMs should arise at the same location, but mutational hotspots exist throughout the genome. At some hotspots, recurrent site mutations occur, where the same single nucleotide substitution occurs at the same position in multiple unrelated individuals. The objective of this study was to characterize the molecular properties of recurrent site *de novo* mutations (rsDNMs) in humans.

DNMs were identified in children of 1,837 families from the Simons Simplex Collection using two variant callers. This dataset consists of 30-fold whole-genome sequences obtained from blood DNA from families of two parents and two children, one child affected with autism and one unaffected. Mutations were classified using data from the UCSC Genome Browser and enrichment was calculated using the Chi squared test. Approximately one sixth of mutations were assigned a parent of origin based on the presence of inherited variants on the same sequencing reads.

From the 3,674 children, I identified over 19,000 rsDNMs, of which 1,079 were found in more than 5 individuals. There was no significant difference in the number of mutations recurrent only in affected and unaffected siblings. CpG dinucleotides and regions mapping to the last 5 Mbp of human chromosomes were enriched for rsDNMs. I observed a significant enrichment of rsDNMs in repetitive DNA, including short tandem repeats and SVAs, even after controlling for high-sequence identity regions with reduced mapping quality. There were almost equal numbers of maternally inherited (n=313) and paternally inherited (n=302) rsDNMs, even though there were more than twice as many paternally inherited nonrecurrent mutations. Understanding this pattern of rsDNMs is critical for characterizing germline mutational processes and identifying mutational hotspots that may have been excluded in the modeling of disease and evolutionary processes.

Mutation Rate Evolution

SMBE-PO-375

Inferring the genomic landscape of mutation rates from polymorphism data

G. V. Barroso [1,*], J. Y. Dutheil [1]

[1]Evolutionary Genetics, Max Planck Institute For Evolutionary Biology, Ploen, Germany

Abstract: In sexually-reproducing species, the distribution of diversity along the genome is influenced by variation in (1) branch lengths of genealogies as a result of genetic drift; (2) effective population size as a result of natural selection and gene-flow; (3) recombination rate, via GC-biased gene conversion and modulation of linked selection; and (4) the rate of *de novo* mutations generating polymorphism. Quantifying the relative importance of these factors in shaping patterns of diversity is a major goal of population genomics, however, spatial variation in the mutation rate has largely been neglected by empirical studies, in part due to its difficult estimation. This is problematic because regions of the genome with differential mutation rates will either mimic or dilute the signal of selection, leading to false positives and negatives in genome-wide scans. Here we present a new statistical model (iSMC) that infers the genomic landscape of mutation rates from polymorphism data while accounting for the effect of demography and recombination rate variation. Our simulation study demonstrates that iSMC has high accuracy in diverse scenarios. Using population genomic datasets of different organisms, we find that spatial variation in mutation rates is a significant explanatory factor of the distribution of diversity in natural populations. This result suggests that mutation rate heterogeneity should be more often incorporated in data analysis. Our explicit model of the mutation landscape allows parametric inference from polymorphism data, thus fostering research in species where large-scale sequencing of pedigrees is not feasible.

Mutation Rate Evolution
SMBE-PO-374
Spatial Patters of Substitutions in Bacterial Genomes
D. F. Lato [1,*], G. B. Golding [1]
[1]Biology, McMaster University, Hamilton, Canada

Abstract: Increasing evidence supports the notion that different regions of a genome have distinct molecular properties. This variation is abundant in bacterial genomes where essentiality and gene expression decrease with distance from the origin of replication, while at the same time mutation rate increases. There is limited research on how molecular trends such as substitution rates vary between the assortment of genomic structures, for example linear, circular, and multi-repliconic genomes. In this work, we mapped extant and ancestral substitutions to the phylogenies of *Escherichia coli*, *Bacillus subtilis*, *Streptomyces*, and *Sinorhizobium meliloti*, quantifying how many substitutions were at each coding and non-coding position of the genome. Previous studies indicate that the number of substitutions should increase with distance from the origin. Our analysis instead demonstrates the opposite, that the number of substitutions decreased when moving away from the origin of replication in six of the ten replicons analyzed. The replicons that did not follow this trend were pSymB of *S. meliloti* and the chromosome of *Streptomyces* where we found the number of substitutions increase with increasing distance from the origin. We explore how these substitution trends impact the functional categories of genes and their placement within a bacterial genome.

Mutation Rate Evolution

SMBE-PO-372

Including genetic recombination in phylogenetic reconstruction to decipher the evolutionary history of Hepatitis B virus.

A. Kocher [1,*], T. Vaughan [2], A. Herbig [3], J. Krause [3], D. Kühnert [1]

[1]Transmission, Infection, Diversification & Evolution group, Max Planck Institute for the Science of Human History, Jena, Germany, [2]Biosystem science and eng., ETH Zürich, Basel, Switzerland, [3]Archeogenetics, Max Planck Institute for the Science of Human History, Jena, Germany

Abstract: Hundreds of millions of people from all continents are currently infected by Hepatitis B virus (HBV). Despite the availability of numerous full-genome sequences, including dated ancient HBV sequences from archaeological remains, much mystery persists regarding the evolutionary history of this virus. In particular, most existing phylogenetic reconstructions suggest complicated biogeographical scenarios and are hardly compatible with current knowledge of prehistoric human migrations.

HBV often causes long-lasting chronic infection, and is present at high prevalence in many regions of the world. This creates frequent opportunities for co-infection and genetic recombination, which is known to occur even in genetically distant lineages. This process is thought to have significantly shaped current HBV diversity and some of the main genotypes have been identified as hybrids. Genetic recombination may introduce contradicting phylogenetic signal and bias in molecular clock estimates, hence known recombining sites are usually excluded from phylogenetic analysis.

Our study aims at shedding light on the evolutionary history of HBV by explicitly accounting for genetic recombination in phylogenetic inference. Taking profit of the recently recovered ancient genomic data, we use a time-calibrated Bayesian approach that allows joint estimation of recombination events and phylogenetic relationships. Our preliminary results suggest that we can identify previously suspected as well as unknown inter-lineage recombination events, and provide a more parsimonious phylogenetic hypothesis for the evolutionary history of HBV.

Mutation Rate Evolution

SMBE-PO-377

Cytosine methylation affects mutation rate beyond the modified base itself

V. Kusmartsev [1 2,*], T. Warnecke [1 2]

[1]MRC London Institute Of Medical Sciences (LMS), [2]Institute of Clinical Sciences (ICS), Imperial College London, London, United Kingdom

Abstract: DNA modifications can affect the mutation rate of the bases they modify. In particular, methylated cytosines at CpG sites have long been known to spontaneously deaminate at higher rates than their unmethylated counterparts. Intriguingly, data from NMR, molecular dynamic simulations, and other studies suggest that methylation can have an impact on the mechanical properties of the DNA helix that goes beyond the modified base. As a consequence, DNA methylation might impact both lesion formation and repair dynamics beyond the methylated base. There have been prior experimental and genomic studies suggesting methylation incurs a mutational liability to proximal bases in human, but they did not control for potential confounding effects. Here, we investigate this question in humans, Arabidopsis and rice. We compare mutation rates around methylated and unmethylated sites using rare single nucleotide polymorphisms as a proxy. Matching methylated and unmethylated sites for sequence context and chromatin state, we find that methylation in human has a mild protective effect on mutational incidence around the focal CpG. This is in conflict with previous reports, which we show suffered from unreliable SNP calls. In contrast to humans, we find a large increase in relative mutation risk associated with methylation in Arabidopsis and rice. We trace the opposite effects of methylation to differences in the methylation machinery between humans and plants. Specifically, regions of the Arabidopsis genome with greatest active demethylation show a lower mutational risk, compared to regions with greater RNA-directed DNA methylation. Our findings suggest that methylation does indeed affect mutation rates beyond the methylated base itself and does so in a manner that implicates the action of species-specific DNA methylation and demethylation machineries. This work highlights the importance of accounting for DNA modifications and their dynamics when characterising the genomic distribution of mutation rates.

Mutation Rate Evolution
SMBE-PO-376
Characterization of the mutations in the C-terminal region of the Cystic Fibrosis transmembrane conductance regulator (CFTR)

O. Jefri*, R. Ford

Abstract: Cystic fibrosis is a major inherited disease associated with morbidity and mortality around the globe. CFTR is an important ion channel that regulates the movement of chloride ions. Thus, mutations in the CFTR gene disrupt the transport of epithelial fluid and result in the development of classical symptoms such as impaired lung function. Among the protein domains involved in CFTR localisation and function, mutations affecting the C-terminal region have been implicated in cystic fibrosis. Understanding the impact of the mutations is important in developing effective therapeutic approaches.
This study determined the impact of six cystic fibrosis mutations related single amino acid substitutions on the overall conformation of the C-terminal peptide and on the interaction between CFTR and the scaffold protein NHERF1, with specific focus on the PDZ1 domain. The effects of single amino acid substitution were investigated by several biochemical and biophysical analyses. The data showed that the mutations did not affect the successful creation, amplification and production of the C-terminal^{42aa} CFTR peptides nor their overall secondary structure. The PDZ1 domain of the NHERF1 was also successfully created, expressed and purified.
The single amino acid mutations were shown to alter this protein-protein interaction in a predictable and unpredictable way compared to robust binding of the wild-type C-terminus CFTR with NHERF1 PDZ1. The phenotypic manifestation of a given CFTR mutation in Cystic Fibrosis is complex and heavily dependent on the impact of the mutation on the different functions and interactome of CFTR.

Mutation Rate Evolution
SMBE-PO-379
Predicting tRNA gene activity using local mutation rate variation

R. Corbett-Detig [1,*], B. Thornlow [1], T. Lowe [1]
[1]Biomolecular Engineering, UC Santa Cruz, Santa Cruz, United States

Abstract: Predicting gene expression in redundant gene families, such as transfer RNAs, is an important challenge in functional genomics. Because the mature transcripts are often identical, standard approaches for measuring gene expression levels from read counts are insufficient. Nonetheless, despite their identical sequences, tRNAs exhibit extremely high variation in their expression. Among the hundreds of tRNA genes in the human genome, many are constitutively expressed in all tissues, while many others appear to be entirely dormant, and still others are expressed in only a subset of tissues. In some cases, gene activity can be inferred indirectly from chromatin accessibility data, but these data are not available for the vast majority of species, and therefore alternative approaches are needed for predicting gene expression levels.

We recently discovered that tRNA genes experience mutation rates approximately 10 times greater than the genome-wide average. The distribution and profile of the mutations in and near to tRNA genes indicate that this excess is a result of transcription-associated mutagenesis, a result of mutations introduced during transcription. Consequently, sequence divergence and polymorphism near to tRNA genes is strongly correlated with gene expression levels.

Here we exploit this signal to predict which tRNA genes are active across mammalian genomes. We develop a random forest classifier that uses local sequence features around each tRNA gene to infer its level of expression. After training this model on the human genome, we are able to classify actiivty of tRNAs in the mouse genome with 92% accuracy (despite 75 million years of divergence!). Additionally, we use a full set of tRNA gene orthologs across placental mammals in conjunction with our expression classifications to infer tRNA gene function evolution across the mammalian phylogeny. We find that changes between tRNA gene expression states are rare at orthologous loci indicating that genomic environment is the key determinant of tRNA gene expression. More generally, our results indicate that local mutation rate variation is a potentially powerful tool for functional genomic prediction.

Novel insights into evolutionary genetics from emerging technologies

SMBE-PO-413A

Exploring the functions of polymorphisms driving long-term balancing selection in humans

K. S. Velazquez-Arcelay [1,*], J. A. Capra [1 2 3 4]

[1]Biological Sciences, [2]Biomedical Informatics and Computer Science, [3]Genetics Institute, [4]Center for Structural Biology, Vanderbilt University, Nashville, Tennessee, United States

Abstract: Long-term balancing selection (LTBS) maintains allelic variation at intermediate frequencies in a population over long periods of time. In extreme instances, balancing selection has preserved multiple alleles over millions of years and through speciation events. For instance, the major histocompatibility complex (MHC) locus contains a trans-species polymorphism known to be under LTBS in primates, and host-pathogen interactions are thought to be a common source of LTBS. Several hundred instances of long-term balancing selection have been identified between the human and chimpanzee genomes. However, the functions and adaptive roles remain unknown for nearly all of them, and many are located in non-coding regions of the genome, making annotation challenging. We characterized previously identified trans-species polymorphisms likely under balancing selection since the common ancestor of humans and chimpanzees. To identify candidate functions underlying LTBS in humans, we integrated data from protein structures, genome-wide functional assays, eQTL studies, genome-wide association studies (GWAS), and phenome-wide association studies (PheWAS). We identify potential functions for 104 of 125 LTBS loci considered. These include: 7 that modify protein structure and/or function; 96 with evidence of gene regulatory function from GTEx or genome-wide functional genomics data; 37 with evidence of influencing traits from GWAS and PheWAS. These analyses revealed association with many immune system phenotypes, including response to pathogens, but also a range of other phenotypes, including cognitive performance, addiction, and cancer. We provide detailed functional hypotheses for four example LTBS regions with multiple lines of evidence supporting their function outside of the immune system. Overall, there was no significant enrichment for any specific functional annotations among the regions we analyzed. Our results suggest that interactions with pathogens are a main driver of LTBS in humans, but a range of additional functional constraints beyond the immune system can drive balancing selection. This highlights that substantial work is needed to understand the functional basis of balancing selection in humans.

Novel insights into evolutionary genetics from emerging technologies

SMBE-PO-418

The population and quantitative genetics of polymorphism and divergence for structural variation

M. Chakraborty [1,*], S. J. Macdonald [2], A. D. Long [1], J. J. Emerson [1]

[1]Ecology and Evolutionary Biology, University Of California Irvine, Irvine, [2]Molecular Biosciences, University of Kansas, Lawrence, United States

Abstract: Mutations that alter genome organization can lead to profound changes in functional elements, driving genome evolution, leading to genetic diseases, or serving as raw material for adaptation. However, such mutations remain elusive, with 40-80% of such complex events evading discovery, concealing their importance to genome and phenotypic evolution. Recent advances in long read sequencing and assembly algorithms facilitate high quality *de novo* assemblies of metazoan genomes. To create the first complete polymorphism and divergence map of genomewide SVs for a metazoan organism, we performed *de novo,* reference quality assembly of 14 *D. melanogaster* strains and three of its sibling species using long reads, following an assembly approach we developed. We developed software to identify structural variants (SVs) and build a comprehensive SV map. Our map identified thousands of SVs, many of them hidden, associated with candidate genes for complex phenotypes. These genes, including those for insecticide resistance, lifespan, and reproductive success, often harbor multiple rare SV alleles. Indeed, we show that such QTL candidates are statistically enriched for SVs, an observation that complicates genome wide association mapping. We used population genetics approaches to estimate the strength of selection on SVs using the histogram of allele frequencies, or the site frequency spectrum (SFS). The SFS analyses uncover pervasive purifying selection, with transposon insertions being most deleterious, followed by tandem duplications. We then estimated the rates of adaptation in SVs by integrating SV polymorphisms with interspecific SV variation derived from new reference quality long read *de novo* assemblies *D. simulans*, *D. sechellia*, and *D. mauritiana*, sibling species of *D. melanogaster*. These results underscore the importance of high quality genomes in comprehensive and accurate SV genotyping and their essentiality in understanding the evolution of genomes and phenotypes.

Novel insights into evolutionary genetics from emerging technologies

SMBE-PO-393

Development and optimisation of a real-time sequencing and analysis pipeline for Norovirus outbreaks in a clinical setting.

A. N. O Toole [1,*], L. Meredith [2], I. Goodfellow [2], A. Rambaut [1]

[1]University of Edinburgh, Edinburgh, [2]University of Cambridge, Cambridge, United Kingdom

Abstract: Norovirus is a leading cause of acute gastroenteritis worldwide, directly costing health-care systems $4.2 billion annually (Bartsch et al., 2016). Infections are normally self-limiting, with symptoms resolving within 2-4 days, however viral shedding and infectivity can continue for weeks following the ceasing of symptoms. Complications and chronic infection are common in the young, elderly and immunocompromised. Norovirus outbreaks in healthcare settings are complex, often with multiple strains circulating, and as it is an incredibly diverse virus, it is often very challenging to accurately reconstruct the epidemiology of these outbreaks using PCR diagnostics alone, which hinders control efforts. As Norovirus is a small, but fast-evolving RNA virus, whole-genome sequencing of the virus can help inform epidemiology. We present a novel approach to in-hospital epidemiology using multiplexed, tiled PCR amplification, the MinION sequencer and a custom analysis pipeline to implement real-time sequencing in a clinical setting. Using the MinION, rather than other sequencing technologies, allows small-scale sequencing to be carried out and makes it unnecessary to batch samples in order for sequencing to be cost-effective. We also trial use of within host level information to infer epidemiological links. We use sequence data to assess whether cases of chronic norovirus infection, which show striking temporal patterns of molecular evolution, act as a source of infection within a clinical setting. With this novel analysis pipeline, we address the challenges posed by a recombining, highly diverse virus with limited reference availability.

Bartsch, S. M., et al. (2016), `Global Economic Burden of Norovirus Gastroenteritis.', PloS one.

Novel insights into evolutionary genetics from emerging technologies
SMBE-PO-392
The evolution of fitness effects in a long-term experiment with bacteria
A. Couce [1,*], O. Tenaillon [2]
[1]Imperial College London, London, United Kingdom, [2]INSERM, Paris, France

Abstract: How the fitness effect of mutations changes with adaptation is one of the most fundamental questions in evolutionary genetics, being relevant to many central theories such as the origin of sex, the maintenance of genetic variation or the reproducibility of evolution. However, most attempts so far to address this issue have been limited by their small sample size or by a lack of characterization of the mutants employed. Here we circumvented these limitations by analyzing high-throughput sequencing data of transposon mutant libraries generated on bacteria from Lenski's Long-Term Evolution Experiment. One main finding is that the genome-wide distribution of fitness effects becomes depleted in beneficial mutations in later generations. This observation supports a modular epistasis model of adaptive evolution, in which a pattern of declining adaptability emerges without the need for any epistasis sensu strictu at the genetic level. Nonetheless, we also identified some instances of diminishing returns epistasis affecting particular genes, suggesting that both types of epistasis should be incorporated into any theory which attempts to make statistical predictions about evolution.

Novel insights into evolutionary genetics from emerging technologies

SMBE-PO-395

Toward understanding functional implications of genetic variation in splicing within and between species enabled by new technology

L. Mcintyre[*], A. M. Morse[1], A. Nanni, F. J. Pardo Palacios, J. Wedow, N. Choquette, E. Ainsworth, J. R. Newman, S. Nuzhdin, A. Conesa

[1]University of Florida, Gainesville, United States

Abstract: Eukaryotic genes are often translated into multiple mRNAs through Alternative Splicing (AS) and differential UTR processing. Different transcripts may have different functional properties and/or encoded proteins. Splicing is the mechanism by which Drosophila sex is determined, over 90% of the multi-exon genes in humans are alternatively spliced and AS is involved in numerous key biological processes. Despite the popularity of short read isoform reconstructions, there are limitations in short reads in their the ability to accurately assemble isoforms. Long read sequencing of transcriptomes offers an unprecedented ability to study the evolution of transcriptomes, but are not without their own technical challenges. We have performed two experiments, one on the model organism *Drosophila* and another on *Zea Maize*. In *D. mel* and *D. sim* adult male and female head tissue was assayed with PacBio, Illumina and ChIP allowing us to create a comprehensive picture of splicing differences between the sexes and species. In Maize we profiled leaf tissue from 5 genotypes in ambient and ozone conditions with PacBio and Illumina in order to understand whether the profound differences in the Maize genome among different genotypes was reflected in the transcriptome and whether diverse inbreds may have a differential transcriptional response to climate change. By leveraging the common challenges to transcriptome assembly in both experiments we were able to develop estimates of isoform quantities and test functional implications of differential isoform use between conditions.

Novel insights into evolutionary genetics from emerging technologies

SMBE-PO-394

fastGLOBETROTTER: efficient identification and dating of admixture events inference in large-scale population data cohorts

P. Wangkumhang [1,*], G. Hellenthal [1]

[1]UCL Genetics Institute, University College London, London, United Kingdom

Abstract: Intermixing, or "admixture", among populations, e.g. due to past migrations, has had a major influence in shaping genetic diversity. The software *GLOBETROTTER* (Hellenthal et al 2014) shows increased precision over other available techniques for characterising admixture events due to modeling haplotype information, i.e. associations among tightly linked Single Nucleotide Polymorphisms (SNPs). However, *GLOBETROTTER* also has increased computational demands, and hence cannot cope efficiently with the increasing availability of large, genetically homogeneous cohorts sampled from relatively narrow geographic areas. Therefore, we present a new statistical method, *fastGLOBETROTTER,* that both reduces computational time by a factor of 4-20-fold and increases accuracy relative to GLOBETROTTER. This new technique can exploit large-scale cohorts to provide more precise estimates of admixture dates and sources.

Through application of *fastGLOBETROTTER* to over 6000 European individuals, using over 2500 individuals as ancestry surrogates, we report new insights into admixture events across Western Europe. These include admixture dated to ~500-600CE from sources carrying DNA related to present-day West Asian and North African populations found in individuals within Belgium, France and parts of Germany. We also detect admixture from sources carrying DNA related to East Asians in individuals within Finland, Norway and Sweden at different times starting ~1900 years ago. Through these applications and additional simulations, we illustrate the time depths for which these approaches can reliably detect such past intermixing and the extent to which additional samples can unearth previously hidden events.

Novel insights into evolutionary genetics from emerging technologies

SMBE-PO-396

Single-molecule assembly of the Basket Willow, Salix viminalis, reveals the earliest stages of sex chromosome expansion

P. Almeida [1,*], E. Proux-Wera [2 3 4], A. Churcher [3 5], L. Soler [3 6], J. Dainat [2 6], P. Pucholt [7 8], S. Berlin [8], J. Mank [1 9 10]

[1]Department of Genetics, Evolution & Environment, University College London, London, United Kingdom, [2]Science for Life Laboratory, Uppsala University, [3]National Bioinformatics Infrastructure Sweden, [4]Department of Biochemistry and Biophysics, Uppsala University, Uppsala, [5]Department of Molecular Biology, Umeå University, Umeå, [6]Department of Medical Biochemistry and Microbiology, [7]Department of Medical Sciences, Section of Rheumatology, Uppsala University, [8]Department of Plant Biology, Swedish Agricultural University, [9]Department of Organismal Biology, Uppsala University, Uppsala, Sweden, [10]Department of Zoology, University of British Columbia, Vancouver, Canada

Abstract: Sex chromosomes are known to have evolved independently multiple times throughout the eukaryotes and are considered a prime example of convergent genome evolution. The current model of sex chromosome evolution posits that sex chromosomes evolve when recombination is halted between a pair of autosomes and this leads to a range of non-adaptive modifications, including the accumulation of repetitive sequence and gene loss on the sex-limited chromosome. However, because studies on sex chromosomes have primarily focused on old and highly differentiated sex chromosomes, the causes of recombination suppression and the pace at which degeneration subsequently occurs remain unclear. Here, we use long- and short-read single molecule sequencing approaches to assemble and annotate a draft genome of the basket willow, *Salix viminalis*, a species with a female heterogametic system at the earliest stages of sex chromosome emergence. Our single-molecule approach allowed us to phase a large fraction of the non-recombining region of the sex chromosomes and to perform detailed phylogenetic, sequence divergence and synteny analyses of the emerging Z and W haplotypes. We detected very low levels of Z/W divergence and found that gene order is still largely conserved between the Z and W homologs, albeit with a few genes apparently lost from the W and other W genes that are not syntenic with the Z. Furthermore, multiple lines of evidence suggest that selection against recombination is a more gradual process at the earliest stages of sex chromosome formation than would be expected from an inversion. Our results highlight the relevance of single molecule sequencing methods in understanding the earliest genomic consequences of recombination suppression as well as valuable insights into the initial stages of sex chromosome formation.

Novel insights into evolutionary genetics from emerging technologies

SMBE-PO-420

Mitogenome phylogeny and the evidence of heteroplasmy in fireflies extracted from short read sequencing data

P. Wonnapinij [1 2,*], A. Sriboonlert [1]

[1]Department of Genetics, Faculty of Science, Kasetsart University, [2]Omics Center for Agriculture Bioresources, Food and Health, Kasetsart University, Bangkok, Thailand

Abstract: Fireflies are mesmerizing bioluminescent insects belonging to the family Lampyridae, order Coleoptera. Despite 2000 species of fireflies worldwide, only 16 species have their mitogenome sequences available in GenBank database. This study aimed to gain insight into the evolutional relationship of fireflies using whole mitogenome data, with an emphasis on the fireflies species collected from Thailand. We reconstructed mitogenome sequences of 15 fireflies from short read whole genome sequencing data. These mitogenome sequences were annotated, compared and applied for phylogenetic analysis. Morphological based species identification suggested that 10 out of 15 fireflies belonging to 6 genera: *Pteroptyx*, *Sclerotia*, *Pyroceolia, Australoluciola*, *Curtos*, and *Luciola,* while the other 5 fireflies are unreported species potentially belong to genus *Trisinuata* and *Medeopteryx.* The comparative genome analysis showed that all Lampyridae mitogenomes are similar in size, gene content and gene arrangement. The mitochondrial phylogenomic analysis showed that members of *Pteroptyx* do not form a monophyletic clade. Even though the topology of *mtCOI* phylogeny differs from that of mitogenome phylogeny, this pattern holds. Further analysis that aimed to identify intra-individual variation showed that some fireflies carry mitochondrial DNA (mtDNA) heteroplasmy. Hence, this study not only helps us gain insight into the mitogenome evolution of fireflies, especially within subfamily Luciolinae but also indicates slightly disagreement between mtDNA sequence and morphological based genus classification. The observation of mtDNA heteroplasmy also suggests that the high throughput sequencing technology increases the probability of detecting this feature of mtDNA.

Novel insights into evolutionary genetics from emerging technologies
SMBE-PO-408
Adaptive evolution and function of uORFs (upstream open reading frames) in Drosophila
J. Lu*

Abstract: Upstream open reading frames (uORFs) in the 5' untranslated regions (UTRs) of messenger RNAs can potentially inhibit translation of the downstream regions that encode proteins by sequestering protein-making machinery the ribosome. Moreover, mutations that destroy existing uORFs or create new ones are known to cause human disease. Although mutations that create new uORFs are generally deleterious and are selected against, many uORFs are evolutionarily conserved across eukaryotic species. To resolve this dilemma, we used extensive mRNA-Seq and ribosome profiling to generate high-resolution genome-wide maps of ribosome occupancy and translational efficiency (TE) during the life cycle of the fruit fly *D*. *melanogaster*. This allowed us to identify the sequence features of uORFs that influence their ability to associate with ribosomes. We demonstrate for the first time that the majority of the newly fixed uORFs in *D*. *melanogaster*, especially the translated ones, are under positive Darwinian selection. We also show that uORFs exert widespread repressive effects on the translation of the downstream protein-coding region. We find that many uORFs are transcribed or translated in a developmental stage-, sex-, or tissue-specific manner. Our results suggest that during *Drosophila* development, changes in the TE of uORFs, as well as the inclusion/exclusion of uORFs, are frequently exploited to inversely influence the translation of the downstream protein-coding regions. Our study provides novel insights into the molecular mechanisms and functional consequences of uORF-mediated regulation.
Reference: Zhang H#, Dou SQ#, He F, Luo JJ, Wei LP, and Lu J* (2018). Genome-wide maps of ribosomal occupancy provide insights into adaptive evolution and regulatory roles of uORFs during Drosophila development. PLoS Biology. 16(7):e2003903.

Novel insights into evolutionary genetics from emerging technologies

SMBE-PO-401

Satellite DNA evolution in songbirds

V. Peona [1,*], M. Weissensteiner [1], A. Suh [1]

[1]Department of Ecology and Genetics, Uppsala University, Uppsala, Sweden

Abstract: Single-molecule technologies are now helping us to investigate the genomic 'dark matter', namely all those genomic regions that are missing or under-represented in most short read assemblies. Among these dark genomic corners, peri-centromeric and centromeric regions are the most elusive because (often) they are made of highly repetitive and homogeneous satellite DNA.

In birds, satellite DNA has been mostly studied in chicken while there are few studies in songbirds, e.g. in crows. Here we widen the study of satellite to the available genomes of songbirds. We take advantage of both strengths and weaknesses of current sequencing technologies to develop a new method to find satellite repeats and investigate their still unknown diversity and evolution.

Previous studies found that the chicken genome harbours short satellite repeat units of some hundred base pairs whereas in crow we find huge satellite repeat units of the size of about 14 kb (crowSat1).

The difference in size is striking and it is maintained in many songbird species. In particular we found that the majority of satellite repeats are homologous to one another, although with large structural differences, and to a small unit of crowSat1. Finally, thanks to the chromosome-level assemblies available and recombination landscapes, we were able to identify chromosome-specific satellite repeats and thus localise putatively different centromeres within the same species.

Novel insights into evolutionary genetics from emerging technologies
SMBE-PO-405
Structural genomic divergence and introgression during hominin evolution
S. M. Yan [1,*], R. C. McCoy [1]
[1]Department of Biology, Johns Hopkins University, Baltimore, United States

Abstract: Structural variation (SV) is a prominent source of genomic diversity within and between populations. These large (>50 bp) insertions, deletions, and inversions have the potential to impact phenotypes via disruption of regulatory and protein-coding sequences. Though these variants are poorly resolved by short reads, the development of long-read sequencing has dramatically expanded the catalog of SVs in contemporary human populations. Unfortunately, degradation of ancient DNA precludes application of long-read sequencing for discovering these variants in archaic hominins. By consequence, the role of SV during hominin divergence remains poorly characterized.
We leveraged recent long-read-based catalogs to discover SV alleles that persist in modern humans as a result of ancient admixture with Neanderthals. We identified 270 candidate introgressed SV alleles that are absent or nearly absent among African populations, polymorphic among non-African populations, flanked by introgressed single nucleotide polymorphisms, and in strong linkage disequilibrium with introgressed haplotypes. These include 143 deletions and 127 insertions, spanning totals of 113 and 96 kb, respectively. We further applied k-mer alignment to variant graphs to genotype Neanderthal samples at these SV loci. A subset of these SVs overlap protein-coding genes or regulatory elements. These include a 4.5 kb deletion spanning an intron-exon boundary of the transcriptional regulator *ZNF454* and a 349 bp insertion in a conserved portion of the 3' UTR of the growth hormone receptor (*GHR*). Many introgressed insertions represent transposable elements, with LINEs and Alus comprising 77% of the inserted sequence. Together, our results highlight SV as a potent substrate for hominin evolutionary divergence.

Novel insights into evolutionary genetics from emerging technologies

SMBE-PO-406

Decode the timing of plant germline segregation using a high-resolution DNA barcoding system

F. Xia [1,*], K. Liu [1], X. He [1]

[1]School of life sciences, Sun Yat-sen University, Guangzhou, China

Abstract: Whether plants have a segregated germline is a fundamental and unsolved question of plant development and evolution. Not like animals, whose gametogenesis occurred early in development so as to somatic mutations cannot be inherited, most plants were believed to lack a developmentally defined germline. However, such prevailing wisdom was supported by little direct evidence and disabled to explain the escape of mutational meltdown and longevity of plant. Efforts to spy this process never stopped but limited tracing resolution kept the truth still veiled. Here, we introduced a powerful DNA barcoding system into model plant Arabidopsis to record cell lineage of the whole plant. By reconstruing a high-resolution genealogy tree between germline and somatic cells, we found that Arabidopsis germ cells shared the same lineage with its physical related somatic cells. Further, the branch lengths of cell genealogy tree were measured to demonstrate the timing of germline segregation. Our data provided a direct evidence that Arabidopsis does lack a segregated germline during early developmental stage. This finding suggests that somatic-to-germline switch has independently evolved in plants and animals.

Novel insights into evolutionary genetics from emerging technologies

SMBE-PO-402

Chromosome-scale genome assembly of the mangrove horseshoe crab provides new insights into its genome duplication history

P. Shingate [1,*], V. Ravi [1], B.-H. Tay [1], A. Prasad [1], L.-M. Yap [2], T. Qian [3], H. Y. Tong [4], K. Tun [4], F. Rheindt [3], B. Venkatesh [1]

[1]Comparative and medical genomics laboratory, Institute Of Molecular And Cell Biology, A*STAR, [2]School of Applied Sciences, Republic Polytechnic, [3]Department of Biological Sciences, National University of Singapore, [4]National Parks Board, Singapore, Singapore

Abstract: Horseshoe crabs are marine arthropods belonging to the subphylum Chelicerata and order Xiphosura. They are phylogenetically closer to spiders and scorpions than to crabs which are crustaceans. Horseshoe crabs have existed for ~450 million years but exhibit little morphological change throughout evolution. Additionally, they show a low level of species diversity with only four extant species. Yet, previous studies have suggested that horseshoe crabs have experienced two rounds of whole-genome duplication, independent of the spider + scorpion clade. However, these studies were based on highly fragmented genome assemblies derived from short reads. A recent high contiguity assembly of the Chinese horseshoe crab (*Tachypleus tridentatus*) based on Oxford Nanopore reads (Gong et al. 2019 *Mol Ecol Resour*) identified only two Hox clusters in the genome suggesting only one round of whole-genome duplication. To resolve this issue conclusively, we have generated a high-quality chromosome-scale genome assembly for the mangrove horseshoe crab (*Carcinoscorpius rotundicauda*) using single molecule real-time PacBio reads and Hi-C data. The genome assembly of the mangrove horseshoe crab spans 1.72 Gb with contig N50 value of 10.5 Mb and scaffold N50 value of 93 Mb. Analysis of this highly contiguous genome assembly and comparison with spider and scorpion genomes have provided new insights into the whole-genome duplication history of horseshoe crabs.

Novel insights into evolutionary genetics from emerging technologies

SMBE-PO-404

Recent Samoan Population History Suggests Dynamic Population Size Changes and Migrations

D. N. Harris [1 2 3,*], M. D. Kessler [1 2 3], A. C. Shetty [1 2 3], D. E. Weeks [4 5], R. L. Minster [4], S. Browning [6], E. E. Cochrane [7], R. Deka [8], N. L. Hawley [9], M. S. Reupena [10], T. Naseri [11], T.-O. For Precision Medicine (topmed) Consortium, T. Population Genetics Working Group, S. T. McGarvey [12 13], T. D. O'Connor [1 2 3]

[1]Institute for Genome Sciences, [2]Department of Medicine, [3]Program in Personalized and Genomic Medicine, University of Maryland School of Medicine, Baltimore, [4]Department of Human Genetics, Graduate School of Public Health, [5]Department of Biostatistics, Graduate School of Public Health, University of Pittsburgh, Pittsburgh, [6]Department of Biostatistics, University of Washington, Seattle, United States, [7]Anthropology, The University of Auckland, Auckland, New Zealand, [8]Department of Environmental Health, University of Cincinnati College of Medicine, Cincinnati, [9]Department of Epidemiology (Chronic Disease), Yale University School of Public Health, New Haven, United States, [10]Lutia i Puava ae Mapu i Fagalele, [11]Ministry of Health, Apia, Samoa, [12]International Health Institute, Department of Epidemiology, Brown University School of Public Health, [13]Department of Anthropology, Brown University, Providence, United States

Abstract: The Samoan nation in Remote Oceania consists of an archipelago, which archaeological evidence suggests was peopled approximately 3,000 years ago. There are numerous questions regarding Samoa's population history, such as the early population size and a potential recent population replacement. Therefore, we analyzed 1,197 high coverage genomes from Savai'i and Upolu (the main Samoan islands), through the Trans-Omics for Precision Medicine program, to model Samoa's population history. We find that Samoans have admixture from Papuan speaking, African, East Asian, and European sources. We also identify a Denisovan admixture signal in many Samoan individuals, which is tightly correlated to their Papuan ancestry. This suggests that Papuan admixture introduced Denisovan haplotypes to Samoans. Through principal component analysis, we identify small, but significant, separation of individuals from the two islands, even though they are geographically proximate. Rare variant sharing demonstrates clear population structure, where individuals from rural regions share more rare variants with individuals from the urban regions than with individuals from the other rural regions. This is consistent with Samoa's recent urbanization and supports the hypothesis that urbanization has a greater influence on Samoan population structure than biogeography. We also use Identity-by-Descent sharing and the IBDNe methodology to reconstruct the effective population size history within the last 100 generations and find low effective population size from 100 to 35 generations ago, which is consistent with the archaeological hypothesis of initial low population size. There is an increase in effective population size from 35 to 10 generations ago. This population expansion parallels archaeological evidence for increases in settlements around this time, however is also consistent with a proposed population replacement 2,000 to 1,500 years ago. The population size then crashes at 10 generations ago, likely due to European contact. Overall, we provide the most detailed model of human demography in an Oceanic group.

Novel insights into evolutionary genetics from emerging technologies

SMBE-PO-410

Incipient asexual reproduction in Caenorhabditis nematodes

P. Lamelza [1 2], J. M. Young [3,*], L. M. Noble [4 5], A. Isakharov [1], M. Palanisamy [1], M. V. Rockman [4], H. S. Malik [3 6], M. Ailion [1]
[1]Department of Biochemistry, [2]Molecular and Cellular Biology Graduate Program, University of Washington, [3]Division of Basic Sciences, Fred Hutchinson Cancer Research Center, Seattle, [4]Department of Biology and Center for Genomics and Systems Biology, New York University, New York, United States, [5]Institut de Biologie, Ecole Normale Superieure, CNRS, INSERM, Paris, France, [6]Howard Hughes Medical Institute, Fred Hutchinson Cancer Research Center, Seattle, United States

Abstract: Most animal species reproduce by sex. Theory predicts there are advantages to being able to switch reproduction between sexual and asexual modes, but facultative sex is rare in animals, implying that there are strong selective pressures that prevent asexuality arising from an obligately sexual ancestor. A better understanding of these selective pressures requires studying the transitional forms between sexual and asexual modes of reproduction. One of the key steps in the evolution of asexuality from a sexual ancestor is the transition from haploid to diploid maternal inheritance. Here we report that interspecific hybridization between two sexual *Caenorhabditis* nematode species (*C. nouraguensis* females and *C. becei* males) results in two types of viable offspring. The first class consists of fertile asexual (i.e. gynogenetic) offspring, which inherit a diploid maternal genome and fail to inherit a paternal genome. The second class consists of sterile hybrid offspring, which inherit both a diploid maternal genome and a haploid paternal genome. Using whole-genome sequencing of individual viable worms, we show that diploid maternal inheritance in both asexual and hybrid offspring results from the inheritance of two random homologous chromatids from *C. nouraguensis* oocytes. This genetic mechanism of diploid maternal inheritance is indistinguishable from that of many obligately asexual species. Furthermore, we show that intraspecies *C. nouraguensis* crosses can also result in a low frequency of diploid maternal inheritance and asexual reproduction. Thus, *C. nouraguensis* provides a powerful, genetically tractable model to study the evolutionary origins of asexuality from obligately sexual species.

Novel insights into evolutionary genetics from emerging technologies

SMBE-PO-399

De novo assembly of a Tibetan genome and identification of novel structural variants associated with high altitude adaptation

Y. He [1,*], O. NA [2], H. Lou [3], C. Cui [2], L. Deng [3], S. Xu [3], X. Qi [1], B. Su [1]

[1]Kunming Institute of Zoology, Chinese Academy of Sciences, Kunming, [2]High Altitude Medical Research Center, School of Medicine, Tibetan University, Lhasa, [3]Key Laboratory of Computational Biology, Shanghai Institutes for Biological Sciences, Chinese Academy of Sciences, Shanghai, China

Abstract: Structural variants (SVs) may play important roles in human adaption to extreme environments such as high altitude but have been under-investigated. Here, combining long-read sequencing with multiple scaffolding techniques, we assembled a high-quality Tibetan genome (ZF1), with a contig N50 length of 24.57 mega-base pairs (Mb) and a scaffold N50 length of 58.80 Mb. The ZF1 assembly filled 80 remaining N-gaps (0.25 Mb in total length) in the reference human genome (GRCh38). Markedly, we detected 17,900 SVs, among which the ZF1-specific SVs are enriched in GTPase activity that is required for activation of the hypoxic pathway. Further population analysis uncovered a 163-bp intronic deletion in the *MKL1* gene showing large divergence between highland Tibetans and lowland Han Chinese. This deletion is significantly associated with lower systolic pulmonary arterial pressure, one of the key adaptive physiological traits in Tibetans. Moreover, with the use of the high quality *de novo* assembly, we observed a much higher rate of genome-wide archaic hominid (Altai Neanderthal and Denisovan) shared non-reference sequences in ZF1 (1.32%>1.53%) compared to other East Asian genomes (0.70%>0.98%), reflecting a unique genomic composition of Tibetans. One such archaic-hominid shared sequence, a 662-bp intronic insertion in the *SCUBE2* gene, is enriched and associated with better lung function (the FEV1/FVC ratio) in Tibetans. Collectively, we generated the first high-resolution Tibetan reference genome, and the identified SVs may serve as valuable resources for future evolutionary and medical studies.

Novel insights into evolutionary genetics from emerging technologies

SMBE-PO-400

Ultra-deep sequencing reveals the full spectrum and frequency of structural variations in populations of three large double stranded DNA viruses

V. Loiseau [1,*], E. Herniou [2], Y. Moreau [2], N. Lévêque [3], R. Cordaux [4], C. Gilbert [1]

[1]laboratoire EGCE, CNRS, Gif-sur-Yvette, [2]University of Tours, CNRS, Tours, [3]Laboratoire de Virologie, CHU de Poitiers, [4]Laboratoire EBI, CNRS, Poitiers, France

Abstract: Structural variations (SVs), including deletions, insertions, inversions, duplications and translocations, are increasingly recognized as an important source of genomic variation underlying adaptation in prokaryotes and eukaryotes. In viruses, comprehensive studies of SVs are still scarce. Here, we characterize the full spectrum and frequency of SVs in populations of a herpesvirus infecting human (HCMV), a baculovirus used in biocontrol (AcMNPV) and an iridovirus used as model in invertebrate antiviral immunity (IIV6). We developed an approach combining four existing SV detection pipelines and applied it to ultra-deep Illumina sequencing datasets of HCMV (20,000X), AcMNPV (196,000X) and IIV6 (123,000X). We found a total of 1,266, 2,410 and 799 different SVs in HCMV, AcMNPV and IIV6 respectively, with a majority (32 to 59%) of deletions or inversions in the three viruses. We also PacBio-sequenced the AcMNPV genome at 203,000X depth and ran two long-read SV callers on these data, which allowed us to recover a large number of SVs found in the Illumina dataset and new ones, including 87 insertions ranging from 99 to 18,987 bp. While for the three viruses, the vast majority (>88%) of SVs have a population frequency below 0.1%, the cumulative SV frequency reaches no less than 84%, 53% and 46% in HCMV, AcMNPV and IIV6 populations, respectively. Finally, we analyzed 20 published experimental AcMNPV populations extracted from two highly permissive moth species. We found an idiosyncratic pattern of SV gain, loss and frequency variation among populations, suggesting genetic drift as the main force governing SVs evolution in these populations. Overall this study establishes SVs as a major facet of large dsDNA virus biology and provides a solid foundation to further assess the role of these SVs in viral adaptation to less permissive hosts.

Novel insights into evolutionary genetics from emerging technologies

SMBE-PO-422

Whole-genome PhyloCSF reveals 144 high-confidence human protein-coding genes

T. Hunt [1,*], J. Mudge [1], I. Jungreis [2 3], J. M. Gonzalez [1], J. Wright [4], A. Frankish [1], M. Kellis [2 3], P. Flicek [1]

[1]European Bioinformatics Institute, European Molecular Biology Laboratory, Cambridge, United Kingdom, [2]Computer Science and Artificial Intelligence Laboratory, MIT, [3]Broad Institute, Cambridge, United States, [4]Functional Proteomics, Division of Cancer Biology, Institute of Cancer Research, London, United Kingdom

Abstract: The most widely appreciated role of DNA is to encode protein, yet despite nearly twenty years of intense study the exact portion of the human genome that is translated remains to be ascertained. Previously, we developed PhyloCSF to identify the evolutionary signature of protein-coding regions using multi-species genome alignments, and this tool has been widely used to study known transcripts in a variety of species. Here, we present and analyse the first whole-genome human PhyloCSF dataset, and provide corresponding datasets for the fly, worm, and mosquito genomes. We devise a new workflow that allows manual annotators guided by machine-learning algorithms to efficiently discover novel conserved protein-coding regions. As part of the GENCODE gene annotation project, we analyse over 1000 high-scoring human PhyloCSF regions, and confidently add 144 conserved protein-coding genes and 169 pseudogenes to the GENCODE geneset, as well as additional coding sequences within 236 previously-annotated protein-coding genes. The majority of these annotations represent new discoveries, and most identifications were dependent on the extension of the GENCODE transcript catalog through the integration of modern transcriptomics datasets. We show that simultaneous comparative annotation of other vertebrate genomes is essential to remove spurious ORFs and to distinguish coding from pseudogenic sequences. We find support for the translation of several of the novel coding genes in mass spectrometry datasets and use single nucleotide variants to show that purifying selection in the novel coding regions has continued in the human population. Altogether, our PhyloCSF datasets and algorithms will be important tools for researchers seeking to interpret these genomes, while our novel annotations present exciting loci for further experimental characterisation.

Novel insights into evolutionary genetics from emerging technologies

SMBE-PO-424

Fine-scale genomic investigation using long-read sequencing reveals long terminal repeat retrotransposon activity in the marine diatom Phaeodactylum tricornutum

G. V. Filloramo [1,*], B. A. Curtis [1], J. M. Archibald [1]

[1]Dalhousie University, Halifax, Canada

Abstract: The marine pennate diatom *Phaeodactylum tricornutum* is a popular model system for exploring genetic factors underlying the remarkable diversity and success of diatoms in the world's oceans. Previous genomic studies have indicated that transposable elements (TEs) are abundant in *P. tricornutum,* comprising ~6% of the 27.4 Mb genome. The majority of TEs present in *P. tricornutum* are long terminal repeat retrotransposons (LTR-RTs), which are characterized by long terminal repeat sequences flanking the central coding domain and replication via reverse transcription of an mRNA intermediate. Using Oxford Nanopore long-read sequencing, we have discovered that *copia*-type LTR-RTs are unusually active in the *P. tricornutum* genome. In addition to confirming the 42 previously reported *copia*-type LTR-RT loci, we identified 34 previously overlooked LTR-RTs in the existing *P. tricornutum* reference genome. Furthermore, we identified over 300 novel *copia*-type LTR-RT insertions in our *de novo* long-read assembly, indicating rapid expansion of this superfamily in the *P. tricornutum* genome. Consistent with these findings, we have detected *copia*-type LTR-RTs that are both chromosomally integrated as well as non-integrated, suggesting that at least some of the elements are undergoing reverse transcription. The potential impact of LTR-RTs on the structure, regulation and diversity of diatom genomes will be discussed.

Novel insights into evolutionary genetics from emerging technologies

SMBE-PO-423

De novo eukaryotic genome assembly from subnanogram input samples with nanopore sequencing

C. Laumer [1 2,*], J. Marioni [3 4], E. Birney [5]

[1]Marioni Research Group, EMBL / EBI, [2]Marioni Research Group, Wellcome Sanger Institute, [3]Marioni Research Group, EMBL-EBI, Hinxton, [4]Marioni Research Group, CRUK Cambridge Institute, Cambridge, [5]Birney Research Group, EMBL-EBI, Hinxton, United Kingdom

Abstract: Despite recent advances, nanopore sequencing still requires large amounts of input DNA, and base accuracy remains limited both at the single read and consensus sequence level. For many species, this necessitates either amplification or pooling prior to sequencing, with the latter strategy potentially increasing library heterozygosity and complicating de novo assembly. Here, we compare amplification strategies compatible with ONT-driven genome assembly. We demonstrate that via optimized library preparation and long-range PCR amplification, we can completely cover the genome with a Poisson-like distribution of 1-D reads with mean size between 2-8 kbp (depending on the required number of cycles). More exciting still, we have developed strategies for efficiently circularizing libraries of multikilobase genomic inserts, enabling rolling circle library amplification. In our hands this process can convert subnanogram inputs into tens of micrograms of DNA with mean size in excess of 50 kbp, representing concatemeric direct repeats of the insert. This gives millions of error-correctable reads per MinION flow cell, enabling whole-genome long read coverage with single read accuracy well in excess of Q20. We comment on different sources of coverage bias, and discuss the mitigation thereof. To demonstrate the potential of these methods in enabling genome assembly across the Tree of Life, we show contiguous assemblies from single adult worms of *C. elegans* as well as its congener *C. brenneri*, one of the most heterozygous eukaryotic genomes described to date.

Novel insights into evolutionary genetics from emerging technologies
SMBE-PO-421
A-to-I RNA editing shapes the distribution of damaging variants
T.-L. Mai [1,*], T.-J. Chuang [1]
[1]Genomics Research Center, Academia Sinica, Taipei, Taiwan

Abstract: Mutation, providing the major source of genetic variability within a population, is one of the fundamental driving forces of evolution. Generally, most deleterious nonsynonymous substitutions at functionally important loci would be eliminated by natural selection. However, recent studies have observed that each healthy individual could carry hundreds of missense mutations predicted to be severely deleterious or disease causing. The reason that strongly detrimental variants are preserved in a population and not eliminated by negative selection remains unclear. We therefore hypothesized that adenosine-to-inosine (A-to-I) RNA editing, a very common co-/post-transcriptional modification mechanism in metazoans, associates with the burden of deleterious A/G polymorphisms in a population. To test this assumption, we conduct a population-based analysis by integrating genomic and transcriptomic sequencing data. We show that nonsynonymous editing activities (prevalence/level) are negatively correlated with the deleteriousness of A-to-G genomic changes and positively correlated with that of G-to-A genomic changes within human population. Our results also reveal a negative correlation between nonsynonymous editing activities and allele frequency of A within the population, and this correlation becomes stronger when editing sites are located in evolutionarily/functionally important loci. In addition, we found that the proportion of rare missense G-to-A mutations and of the binding motif of editing enzymes at the variants increased with the increasing deleteriousness. Taken together, our results suggest a new insight of nonsynonymous A-to-I RNA editing that is associated with the increased burden of G-to-A missense substitutions in healthy individuals, thereby expanding its potential in pathogenomics studies.

Novel insights into evolutionary genetics from emerging technologies
SMBE-PO-409A
Comparative genomics analysis revealed novel insight into different characteristics among coral lineages
H. Ying*, I. Cooke, G. Huttley, D. Hayward, W. Wang, E. Ball, S. Forêt, D. Miller

Abstract: As the building block of coral reefs, stony corals have undergone extensive diversification and have thrived over the last 200 million years. There are nearly 1000 coral species exhibiting a variety of morphological and physiological attributes, but the genomic basis is largely unknown. This situation is staring to change with many coral genomic projects undertaken over the last ten years. Conventionally, most extant corals fall into two deeply diverged clades known as the Complexa (complex corals) and Robusta (robust corals). Here, we describe five newly sequenced corals genomes including three complex corals and two robust corals. A well-structured HOX-related gene cluster containing seven homeobox genes was found to be highly conserved in all coral genomes. Conversely, the expansion of HSP20 gene family appeared to be species-specific and has an apparent correlation with stress tolerance. The most unexpected finding from the comparative work was that unique among animals, the complete gene set for histidine biosynthesis pathway is retained in robust corals, which has important implications for coral nutrition and symbiosis. The second comparative study was performed on two closely related coral species, *Acropora millepora* and *Acropora digitifera. Acropora* is of particular significance because it is the dominant genus of reef-building corals in the Indo-Pacific and was believed to undergo a series of rapid radiations. The two genomes shared 95% nucleotide similarity and 98% protein coding sequence similarity. Approximately 1,500 protein-coding genes were identified as fast-evolving genes, among which immunity and miRNA metabolism related genes were highly over-represented. These observations provide insight on how different corals achieved adaptation to their environment and build better understanding why corals were successful in the past and how we can protect them in the future.

Novel insights into evolutionary genetics from emerging technologies

SMBE-PO-425

A K-mer Based Approach to Characterize the Dark Matter of the Arabidopsis thaliana Genome

C. J. Fiscus [1,*], D. Koenig [1]

[1]Botany and Plant Sciences, University of California, Riverside, Riverside, United States

Abstract: Copy number variation of repeats can change rapidly within a species and can drive variation in genome size. Both whole genome assemblies and alignment-based comparisons between genomes are hindered by highly repetitive sequences, resulting in the poor characterization of as much as 90% of some eukaryotic genomes. Here, we simultaneously compare genome content amongst hundreds of samples using K-mer abundance profiles generated from short sequencing reads. The model plant *A. thaliana* can vary up to 18 Mbps in genome size, and we partition this variation by characterizing the repetitive content of more than 1,000 sequenced genomes. We identify extensive variation in satellite repeat abundance as an important contributing factor to genome size variation. Further, we demonstrate that the genetic basis of satellite repeat abundance can be mapped to both *cis* and *trans* loci. This work presents a novel K-mer based approach to cataloging genome content variation from sequencing reads and contributes to the understanding of repeat dynamics in eukaryotic genomes.

Novel insights into evolutionary genetics from emerging technologies
SMBE-PO-412

A new high throughput whole genome approach to simultaneously assay genotype-phenotype associations in bacteria under multiple conditions

S. Bastkowski [1,*], A. Page [1], M. Yasir [1], K. Turner [1], M. Webber [1], I. Charles [1]
[1]Quadram Institute, Norwich, United Kingdom

Abstract: Bacteria have evolved over billions of years to survive in a wide range of environments. Currently, there is an incomplete understanding of the genetic basis for mechanisms underpinning survival in stressful conditions, such as the presence of anti-microbials. Transposon Directed Insertion-site Sequencing (TraDIS) has been proven to be a powerful tool to identify genes and networks, involved in survival and fitness. Here we present a modification of TraDIS, which allows conditional fitness of all genes in the genome to be assessed simultaneously, including essential genes, which cannot be effectively assayed in traditional TraDIS experiments. Our refined methodology uses a transposon with an outward directed inducible promoter. This allows the investigation of the impact of over- and under-expression of each gene in addition to inactivation. In order to extract meaningful signals from this wealth of data, we developed AlbaTraDIS, a software application for rapid large-scale comparative analysis of TraDIS experiments that predicts the impact of inserts on nearby genes. We demonstrate the utility of our new approach by applying it to investigate mechanisms used by *Escherichia coli* grown in different concentrations of the biocide triclosan. As a result, we identified all well-characterised triclosan resistance genes, including the primary target, *fabI*. A number of new loci were also implicated in triclosan resistance and the predicted phenotypes for a selection of these were validated experimentally and results showed high consistency with predictions. Our new approach combined with the rapidly falling costs of sequencing allows for comparisons between experiments to identify commonalities in stress responses to many different conditions.

Novel insights into evolutionary genetics from emerging technologies

SMBE-PO-411

A Robust User Friendly Galaxy Pipeline for Advanced Bayesian Models of Allelic Imbalance

B. Miller [1,*], G. Gamble [1], J. Borgert [1], A. Morse [1], O. Moskalenko [1], F. Marroni [2], L. McIntyre [1]

[1]University of Florida, Gainesville, United States, [2]University of Udine, Udine, Italy

Abstract: The amount of mRNA each allele "transcribes" from DNA can be measured when the sequence between the alleles diverges. Allelic imbalance (AI) occurs when genetic variation in *cis* between the two alleles affects the regulation of mRNA expression and results in differing amounts of mRNA. The first statistical tool to identify AI differences between two "environmental" conditions was published by Novelo et. al (2017). It is based on a novel Bayesian statistical model, which formally tests allelic imbalance between two conditions while accounting for genomic context and bias in read mapping. However, there are numerous complex steps leading up to the deployment of this model. Individual scientists must execute each of these stages separately, a time-consuming process that made the model largely inaccessible. We have translated these individual steps into a robust process in Galaxy (Afgan 2018). Galaxy is an easy to use web interface that can be accessed as a part of NSF funded Cyverse or installed locally on a Linux server or workstation. The entire platform is open source and our workflow, with all segments written in Python/R, is wrapped in Galaxy. The workflow minimizes the amount of time, memory, and room for human error thereby increasing reproducibility. Extensive testing has proved this pipeline a robust and a viable option for analyzing thousands of fastq files in a single action to identify AI in a diploid genome given two conditions. The approach can also be applied to allo-polyploids in the analyses of homeologs (Boatwright et. al. 2018).

Novel insights into evolutionary genetics from emerging technologies
SMBE-PO-416
Post-duplicational dosage maintenance at translation level in yeasts and worms
A. Y.-F. Chang [1,*], B.-Y. Liao [1]
[1]National Health Research Institutes, Taiwan, Zhunan, Taiwan

Abstract: In eukaryotes, expression reduction of duplicated genes for dosage maintenance has been linked to transcriptional control involving epigenetic mechanisms. In light of former studies indicating that gene expression control is predominantly determined at the translation level, we anticipated that post-duplicational dosage maintenance also involves regulation of gene expression at the protein level. Combining transcriptomes and proteomes of unicellular or multicellular eukaryotic models, which are yeasts (*Saccharomyces cerevisiae* vs. *Schizosaccharomyces pombe*) or worms (*Caenorhabditis elegans* vs. *Caenorhabditis briggsae*) respectively, we tested our hypothesis by investigating regulatory properties of lineage-specific duplicated genes. Dosage sensitive genes (DSGs) were found to reduce their rate of protein production more significantly than dosage insensitive genes (DIGs) after duplication events. To determine if such control at protein level is achieved through mechanisms related to the efficacy of protein synthesis or protein degradation, we analyzed ribosome profiling (RiboSeq) data of yeast and worm genes. In both eukaryotic model systems that were investigated, a more prominently reduced translational efficiency was observed in DSGs than DIGs. Furthermore, we observed that the translationally repressed copy of duplicated genes was more likely to be the copy that was transcriptionally inhibited. These results together suggested that dosage control does involve translation, and this translational control is to enhance transcriptional control for maintaining gene dosage during eukaryotic genome evolution.

Novel insights into evolutionary genetics from emerging technologies

SMBE-PO-415

The divergence of maize NAM founders revealed by syntenic LTR retrotransposons

S. Ou [1,*], C. N. Hirsch [2], K. R. Dawe [3], M. B. Hufford [1]

[1]Department of Ecology, Evolution, and Organismal Biology, Iowa State University, Ames, [2]Department of Agronomy and Plant Genetics, University of Minnesota, Saint Paul, [3]Department of Genetics, University of Georgia, Athens, United States

Abstract: Maize (*Zea mays* L.) exhibits extensive molecular variation that is associated with its diverse phenotypes. To understand these associations and facilitate predictive crop breeding, a total of 25 diverse inbred lines were crossed with the B73 inbred line to develop a family of recombinant inbred lines, known as the maize Nested Association Mapping (NAM) population. Due to advances in sequencing and assembly techniques, we were able to generate chromosomal-scale genomes of all 26 NAM founder lines. These high-quality genomes provide an opportunity for complete analysis of repetitive sequences, an impossibility with resequencing data. In this study, we utilized long terminal-repeat retrotransposons (LTR-RTs) found in syntenic genomic regions to study the divergence of the 26 NAM founder genomes. LTR-RTs are interspersed repetitive elements dominating the maize genome. The frequent insertional activity of LTR-RTs creates thousands of genetic footprints that record the genomic divergence of maize. Further, random mutations accumulated in the direct repeat of LTR-RTs serve as ideal molecular clocks to estimate the timing of these events. We reveal that syntenic LTR-RTs can capture genetic variation missed in SNP-based methods, which could be used to pinpoint genome divergence between closely related lines, such as the maize NAM founders. We thus propose a model-free method for unbiased estimation of genomic distance based on syntenic LTR-RTs. Together with the reciprocal hybrids of NAM founders as well as related wild species, we test the association between a wide range of genetic distance and heterosis, providing a potential avenue toward more accurate genomic prediction.

Novel insights into evolutionary genetics from emerging technologies

SMBE-PO-414

Tractor: A framework for well-calibrated genomic analysis of underrepresented admixed individuals.

E. G. Atkinson [1 2 3,*], A. X. Maihofer [4], C. M. Nievergelt [4], K. C. Koenen [2 3], B. M. Neale [1 2], M. J. Daly [1 2 5]

[1]Analytical and Translational Genetics Unit, Massachusetts General Hospital, Boston, MA, [2]Stanley Center for Psychiatric Research, Broad Institute of MIT and Harvard, Cambridge, MA, [3]Epidemiology, Harvard TH Chan School of Public Health, Boston, MA, [4]Psychiatry, University of California - San Diego, La Jolla, CA, United States, [5]Institute for Molecular Medicine Finland, Helsinki, Finland

Abstract: Admixed individuals, whose ancestry comprises components from several ancestral groups, are routinely removed from medical genetic studies due to the challenges of accurately accounting for their complex ancestry and the resulting concern that population structure may infiltrate analyses and bias results. Here, we present a framework and software package, named *Tractor*, to account for this issue and allow admixed individuals to be studied alongside homogenous ones by inferring fine-scale population structure as informed by local ancestry. Incorporating local ancestry information in addition to principal components takes into account subtle differences in admixture patterns that may differ among cohorts even if their genome-wide ancestry fractions are the same. We test our framework in simulation with data from African American individuals of the Psychiatric Genomics Consortium PTSD working group. We demonstrate the notable extent to which *Tractor* boosts signal to discover GWAS loci and characterize the landscape of when using our method gives the most significant power gains. We further demonstrate that this framework gives increased fine-mapping precision by leveraging differences in linkage disequilibrium patterns between ancestral populations to more narrowly map where signal is coming from. This framework could be applied to solve the statistical issues related to admixture across many medical and population genetics activities, such as evolutionary studies using genome-wide selection scans or in the construction of polygenic risk scores. In sum, *Tractor* dramatically advances the existing methodologies for studying admixed individuals and allows for significantly better calibrated study of the genetics of complex disorders in underrepresented populations.

Novel insights into evolutionary genetics from emerging technologies

SMBE-PO-407

A recipe for high-quality primate genome assemblies

M. Escalona [1,*], M. Haukness [1], A. Regier [2], C. Tomlinson [2], M. Kremitzki [2], D. Albracht [2], W. Warren [3], T. Graves-Lindsay [2], I. Hall [2], R. E. Green [1]

[1]Department of Biomolecular Engineering, University of California, Santa Cruz, Santa Cruz, California, [2]McDonnell Genome Institute, Washington University School of Medicine, St. Louis, Missouri, [3] Christopher S. Bond Life Sciences Center, University of Missouri, Missouri, United States

Abstract: New technologies including long-read sequencing and proximity-ligation are being used to generate genome assemblies for many organisms. While there are many tools for exploiting the specific features of these data for genome assembly, there is no currently integrated approach that takes advantage of the complementary features of each data type. In an effort to upgrade the reference genomes for several primates and to provide new de novo genomes for several individual humans, we have designed an iterative pipeline for genome assembly. This pipeline includes several steps including de novo scaffold generation of PacBio data using the Falcon assembler, correction of mis-assemblies using BioNano optical map data, scaffolding using proximity-ligation data, and gap-filling using short and long-read data using a custom approach. Our strategy then iterates the scaffolding and gap-filling steps until convergence. We will present the analysis of several human and non-human primate genome assemblies which achieve multi-megabase contig N50s and chromosome-scale scaffold N50s.

Open Symposium

SMBE-PO-439

Genome-wide analysis on the genetic basis of human head morphology in Japanese

K. Yamaguchi [1,*], K. Koganebuchi [2], T. Sato [3], T. Kannon [3], H. Tsujiguchi [3], K. Hosomichi [3], A. Tajima [3], K. Yamamoto [4], H. Ishida [2], H. Nakamura [3], R. Kimura [2]

[1]School of Natural Sciences and Psychology, Liverpool John Moores University, Liverpool, United Kingdom, [2]Graduate School of Medicine, University of the Ryukyus, Okinawa, [3]Graduate School of Medical Sciences, Kanazawa University, Kanazawa, [4]Kurume University School of Medicine, Kurume, Japan

Abstract: Human head morphology reflects brain development. Therefore, head circumference has been used as a proxy index for infantile brain development. The genetic background of pathological variation in head morphology has been studied well, however, the genetic basis of normal variation in head morphology remains unclear.

We conducted two genome-wide association studies (GWAS) for human head morphology in Japan as well as a meta-analysis on the results. The first GWAS used DNA microarray data of 767 Japanese individuals (429 males, 338 females) living in Okinawa Island of Japan. The second study consisted of 825 Japanese (377 males, 448 females) residing in Ishikawa prefecture of Japan. The meta-analysis found an association between a single nucleotide polymorphism (SNP) located within *CNTNAP2* on chromosome 7 and head breadth measured with a caliper at a conventional genome-wide significant level ($P<5.0x10^{-8}$) after controlling for sex, height, BMI, and genomic principal components (effect of population structure).

CNTNAP2 encodes a member of neurexin family and has been reported as a candidate gene for neurodevelopmental disorders including autism and ADHD. Since a brain size determines the skull size, it is convincing that a gene related to the nervous system also affects the head size and morphology. To elucidate the pathway in which *CNTNAP2* affects head breadth, further studies are required to examine each factor that determine head size, from brain structure to thickness of soft and hard tissue.

Open Symposium
SMBE-PO-514
Mapping natural selection through the D. melanogaster development
M. Coronado-Zamora [1,*], I. Salvador-Martínez [2], D. Castellano [3], A. Barbadilla [1], I. Salazar-Ciudad [4]
[1]Universitat Autònoma de Barcelona, Cerdanyola del Vallès, Spain, [2]University College London, London, United Kingdom, [3]Centre de Regulació Genòmica, Barcelona, [4]Centre de Recerca Matemàtica, Cerdanyola del Vallès, Spain

Abstract: The current genomic era has led to the paradoxical situation in which much more evidence of selection is available on the genome than on the phenotype of the organism, the primary target of natural selection. The first high-resolution map of natural selection showed that natural selection is rampant in the genome of *D. melanogaster* [1,2]. In contrast to Kimura's neutral theory expectations, from 30 to 50% of mutations that become incorporated into the genome of this species are adaptive [3]. Which is the adaptive significance of these mutations? The action of the selection in the whole phenotype of an organism has not yet been approached, neither any study integrating both levels of selection at a genomic scale. Here we carry out an organismal selection-phenotype-genotype integration; specifically we draw an exhaustive map of selection acting on the complete embryonic anatomy of species *D. melanogaster*. We have developed a new approach that integrates high-throughput data of genomic variation from the DGRP [1], gene expression and development [4] to map adaptation over the entire embryo's anatomy of *D. melanogaster*. The selection map [5] indicates that selective constraint is pervasive over most of the embryo's anatomy, specifically affecting anatomical organs related to the nervous and digestive systems. Adaptation is found in the structures that also show evidence of adaptation in the adult, the immune and reproductive systems. It is also found a relaxation of selection in the first stage, attributed to the maternal-effect genes. Finally, we observe that genes that are expressed in one or a few different anatomical structures are younger and have higher rates of evolution, unlike genes that are expressed in all or almost all structures. The integration of population genomics with other phenotypic multi-omics data is essential to obtain a global picture of how does adaptation occur [6].

[1] Mackay, et al. (2012) Nature. 482:173–178
[2] Langley, et al. (2012) Genetics. 192(2):533–598
[3] Eyre-Walker (2006) Trends Ecol Evol. 21:569–575
[4] Tomancak, et al. (2007) Genome Biol. 8(7):R145
[5] Salvador-Martínez, Coronado-Zamora, Castellano, et al. (2018) Mol Bio Evol. 35(1):66–79
[6] Casillas & Barbadilla (2017) Genetics. 205(3):1003–1035

Open Symposium
SMBE-PO-516
How Does Local Selection Affect Antibiotic Resistance Populations?
C. Cole [1,*], M. Barlow [1]
[1]Quantitative Systems Biology, University of California, Merced, Merced, United States

Abstract: The prevalence of infections caused by antibiotic resistant bacteria has been continuously increasing resulting in more than 2 million antibiotic resistant infections every year[1]. The evolution of antibiotic resistance is a selection driven process that is the direct result of antibiotic use. Local hospitals and doctors are responsible for antibiotic prescriptions and therefore the selection of resistance gene populations in a local community. Understanding how local antibiotic resistance populations relate to national populations will allow us to identify the local effect of selection of antibiotic resistance gene populations. Our lab seeks to identify resistance trends over time by comparing localized antibiotic resistance populations from a single hospital with national populations. Using multiplex PCR to identify the most common resistance genes, TEM-1, CTX-M-15, OXA-1, and SHV-2, which are found in our ESBL clinical isolates, we quantified resistance trends for the local hospital over a 3 year period (n=402). We found the following frequencies, TEM-1 (33%), CTX-M-15 (65%), OXA-1 (59%), SHV-2 (7.8%) in the local hospital population. Analysis using our database of clinical isolate genomes from across the US submitted to public databases (n=1948), we found the following frequencies TEM-1 (26%), CTX-M-15 (11%), OXA-1 (8.1%), and SHV-2 (27%). The frequencies of resistance genes from our local population are inconsistent with the frequencies found using our national database. Results from this study imply selection at a local level is stronger for CTX-M-15 and OXA-1 than the selection at the national level. CTX-M-15 is often associated with multidrug resistance and the CTX-M-15, OXA-1 combination results in resistance to ß-lactam ß-lactamase inhibitor combinations [2,3,4]. Conversely, there is stronger selection for SHV-2 at the national level. SHV-2 is associated with hyper-susceptibility to ß-lactam ß-lactamase inhibitor combinations [3]. The inconsistency between local and national resistance gene populations and their resistance mechanism reveals the impact of local selection on resistance gene populations.

Open Symposium

SMBE-PO-429

Speciation and secondary contact in Sulawesi macaque species

Y. Terai [1,*], S. Takuno [1], K. A. Widayati [2], L. H. P. S. Purba [2], X. Yan [3], H. Imai [3], B. Suryobroto [2]

[1]Department of Evolutionary Studies of Biosystems, SOKENDAI (The Graduate University for Advanced Studies), Hayama, Japan, [2]Faculty of Mathematics and Natural Sciences, Bogor Agricultural University, Bogor, Indonesia, [3]Primate Research Institute, Kyoto University, Kyoto, Japan

Abstract: The Sulawesi macaque species endemic to island of Sulawesi (Central Indonesia) have differentiated into seven morphologically distinct species in seven allopatric areas. The evolution of these species has been studied, however the evolutionary process and its genetic basis are still unclear. The remarkable point of these species is that species are hybridizing at the boundary of distribution (hybrid zones), while the hybrid zones have not extended. Therefore, the genetic differences between species are expected at loci responsible for local adaption, as a consequence, for preventing expansion of hybrid zones. Here, we report the process of speciation and genes related to adaptation in Sulawesi macaque species. We determined exome sequences from each of ten individuals of *Macaca tonkeana*, *M. hecki*, *M. nigra,* and *M. nigrescence* that distribute side-by-side. A phylogenetic tree based on exome sequences showed that Sulawesi macaques has split into seven species during a very short evolutionary period after migration to Sulawesi island. Single nucleotide polymorphisms (SNPs) were extracted from the exome sequences, and a handful of SNPs were fixed differences among species. These fixed differences were located in ~150 genes including genes responsible for olfaction, detoxification, hair formation, and reproduction in female. Especially, structural variants of a detoxification gene showed a mosaic distribution in Sulawesi macaques. These results demonstrate that most of genomic regions have not completely differentiated and the small number of genes with fixed differences may be responsible for local adaption and for preventing expansion of hybrid zones.

Open Symposium
SMBE-PO-489
Neanderthals were a genetic reservoir for functional alleles lost in the Eurasian migration out of Africa
D. C. Rinker [1,*], C. N. Simonti [2], E. McArthur [3], D. Shaw [3], E. Hodges [3 4], J. A. Capra [1 3]
[1]Biological Sciences, Vanderbilt University, Nashville, [2]Biological Sciences, Georgia Institute of Technology, Atlanta, [3]Vanderbilt Genetics Institute, [4]Biochemistry, Vanderbilt University, Nashville, United States

Abstract: Neanderthal ancestry persists in the genomes of all modern Eurasian populations. Neanderthal introgression introduced many new alleles into the anatomically modern human (AMH) ancestors of Eurasians, with the most distinct variants being those Neanderthal derived alleles (NDAs) that first appeared on the Neanderthal lineage. However, archaic introgression also[d1] introduced many alleles that had been maintained in Neanderthals since the common African ancestor of Neanderthals and AMHs. Some fraction of those African variants was certainly lost through the effects of genetic bottlenecking as AMHs expanded into Eurasia. Hence, most of the ancestral alleles that now occur exclusively on Neanderthal haplotypes in Eurasians had likely been lost by AMHs as they migrated out of Africa, only to be subsequently reintroduced through archaic admixture. In this study, we systematically identify and characterize these reintroduced alleles (RAs) in modern Eurasian populations.

RAs are present on at least 80% of introgressed haplotypes, with an average of one RA for every two NDAs genome-wide. RAs tend to occur in clusters, sometimes with many RAs in perfect linkage disequilibrium (LD) with a single NDA. This raises the possibility that, for introgressed loci associated with human traits (e.g., via GWAS or eQTL studies), the causal variant(s) may actually be RAs. Indeed, our results show that over 70% of NDA-associated GWAS traits are equally well associated with RAs. Using a range of functional genomic, computational, and MPRA data, we show that many RAs are likely to have gene regulatory functions. For one introgressed eQTL locus, we employ a luciferase reporter assay to demonstrate that RAs influence gene expression independently of the NDA in LD. Finally, we show that RA-dense eQTL loci are enriched in most GTEx brain tissues, suggesting that the regulatory effects of RAs may be pervasive.

Overall, we identify hundreds of thousands of RAs in Eurasian genomes, and provide evidence that some RAs have regulatory functions. This study illustrates how hybridization in the hominin lineage has reintroduced genetic variation across great spatiotemporal intervals, and that we must also consider RAs when evaluating the effects of Neanderthal introgression.

Open Symposium
SMBE-PO-444
Dosage-sensitive genes have restricted expression variation
A. M. Rice [1,*], P. Donnelly [1], A. McLysaght [1]
[1]Trinity College Dublin, Dublin, Ireland

Abstract: Dosage-sensitive genes are often seen to be refractory to variation. One such group, vertebrate ohnologs, paralogs retained from whole genome duplications at the base of the vertebrate lineage, are observed to be refractory to small-scale duplication, depleted on human benign copy number variants (CNVs) and enriched on pathogenic CNVs. This intolerance to copy number change is likely due to a CNV giving rise to a violation of an expression constraint that exists in one or more tissues. While CNVs alter encompassed genes' expression across tissues, expression quantitative trait loci (eQTLs), genomic regions harbouring sequence variants that influence the expression level of one or more genes, can act in a tissue-specific manner. Expression changes due to the presence of eQTLs in unconstrained tissues will not be deleterious and allow dosage-sensitive genes to vary expression while obeying constraints in unaffected tissues. Using eQTLs across 48 tissues from The Genotype-Tissue Expression (GTEx) project, we find that dosage-sensitive genes are enriched for being affected by eQTLs. However, the eQTLs affecting dosage-sensitive genes are biased towards having narrow tissue specificity with these genes having fewer eQTL-affected tissues than dosage-insensitive genes. Additionally, we find that dosage-sensitive genes are depleted for being affected by broad tissue breadth eQTLs, likely due to the increased chance of these eQTLs conflicting with expression constraints and being removed by purifying selection. These patterns suggest that dosage-sensitivity shapes the evolution of these genes by precluding copy number evolution and restricting their evolutionary trajectories to changes in expression regulation compatible with their constraints.

Open Symposium

SMBE-PO-445

Patterns of genome composition in double-stranded DNA viruses

D. Simón [1 2,*], M. Ines Fariello [3], H. Musto [1]

[1]Laboratorio de Genómica Evolutiva, [2]Laboratorio de Virología Molecular, Facultad de Ciencias. Universidad de la República, [3]Institut Pasteur de Montevideo, Montevideo, Uruguay

Abstract: The double-stranded DNA viruses (dsDNA), corresponding to Group I of the Baltimore classification, are replicated using a DNA-dependent DNA polymerase. A great proportion of dsDNA known are prokaryotic viruses, that derive from a diverse group of host bacteria or archaea. Some dsDNA that replicate in eukaryotic cells must enter the nucleus after infection to use the cellular machinery and replicate their DNA. Nucleo-Cytoplasmic Large DNA viruses (NCLDV) are largely independent of the host nucleus because they encode protein components of the replication complex machinery. They are a proposed taxonomic group but not yet accepted by the International Committee on Taxonomy of Viruses. The replication of some NCLDV is in the cytoplasm. DNA deamination, loss of amino groups of their nitrogenous bases, is a spontaneous process, although of low frequency. It can be suffered by all the bases that make up DNA, except thymine. Generally, these modifications are corrected by the DNA repair machinery, because these bases are not recognized by the replication machinery (i.e., Hypoxanthine, Xanthine, Uracil). Conversely, deamination of a methylated CpG generates a thymine, the constituent base of DNA; if it is not repaired before replication, a mutation occurs. Although deaminations of methylated CpG sites occur infrequently, they still represent a relatively important source of mutations. The aim of this communication is to explore the dsDNA diversity and try to describe major evolutionary patterns. Regardless of where DNA replication occurs, other aspects such as host and genome size are also crucial.

Open Symposium

SMBE-PO-440

Paleogenomic study of the pre-contact site of Cañada de la Virgen, Guanajuato, Mexico

M. A. Nieves-Colon [1 2,*], V. Villa-Islas [3], M. C. Ávila-Arcos [3], A. Moreno-Estrada [1], G. Z. Garcia-Moreno [4], K. Sandoval-Mendoza [1]

[1]National Laboratory of Genomics for Biodiversity (LANGEBIO-UGA), Centro de Investigación y de Estudios Avanzados del Instituto Politécnico Nacional, Irapuato, Mexico, [2]School of Human Evolution and Social Change, Arizona State University, Tempe, United States, [3]International Laboratory for Human Genome Research, National Autonomous University of Mexico, Querétaro, [4]National Institute of Anthropology and History, Guanajuato, Mexico

Abstract: Cañada de la Virgen is a pre-contact archaeological site located in northeast Guanajuato state, within Central Mexico. Archaeological study indicates the site was occupied between 540 and 1050 A.D. During this time, monumental structures built in the traditional Mesoamerican style were erected, including a pyramid and other ceremonial structures. However, which ethnic group, or groups, inhabited the site is still under debate. Thus, in this study we use paleogenomics to gain additional insight into the population history of the ancient Indigenous communities of Cañada de la Virgen and assess their potential origins within Mexico. We sampled 19 human skeletal remains excavated from several parts of the site, including two individuals who pre-date the construction of the monumental pyramid complex. Shotgun sequencing indicated six of the sampled individuals had over 1% endogenous DNA content. We recovered low-coverage nuclear DNA and complete mitochondrial DNA data for four of these individuals. Through principal components analyses, we compared the ancient nuclear genotypes to genome-wide data from present-day Native Mexicans. Three of the Cañada de la Virgen individuals cluster with populations from Central-West and East Mexico, but one clusters with present-day Maya groups. Additionally, we identified two mitochondrial haplogroups commonly found in Mesoamerica among the remains: A2ai and B2. These preliminary results lead us to propose that populations at Canada de la Virgen may have had multiple origins within Mexico. However, additional analyses of higher depth nuclear enriched data are needed to further test this hypothesis.

Open Symposium

SMBE-PO-435

Extracting complementary mitogenomic data from target enrichment experiments: a case study with 501 ant UCE libraries

R. Allio [1,*], J. Romiguier [1], B. Nabholz [1], F. Delsuc [1]

[1]Institut des sciences de l'évolution de Montpellier (ISEM), CNRS, EPHE, IRD, Université de Montpellier, Montpellier, France

Abstract: Next generation phylogenomics in which phylogenetic relationships among species are inferred from thousands of genomic markers is on the rise. Because of the combination of the advantages of NGS and target sequence amplification facilities (universal priors), one capture sequencing approach consisting of sequencing nuclear regions flanking and including ultraconserved DNA elements (UCEs) has become popular. Many recent studies have used phylogenomic reconstruction based on UCEs sequencing method. On the other side, mitochondrial genome (mtDNA) was and still is a marker of choice in phylogenetic inference and species identification. However, mtDNA is rarely assembled and used in conjunction with nuclear markers in capture-based studies. In addition, despite an increasing availability of UCEs data for numerous non-model species, mtDNA is still under-represented for numerous species.
Here, focusing on Formicidae for which 501 UCEs sequencing data are available whereas only 29 mtDNA are assembled and annotated, we have developed a pipeline called MitoFinder to extract, assemble, and annotate mtDNA from UCE libraries. To perfect our pipeline, we compared the efficiency of four different assemblers for both assemble UCEs and mtDNA. Using MitoFinder, we were able to reconstruct both mitochondrial and UCEs phylogenies for 501 species of ants and we provide a third phylogenomic inference using both nuclear and mitochondrial markers.
Overall, by assembling 501 mtDNA for Formicidae, our results support that mtDNA can be efficiently assembled from UCE data. Finally, we show that mitochondrial and nuclear markers contribute differently to phylogenomic reconstruction and using both of them could improve phylogenomic inference. In the light of the value of the mitochondrial signal, this study show that the extraction of this additional marker from UCE data should be more usual.

Open Symposium
SMBE-PO-437
Relaxed selection on the CFTR gene in muroid rodents
A. D. Stump [1,*]
[1]Biology, Adelphi University, Garden City, United States

Abstract: Cystic fibrosis (CF) is caused by mutations in the gene encoding CFTR, a transmembrane chloride channel. Laboratory mice have long been used as models for CF research, despite phenotypic differences in the effects of loss of CFTR function. Presented will be several lines of evidence that the CFTR gene has been evolving under significantly relaxed selection in a superfamily of rodents, the Muroidea, which includes the laboratory mouse. This evidence includes substantial amino acid sequence divergence between muroid CFTR orthologs and those from other mammals, d_N/d_S analyses consistent with relaxed selection, and reduced codon bias in muroid CFTR orthologs relative to other rodents. The full phylogenetic distribution of this relaxed selection is uncertain at this point, although it clearly does not extend to the rodent suborders Sciuromorpha and Caviomorpha (which includes the guinea pig Cavia porcellus). These results provide a critical perspective on the use of mice as models for CF research, and point to alternatives (including other rodents) that would be predicted to better match human CF pathology. More broadly, these results are a reminder of the importance of an evolutionary perspective in biomedical research.

Open Symposium
SMBE-PO-517
Accurate inference of tree topologies from multiple sequence alignments using deep learning
A. Suvorov [1,*], J. Hochuli [2], D. R. Schrider [1]
[1]Department of Genetics, [2]Biological & Biomedical Sciences Program, The University of North Carolina at Chapel Hill, Chapel Hill, United States

Abstract: Reconstructing the phylogenetic relationships between species is one of the most formidable tasks in evolutionary biology. Multiple methods exist to reconstruct phylogenetic trees, each with their own strengths and weaknesses. Both simulation and empirical studies have identified several "zones" of parameter space where accuracy of some methods can plummet, even for four-taxon trees. Further, some methods can have undesirable statistical properties such as statistical inconsistency and/or the tendency to be positively misleading (i.e. assert strong support for the incorrect tree topology). Recently, deep learning techniques have made inroads on a number of both new and longstanding problems in biological research. Here we designed a deep convolutional neural network (CNN) to infer quartet topologies from multiple sequence alignments. This CNN can readily be trained to make inferences using both gapped and ungapped data. We show that our approach is highly accurate, often outperforming traditional methods, and is remarkably robust to bias-inducing regions of parameter space such as the Felsenstein zone and the Farris zone. We also demonstrate that the confidence scores produced by our CNN can more accurately assess support for the chosen topology than bootstrap and posterior probability scores from traditional methods. While numerous practical challenges remain, these findings suggest that deep learning approaches such as ours have the potential to produce more accurate phylogenetic inferences.

Open Symposium
SMBE-PO-433
Dispensability and strict co-evolution of the MHC II pathway in teleosts.
A. Dubin [1,*], T. E. Jørgensen [1], T. Moum [1], S. D. Johansen [1], L. M. Jakt [1]
[1]Genomics group, Faculty of Biosciences and Aquaculture, Nord University, Bodø, Norway

Abstract: Although the gnathostome adaptive immune system increased in complexity with evolution, it is thought that the acquisition of all required cellular processes, tissues and genes happened quickly as most components are present across all jawed vertebrates, from the chondrichthyes to the bony aquatic and terrestrial vertebrates. However, whole genome studies in marine fish found that certain species lack important components of the immune system, revealing alternative immune strategies. It has been reported that members of the Syngnathus genus (pipefishes) and the entire Gadiformes order lack genes coding for the MHC II arm of the adaptive immune system.The MHC II pathway is responsible for defence against extracellular threats and immunisation through the presentation of exogenous peptides to T helper cells. We use two high-coverage draft genome assemblies of White-bellied anglerfish (*Lophius piscatorius*) to provide evidence for the absence of all genes encoding MHC II components (CD4, CD74 A/B, and MHC II α/β) in the genome, indicating loss of the classical MHC II pathway in this species. Our finding contrasts with prior identification of the functional MHC II pathway in the sister anglerfish species *Antennarius striatus*. This argues that loss of the complete MHC II pathway can occur within a relatively short evolutionary time scale and confirms the dispensability of this immune pathway in teleosts. We suggest that the MHC II pathway, including nonclassical molecules, acts as an independent evolutionary module whose components have no external functions, and that loss of any essential MHC II gene rapidly results in the loss of the remaining genes.

Open Symposium

SMBE-PO-434

Examining differentiated genotypes in the Tangier Island genetic study

S. Chavan [1,*], J. A. Shortt [1], T. Brunetti, N. Rafaels [1], M. Campbell [1], R. A. Mathias [2], C. R. Gignoux [1], K. C. Barnes [1], M. Daya [1]
[1]Department of Medicine, University of Colorado, Aurora, [2]Department of Medicine, Johns Hopkins University, Baltimore, United States

Abstract: Human population isolate communities may yield unique insights into variants and pathways involved in disease. From 1993 to 2001, we recruited ~60% of individuals living on Tangier Island, Virginia into a genetics study. The remote island can be reached by 45-minute boat ride or by small aircraft, but limited occupational opportunity has minimized immigration to the island in the 20th century. Previously, we used genealogical data to reconstruct the extended pedigree of 3,512 individuals to the settlement of this community back to its founding by British settlers in 1722. Analysis of pedigree data as well as estimated identical-by-descent sharing of genetic material have demonstrated limited gene flow into the population since its settlement. In this study, we aimed to characterize genotypes with large frequency differences between the Tangier Island population and its putative source population, England.

Samples were genotyped on the Illumina Multi-Ethnic Genotype Array (MEGA). After quality control, 399 samples were available for processing, including a maximum unrelated set of 85 individuals that was used for allele frequency estimation. Affymetrix 500k array data from the Population Reference Sample (POPRES) project was obtained to identify clinically relevant variants more common in Tangier Island compared to the United Kingdom (UK). The Tangier Island and POPRES array data were imputed on the Michigan Imputation Server to the Haplotype Reference Consortium panel, and a dataset containing SNPs present in the ClinVar database, and having imputation quality scores (Rsq) > 0.99, were used to identify clinically relevant variants. We characterized the frequencies of mitochondrial haplogroups, Y-chromosome haplogroups, and HLA genotypes, and compared the frequencies of clinically relevant variants between the Tangier Island population and POPRES.

The most common mitochondrial haplogroups observed in the Tangier Island call set are H (38%), U4 (17%), I (12%) and K (12%), with the U4 group being observed much more frequently in Tangier Island compared to England (2%). The most commonly observed Y haplogroups are R1b (75%), R1a (9%) and I (16%), which is different to the distribution of these haplogroups in England (frequencies of 62%, 5% and 25%, respectively). Multiple HLA alleles that are rare in England but common on Tangier Island were also identified: DBP1*02:02 (0.6% vs. 10.8%), DBP1*05:01 (1.8% vs. 8.5%), DRB1*01:02 (0.07% vs. 20.45%), DRB1*01:03 (3% vs. 10.8%), DRB1*04:02 (0.8% vs. 8.5%), B*14:01 (1.0% vs. 22.2%) and A*02:05 (1% vs. 20.5%). Of the 596 Clinvar variants more common in Tangier Island compared to the UK, 21 allele frequency differences were statistically significant (p-value < 0.05 after correcting for 596 1-sided statistical tests). However, none of these variants were pathogenic or likely pathogenic.

In conclusion, we identified several variants with large allele frequency differences between Tangier Island and England. These allele frequency differences are likely due to founder effects and the genetic isolation of the Tangier Island population.

Open Symposium

SMBE-PO-507

Genetic isolation of the Tangier Island population

M. Daya [1,*], T. Brunetti [1], C. Cox [1], C. Arehart [1], J. A. Shortt [1], S. Chavan [1], N. Rafaels [1], G. van Eeden [2], M. Campbell [1], R. A. Mathias [3], C. R. Gignoux [1], K. C. Barnes [1]

[1]Department of Medicine, University of Colorado, Aurora, United States, [2]Department of Biomedical Sciences, Stellenbosch University, Tygerberg, South Africa, [3]Department of Medicine, Johns Hopkins University, Baltimore, United States

Abstract: Tangier Island, Virginia, is a 3 square mile island located in Chesapeake Bay, 12 miles from the nearest mainland port. During 1993-2001, we recruited ~60% of islanders into a genetics study, and used genealogical data to create a pedigree of 3,512 individuals, connected to one large extended pedigree founded in 1722. In this study, we aimed to quantify the degree to which Tangier Island has been genetically isolated, and to characterize the genetic distance between Tangier Island and other modern-day European groups.

Samples were genotyped on Illumina's Multi-Ethnic Genotyping Array (MEGA). After quality control, 399 samples were available for analysis. We used phased genotypes to estimate runs of homozygosity (ROH) and segments of genetic material shared identical-by-descent (IBD), and used these estimates to infer historical patterns of effective population size. Using publicly available datasets consisting of 2,376 modern-day European populations (POPRES), we characterized the population structure of Tangier Island with principal components analysis (PCA), the fixation index (F_{st}), and IBD sharing.

A drastic reduction in estimated effective population size from ~8 million to ~1600 individuals occurred 13 generations before sample collection, which coincides with the period when the island was first settled. A further reduction in population size was estimated at 7-8 generations, consistent with a previous pedigree analysis of multi-generational genetic diversity in Tangier Island. Larger ROH are present in Tangier Island compared to POPRES (mean total ROH length of ~59MB vs. ~31MB). PCA of Tangier Island and POPRES revealed strong correlations between the total ROH length in Tangier Island and PC1 and PC2 (Spearman r=-0.62 and 0.63, respectively). The observed PCA clusters were consistent with estimates of F_{st} between groups, with Tangier Island falling closest to the United Kingdom (UK) and Swiss-French POPRES groups (F_{st}=0.065 for both distances), and furthest from Southwest Europeans (F_{st}=0.068). The average number of IBD segments shared between a Tangier Island subject and individuals from POPRES revealed that Tangier Island shares most genetic material IBD with the UK (mean per Tangier Island-UK pair = 5.19, mean per Tangier Island-Swiss-French pair = 4.94).

In summary, the genetic isolation of the Tangier Island population, evident from historical records and previous analysis of pedigree data, was confirmed by our genetic analyses. Likely due to its isolation, the population is genetically distinct from other European populations. Consistent with historical records, the Tangier Island population is genetically closest to the UK. In future, we plan to further refine our comparative population analysis by incorporating data from the People of the British Isles (PoBI) study.

Open Symposium
SMBE-PO-502
The genome rearranged: varied evolutionary trends shape genome architecture in rodents
T. D. Brekke [1,*], J. F. Mulley [1]
[1]School of Natural Sciences, Bangor Univeristy, Bangor, United Kingdom

Abstract: Rodents vary hugely in karyotype but it is unclear why they are so susceptible to chromosomal rearrangements compared to other mammals, nor why certain clades seem to be particularly predisposed to certain types of rearrangements. For instance, house mice harbor many Robertsonian translocations and classical cytological research suggests that Mongolian gerbils may as well. Hamsters have very few chromosomes in general, suggesting they experience more fusions than fissions, and voles have an extremely large X chromosome. We have used synteny and genetic linkage to predict the genome structure of Mongolian gerbils and Chinese hamsters, which in addition to the chromosome-scale genomes of house mice, rats, voles, and deermice, provide a detailed understanding of the patterns of rearrangements in two of the major rodent lineages: Muridae and Cricetidae. Using these data we show that gerbils have more rearrangements than most other rodents, and that these are not simply Robertsonian fusions as previously claimed. Indeed gerbils harbor far more rearrangements than any other rodent. We also show that while the X chromosome almost exclusively undergoes inversions, in gerbils and voles there are instances of an X-linked region translocating to an autosome. These genomes shed light on lineage-specific patterns of genome rearrangements in rodents and establish gerbils as a novel model for chromosomal evolution in the genomic era.

Open Symposium

SMBE-PO-511

Demographic processes in the territory of Eastern Baltic from the earliest inhabitants to modern times

K. Tambets [1,*], L. Saag [1 2], M. Laneman [3], L. Varul [4], A. Kriiska [3], M. Malve [3], H. Valk [3], A. Kushniarevich [1], C. L. Scheib [1], L. Saag [1], S. Rootsi [1], E. Metspalu [1], A. Solnik [1], T. Reisberg [1], E. D'Atanasio [5], E. R. Crema [6], M. Reidla [1], J. Parik [1 2], D. Díez-del-Molino [7 8], M. G. Thomas [9 10], V. Lang [3], R. Villems [1 2], T. Kivisild [1 5], M. Metspalu [1]

[1]Institute of Genomics, [2]Institute of Molecular and Cell Biology, [3]Institute of History and Archaeology, University of Tartu, Tartu, [4]School of Humanities, Tallinn University, Tallinn, Estonia, [5]Department of Human Genetics, KU Leuven, Leuven, Belgium, [6]Department of Archaeology, University of Cambridge, Cambridge, United Kingdom, [7]Department of Bioinformatics and Genetics, Swedish Museum of Natural History, [8]Department of Archaeology and Classical Studies, Stockholm University, Stockholm, Sweden, [9]Research Department of Genetics, Evolution and Environment, [10]UCL Genetics Institute, University College London, London, United Kingdom

Abstract: This interdisciplinary project deals with the studies of the temporal population dynamics of the eastern coast of the Baltic Sea, in the territory of present-day Estonia. The region has witnessed several population shifts since people reached its northern part during the Mesolithic 9000 BC. The first people were genetically most similar to the Western hunter gatherer group of Europe [1-3], but considerable shifts occurred within the Neolithic period: first with the arrival of the Comb Ceramic culture people 3900 BC, which brought along Eastern hunter-gatherer genetic ancestry [1-4] and then with Corded Ware culture people of Ponto-Caspian steppe origin [4–7]. The latter wave brought farming into the Eastern Baltic 2800 BC, contrary to most parts of Europe where the Neolithic transition was mediated by Aegean early farmers [5–10].

We present new genomic data from Estonian Late Bronze Age (BA) stone-cist graves (1200–400 BC) and Pre-Roman Iron Age (IA) *tarand* graves (800/500 BC–50 AD). The cultural background of the BA stone-cist graves indicates strong connections to the west as well as to the east, the IA *tarand* layer has been proposed to mirror the culture of Uralic peoples of the Volga-Kama region [11]. To characterize the genetic ancestry of individuals from the so far unstudied cultural layers, we extracted DNA from the teeth of 23 individuals from Late BA Estonia, 14 from Pre-Roman IA Estonia and 7 from medieval Estonia. We analyzed the data in the context of modern and ancient samples.

One of the most notable genetic features of Eastern Baltic populations is a high frequency of Y chromosome haplogroup (hg) N3a; a characteristic shared with mostly Uralic-speaking populations in Europe and several populations all over Siberia [12]. The gene flow from Siberia to western Uralic-speaking populations has also recently been inferred using autosomal data [13]. However, available ancient DNA data has not revealed hg N lineages in Eastern Baltic samples [1–4]. We show that the Siberian component was added to the gene pool of the Eastern Baltic during the transition from BA to IA at the latest. Considering the archaeological context of the samples, this seems to have followed the so-called southwestern route from the Volga-Ural region to the Eastern Baltic [11] and its timing coincides with the hypothesized arrival of westernmost Uralic/Finnic languages in the Eastern Baltic [14], supporting the idea that the spread of these languages was mediated by IA migrants from the east.

We also show that phenotypic characteristics often associated with modern Northern Europeans (light eyes, hair and skin pigmentation, lactose tolerance) can be traced back to the BA in the Eastern Baltic.

References:

1. Jones, E.R. et al. (2017). Curr. Biol. CB 27, 576–582
2. Mittnik, A. et al. (2018). Nat. Commun. 9, 442
3. Mathieson, I. et al. (2018). Nature 555, 197–203
4. Saag, L. et al. (2017). Curr. Biol. CB 27, 2185-2193
5. Lazaridis, I. et al. (2016). Nature 536, 419–424
6. Lazaridis, I. et al. (2014). Nature 513, 409–413

7. Mathieson, Iet al. (2015). Nature 528, 499–503
8. Skoglund, P. et al. (2014). Science 344, 747–750
9. Olalde, I. et al. (2015). MBE. 32, 3132–3142
10. Hofmanová, Z. et al. (2016). PNAS. 113, 6886–6891
11. Lang, V. (2018). UT Press
12. Ilumäe, A.-M. et al. (2016). AJHG. 99, 163–173
13. Tambets, K. et al. (2018). Genome Biol. 19, 139
14. Honkola, T. et al. (2013). JEB. 26, 1244–1253

Open Symposium

SMBE-PO-512

Genetic history of the Maltese archipelago during Neolithic period.

B. Ariano [1,*], V. Mattiangeli [1], L. Cassidy [1], R. Mclaughlin [2], R. Power [3], J. Stock [4], C. Malone [5], B. Mercieca-Spiteri [6], S. Stoddart [4], D. Bradley [1]

[1]Smurfit Institute of Gentics, Trinity College Dublin, Dublin, Ireland, [2]School of Natural and Built Environment, Queen's University Belfast, Belfast, United Kingdom, [3]Department of Ancient History, Macquarie University, Sydney, Australia, [4]Department of Archaeology, University of Cambridge, Cambridge, [5]School of Natural and Built Environment, Queen's University Belfast, Belfast, United Kingdom, [6]The Superintendence of Cultural Heritage, Valletta, Malta

Abstract: The Neolithic period begins in Europe around 8500 years before present (BP) and is characterized by the adoption of farming and domestication of various types of animal. In our project we focus on the structure of the Maltese population during the later part of the Neolithic period. Nine individuals, from 4900 to 4350 years BP, collected from the Xaghra Circle site in the island of Gozo, were sampled. DNA was extracted from both teeth and the inner part of petrous bones giving an average endogenous DNA respectively of: 1.7% for 4 teeth and of 21% for 5 petrous bones. We then used a median of 363,579 SNPs from the Human Origin dataset to compare our samples with 37 ancient individuals from Neolithic and Bronze age period and 604 present-day European individuals already published. PCA analysis places the 3 high coverage Maltese individuals with the early European farmers (EEF) from Germany and Hungary. Further analysis with D-statistics depict that the Maltese population do not resemble any hunter-gatherer population from Caucasus or Eastern Europe, while they show a low affinity with Western European hunter gather individuals (WHG).

Open Symposium

SMBE-PO-457

Title: Genetic consequences of social structure in the golden-crowned sifaka

B. Parreira [1,*], E. Quemere [2], C. Vanpe [3], I. Carvalho [4], L. Chikhi [5]

[1]Institutio Gulbenkian de Ciencia, Oeiras, Portugal, [2]Universite de Toulouse, Touloues, France, [3]Instituto Gulbenkian de Ciencia, Oeriras, [4]Instituto Gulbenkian de Ciencia, Oeiras, Portugal, [5]Laboratorie Evolution & Diversite Biologique, Toulouse, France

Abstract: Title: Genetic consequences of social structure in the golden-crowned sifakaMany mammalian species form social groups with kinship structure where individuals interact with familiar conspecifics during most of their lives. Social groups are far from being sparse aggregations; within social groups individuals cooperate and mate according to complex strategies. Yet, population geneticists have tended to ignore social structure. The traditional genetic approach envisions social groups as relatively small and isolated units of related individuals, and thus many group-living species are believed to be subjected to significant genetic drift, inbreeding and inbreeding-depression effects within social units. This is the case of many studies on lemur species where genetic data are typically analysed under a deme-based approach even though other studies have demonstrated the importance of social structure and identified various mating systems. Here, we investigate inbreeding in an endangered sifaka lemur species (Propithecus tattersalli) restricted to a small region in Northern Madagascar. We measure inbreeding (FIS) at different scales of population structure: social groups, sampling sites and forest patches. Empirically, we found high levels of outbreeding within social groups which can be interpreted as a consequence of inbreeding-avoidance strategies to prevent inbreeding-depression effects. By using simulation framework developed in our group, we simulate social groups without incorporating active inbreeding-avoidance behaviours. Under this model, we show that sifakas are expected to show high levels of genotypic diversity (outbreeding), even though we do not model explicit inbreeding avoidance behaviours. This suggests that demonstrating the existence of active inbreeding avoidance mechanisms may be more difficult than usually assumed.

Open Symposium
SMBE-PO-441
The population genetics of speciation in mouse lemurs
J. Poelstra [1,*], G. Tiley [1], J. Salmona [2], P. Hohenlohe [3], E. Zimmermann [4], J. Ganzhorn [5], A. Yoder [1]
[1]Biology, Duke University, Durham, United States, [2]Laboratoire Evolution & Diversité Biologique, Paul Sabatier University, Toulouse, France, [3]Dept. of Biological Sciences, University of Idaho, Moscow, United States, [4] Institut für Zoologie, University of Veterinary Medicine Hannover, Hannover, [5]Tierökologie & Naturschutz, Universität Hamburg, Hamburg, Germany

Abstract: Understanding the evolution of extant biodiversity largely relies on the inference of historical events, for which genomes provide the most universal and arguably also the most informative data. Several population genetic parameters, in particular, are of key importance for the study of speciation: divergence times, effective population sizes, and migration rates. The accurate inference of these parameters has rapidly become more tractable due to recent developments in coalescent theory and approaches that can deal with large-scale genomic data. With their high diversification rates, the world's smallest primates – mouse lemurs – provide an excellent system to study speciation dynamics. Here, we use whole-genome and RAD sequences to study divergence population genetics in two species complexes of mouse lemurs. While mouse lemurs mostly diverged in isolation, we also detect the presence of post-divergence gene flow, and analyze the consequences of several instances of secondary contact. Most strikingly, we find reproductive isolation among a pair of fairly recently diverged species across one sharp ecological transition, and a complete lack of population genetic structure (despite strong evidence for local adaptation) across a very similar transition, demonstrating the importance of historical contingencies in speciation and ecological adaptation. Furthermore, we infer diversification rates that are even higher than previously thought, which is only partly the result of using a new, direct estimate of the mouse lemur mutation rate. Finally, we find variable trajectories of effective population sizes among species and detect several severe long-term reductions in effective population size, thereby identifying species and regions of heightened conservation concern. These results shed light on the tempo and mode of speciation in mouse lemurs and move towards an understanding of the processes underlying the exceptionally high biodiversity on Madagascar.

Open Symposium

SMBE-PO-448

Mitochondrial variation and provenance of the population of Eastern black rhino in European Zoos

F. Elsner-Gearing [1 2,*], C. Walton [1], M. Pilgrim [2]

[1]The University of Manchester, Manchester, [2]Chester Zoo, Chester, United Kingdom

Abstract: Captive populations are a valuable resource of individuals and genetic variation for dwindling wild populations. As a species, the total population of Black rhino *(Diceros bicornis)* now numbers at around 5,050 individuals and the huge losses it has undergone are reflected by a massive loss of genetic diversity. Many of the founders of the current population of Eastern black rhino (*D. b. michaeli),* held by the zoos of the European Association of Zoos and Aquaria (EAZA), have an unknown origin specified only as from "East Africa". As the establishment of the European population pre-dates the catastrophic decline of *in situ* populations, these animals could therefore represent valuable genetic diversity that has been lost from the wild. Our ongoing project makes use of a set of mitochondrial and nuclear markers to quantify the diversity of this important population and make a direct comparison with wild populations. By incorporating our data into haplotype networks with data from past and present wild populations, we aim to identify the provenance of the founders of the European population. The results of this project will be used to inform management strategies for both *in situ* and *ex situ* conservation efforts and their integration in the form of a One-Plan approach.

Open Symposium
SMBE-PO-451
Tissue hierarchies in plants can efficiently minimize somatic evolution and act as a functional germline
M. Kiss [1,*], I. Derényi [1], G. J. Szöllősi [1]
[1]Biological Physics, Eötvös Loránd University, Budapest, Hungary

Abstract: Plant growth is governed by cell divisions in the apical meristems, which are tissues of undifferentiated cells in the shoot buds. The central zone of each meristem harbors a small group of slowly dividing cells that act as stem cells. From time to time meristems also produce cells that give rise to leaves and a new meristem, called axillary meristem, in the axil of each leaf. These axillary meristems have the potential to turn into apical meristems and start to form new branches, but it is not a priori known which ones.

Recent sequencing studies of within plant genetic variations have revealed that plants are able to limit the accumulation of somatic mutations with unanticipated efficacy. For instance, only a handful of single-nucleotide variants were found between distant branches of a 234-year-old oak [1], as well as between the acorns of several other oak trees [2]. Such a low number of somatic mutations is surprising as it implies that plants manage to keep the number of cell divisions along each lineage low (less than one per year), despite the large number of constantly produced stem cells of the axillary meristems and the inherently stochastic nature of the growth process.

To understand how plants can minimize somatic evolution, we developed a hierarchical meristematic tissue model and used computer simulations to find the optimal parameters that minimize the number of cell divisions along each lineage (leading from the zygote to the cells of the shoot tips). We found that the optimal solution involves slow stem cell divisions in the apical meristems, and an exponentially increasing divisional rate of cells away from the central zone. New axillary meristems are produced from a few cells at the edges of apical meristems via cell divisions along a perfect binary tree.

Longstanding debate has surrounded the existence of a segregated germline in plants [3]. While recent empirical evidence suggests a non-segregated functional germline in plants, it has been unclear how plants are able to produce the steady source of cells with low divisional numbers that are necessary to limit intergenerational mutation rates. Our results demonstrate that there exists an efficient mechanism, which renders germline segregation unnecessary, as the tissue hierarchies underlying plant growth are themselves able to minimize the accumulation of somatic mutations to sufficient extent.

[1] Namrata Sarkar et al.: Low rate of somatic mutations in a long-lived oak tree. Nature Plants, 2017.
[2] Christophe Plomion et al.: Oak genome reveals facets of long lifespan. Nature Plants, 2018.
[3] Robert Lanfear: Do plants have a segregated germline? PLOS Biology, 2018.

Open Symposium
SMBE-PO-450
Detecting novel genetic code reassignments

Y. Shulgina [1,*], S. R. Eddy [1 2]
[1]Harvard University, [2]HHMI, Cambridge, MA, United States

Abstract: Almost all of life uses the same genetic code for translation, suggesting two possibilities: either the genetic code has reached an optimal state, or that changes to the genetic code are so deleterious that it is unable to evolve further. The existence of alternative genetic codes demonstrates that the genetic code can evolve to some degree; however, the evolutionary trajectories by which this happens are still poorly understood. To better study how the genetic code evolves, we have developed a computational tool that can systematically search for new alternative genetic codes among sequenced organisms and characterize their phylogenetic distribution. This tool uses information contained in conserved protein domains to infer the amino acid meaning of each codon. Using this method, as reported independently in two other recent papers, we have identified several yeast species that appear to translate a codon using two differently charged tRNAs with the same anticodon, which may resemble an intermediate stage in genetic code evolution.

Open Symposium

SMBE-PO-453

Strategies for quantitative RNA-seq analyses among closely related species

S. Parekh [1], Z. Karagoez [2], B. Vieth [2], C. Ziegenhain [3], W. Enard [2], I. Hellmann [2,*]

[1]Max Planck Institute for Biology of Ageing, Cologne, [2]Anthropology and Human Genomics, Ludwig-Maximilians University, Munich, Germany, [3]Department of Cell and Molecular Biology, Karolinska Institutet, Stockholm, Sweden

Abstract: With the growing appreciation for the role of regulatory differences in evolution, researchers need to reliably quantify expression levels within and among species. However, for non-model organisms genome assemblies and annotations are often not available or have inferior quality, biasing the inference of expression changes to an unknown extent. Here, we explore the possibility to map RNA-seq reads from diverged species to one high quality reference genome.

As test case, we used a small primate phylogeny ranging from Human to Marmoset spanning 12% nucleotide divergence. To distinguish the effect of sequence divergence and genome quality, we used *in silico* evolved genomes and existing genomes to simulate RNA-seq reads. These were then mapped to the genome of origin (self-mapping) as well as to one common reference (cross-mapping) to infer the quantification biases.We find that the bias due to cross-mapping is small for the closely related great apes (≥4% divergence), and preferable to self-mapping given current genome qualities. For closely related species, cross-mapping provides easy access, high power and a well controlled false discovery rate for both; the analysis of intra-species expression differences as well as the detection of relative differences between species. If divergence increases, so that a substantial fraction of reads exceeds the limits of the mapper used, we find that gene-specific corrections and effect-size cutoffs can limit the bias before self-mapping becomes unavoidable.In summary, for the first time we systematically quantify biases in cross-species RNA-seq studies, providing guidance to best practices for these important evolutionary studies.

Open Symposium
SMBE-PO-472
Genetic study of ancient and recent populations in Central Asia
P. Guarino--Vignon [1,*], C. Bon [1], E. Heyer [1], N. Marchi [1 2]
[1]UMR 7206, Muséum National d'Histoire Naturelle, Paris, France, [2]CMPG, Institute of Ecology and Evolution, University of Berne, Berne, Switzerland

Abstract: Southern Central Asia has been a crossroad for the movements of populations, cultures and goods between Europe, South Asia, East Asia and Middle East since prehistory, and a place of cohabitation between different cultures. Indeed, Central Asia is under both a Northern influence with the Steppe cultures and Southern with the Iranian culture. Furthermore, during the last 2000 years, populations speaking Turko-Mongol languages migrating from the East Asia have partially replace the Indo-Iranian speakers.This led to a strong genetic difference between the modern Indo-Iranian and the Turko-Mongol speaking populations, with various amounts of admixture between them. However, little is known about the origin of the Indo-Iranian populations before 2000 years BP. Based on archaeological data, it has been suggested that they were related to both Central Asia steppe populations (such as Andronovo) and populations from Iran. However, if ancient DNA analyses have shown that migrations from the Steppe have significantly contributed to the modern European genetic diversity, south and east migrations towards India, Central Asia and Mongolia have only been suggested. Thus, despite its rich and complex history, Central Asia lacks genetic studies involving ancient DNA to unravel the peopling of this region.
We propose to test this hypothesis by combining modern genetic data obtained in our lab, to ancient genomes from Eurasia which a substantial number has been recently published. Through genomic analyses, we revealed complex interactions between South Caucasian, Iranian and Levant populations during Protohistory. Using D-statistics, we first showed a clear proximity of modern Indo-Iranian speaking populations from Central Asia with the ancient Steppe populations from the Neolithic to the Iron Age. However, they are not strongly related to ancient Iranians, despite insights from archaeological reports. We also explored, with the help of ADMIXTURE, PCA, D-stats, F3 stats and qpAdm, the relationships between modern Central Asia populations and ancient populations from the Steppes north to Central Asia, the Pontic Steppes, Iran, the Near East, and Caucasius. Our results suggest that ancient populations from north and south Caucasius are closer to modern Central Asia than Iranian or Central Asia Steppes (like Andronovo). Finally, our results show that the history of Central Asia and its peopling is more complex than the common hypothesis of a double influence from its north and its south, but suggest further population migrations from west.

Open Symposium

SMBE-PO-470

Initiatives for genetic data collection from underrepresented countries and populations

K. F. McManus [1,*], M. Moreno [1], A. Shastri [1], S. Micheletti [1], E. Jewett [1], G. D. Poznik [1], K. Bryc [1], J. Mountain [1]
[1]23andMe, Inc., Mountain View, United States

Abstract: Currently available human genetic data do not adequately represent the breadth and depth of human diversity. The lack of genetic data from people of diverse ancestries not only impedes a comprehensive understanding of genetic variation in *Homo sapiens*, but also reduces our chances of identifying genetic variants associated with both common and rare traits, and hinders research into understanding human evolutionary history.

23andMe currently has three initiatives to increase the breadth and quantity of genetic data from diverse populations. First, we invite all of our customers to participate in research, and we ask those who consent to opt-in to tell us about their ancestries. Second, the Global Genetics Project, launched in 2018, seeks to genotype individuals living in the United States who have four grandparents from large populations throughout the world that are underrepresented in 23andMe's research cohort. More than 6,000 participants, with grandparents from 55 countries, have enrolled thus far. Our Population Collaborations Program, also launched in 2018, seeks partnerships with academic researchers working with populations that are not well represented within the United States. Previous collaborations have yielded data from Sierra Leone, The Democratic Republic of the Congo, South Africa, and Angola.

These data have been used to improve ancestry inference for underrepresented groups in Asia and Africa. They will also be used for a variety of current and future ancestry and health research projects to improve understanding of population structure, diversity, and migrations throughout the world and to further health research relevant to all populations, including populations underrepresented in current genetics research.

Open Symposium
SMBE-PO-475
Possible associative overdominance in low recombination regions of the Drosophila genome
B. Charlesworth [1,*], H. Becher [1]
[1]Institute of Evolutionary Biology, University of Edinburgh, Edinburgh, United Kingdom

Abstract: Recessive or partially recessive deleterious mutations can retard the rate of loss of variability by genetic drift at closely linked neutral sites, provided that the product of effective population size and the selection coefficient for the mutations is of order one (associative overdominance, or AOD). It is commonly thought that AOD can only occur in very small populations. However, in regions of the genome with very low rates of genetic recombination, selective sweeps and background selection can reduce the effective population size to a fraction of the genome-wide average. This suggests that AOD could be induced in these regions by weakly selected deleterious mutations. Computer simulations show that this is likely to occur for realistic selection and mutation parameters, for a model of low recombination regions of the *Drosophila* genome. The results are consistent with observations on the degree of distortion of the site frequency spectra at synonymous sites in these genomic regions in *D. melanogaster* and *D. simulans*.

Open Symposium

SMBE-PO-486

Expression of endogenous retroviruses in canine cancer-derived cell lines

A. Jarosz [1,*], E. L. Cech [2], M. L. Day [3], J. V. Halo [1]

[1]Biological Sciences, Bowling Green State University , [2]Biological Sciences, Bowling Green State Univeristy, [3]Bowling Green State University, Bowling Green , United States

Abstract: In contrast to other mammals, the domestic dog genome displays a substantially lower endogenous retrovirus (ERV) presence, at just 0.15% of their nuclear genome (vs. 8-10% in human and mouse). To this day, there have been no confirmed infectious exogenous retroviruses (XRVs) in the dog, or any candid, despite their being constantly challenged from numerous XRVs circulating in prey such as mice or cats. In recent analyses of the canine reference genome, a few copies of ERVs were identified with features characteristic of recent integration, for example the presence of some ORFs and near-identical LTRs. Members of this group are referred to as 'CfERV(Fc-1)' and have been identified to have sequence similarity to the mammalian ERV-Fc/W groups. We have recently discovered and characterized a number of non-reference Fc1 copies in dogs and wild canids, and identified unexpectedly high levels of polymorphism among members of this ERV group. Some of the proviruses we have identified even possess complete or nearly intact open reading frames, identical LTRs, and derived phylogenetic clustering among other CfERV(Fc-1) members. Based on LTR sequence divergence under an applied dog neutral mutation rate, it is thought these infections occurred within as recently as the last ~0.48 million years. There have been previous, but unsubstantiated, reports of reverse transcriptase activity as well as gamma-type C particles in tumor tissues of canines diagnosed with lymphoma. We hypothesize that expression of members of the CfERV(Fc-1) lineage is responsible for those observations in canine cancers. We profiled the expression of individual proviruses in canine cancer cell lines, specifically the *pol*and *env*gene. There was expression of both genes in three canine-cancer derived cells lines that cluster with CfERV(Fc-1) members. Clustering of these sequences also suggest that there is either a new sub lineage of CfERV(Fc-1) or possibly missed polymorphic proviral insertions that are currently assembled as solo LTRs.

Open Symposium

SMBE-PO-473

Early germline specification in bivalves: searching for OSKAR functional homolog

B. Filanti [1,*], G. Piccinini [1], S. Bettini [1], M. Lazzari [1], V. Franceschini [1], M. G. Maurizii [1], L. Milani [1]

[1]Department of Biological, Geological and Environmental Sciences - BiGeA, University of Bologna, Bologna, Italy

Abstract: Germ cells play a unique role in heredity and evolution as carriers of genetic information across generations, but the study of the early stage of germ cell specification is still an overlooked issue.

In this study, we documented the earliest stages of germ cell specification in *Ruditapes philippinarum* (the Manila clam), describing both the mechanism and timing. Initially, we focused our investigation on OSKAR, the necessary and sufficient factor involved in the earliest stage of *Drosophila* germline specification. Despite OSKAR is considered to have no orthologues even among insects, on the basis of bioinformatic analyses we identified a Tudor domain-containing protein 7 (TDRD7) as the possible best candidate acting in the corresponding stage of germline specification in *R. philippinarum*. This hypothesis is supported by domain similarities between OSKAR and TDRD7, as well as *in situ* localization of the latter in putative germ cells in their earliest stage of differentiation from germline stem cells.

In this light, germline specification appears to show evolutionarily conserved features, while maintaining some kind of plasticity, thus allowing differences among different taxa. Although OSKAR orthologues are not known, a corresponding role in other animal taxa has to be carried out, possibly by different proteins with similar functions. We propose TDRD7 as a germline factor acting at early stages of germline specification in *R. philippinarum*, thus a good candidate as OSKAR functional-homolog.

Open Symposium
SMBE-PO-481
Convergent evolution of piscivory in cone snails
S. Abalde [1], M. J. Tenorio [2], C. M. Afonso [3], R. Zardoya [1,*]
[1]Biodiversity and Evolutionary Biology, Museo Nacional de Ciencias Naturales CSIC, Madrid, [2]CMIM y Química Inorgánica-INBIO, Universidad de Cadiz, Puerto Real, Spain, [3]Fisheries, Biodiversity and Conservation, CCMAR-University of Algarve, Faro, Portugal

Abstract: Cone snails (Gastropoda: Conidae) prey on marine worms, snails, and fishes by producing and injecting venom through a harpoon-like radular tooth. Transcriptomics is becoming the tool of choice for cataloguing the different conotoxins that form the venom cocktails of the different cone species. Here, we present the trancsriptome of the venom gland of the amphi-Atlantic piscivorous cone *Chelyconus ermineus*. Transcriptomic data from three individuals and three regions of the venom duct (proximal, medium, and distal) showed a great diversity of peptides produced in the venom as well as compartmentalization of expression. We compared the newly determined conotoxins with those present in the venoms of piscivorous cones from the Indo-Pacific. We found that members of superfamily A, which are hypothesized to be key for preying on fish, are different in *C. ermineus* and in piscivorous Indo-Pacific cones suggesting a convergent evolution of piscivory in cones in agreement with the different relative position of Atlantic versus Indo-Pacific piscivorous species in the phylogeny of the group.

Open Symposium
SMBE-PO-464
Life in the extreme; when did tardigrades colonise Antarctica?
K. Short[*], S. McInnes[1], D. Pisani[2], J. Lozano Fernandez[2], C. Sands[1], P. Convey[1]
[1]British Antarctic Survey, Cambridge, [2]University of Bristol, Bristol, United Kingdom

Abstract: Internal phylogenetic relationships within the Tardigrada are poorly resolved, and the studies that have been undertaken in this area are generally restricted to certain groups within the phylum, as well as rarely including molecular analyses of divergence times. In this ongoing study, we infer the phylogeny of the phylum using 18S, 28S and CO1 sequences obtained from GenBank. To investigate the colonisation history of tardigrades in Antarctica, further newly-available sequences for *Acutuncus antarcticus* and *Mesobiotus furciger* for different geographical regions were added. Trees were created using Maximum Likelihood and Bayesian methods, with divergence times estimated using a relaxed clock model. These were then mapped to a biogeography matrix to estimate biogeographic patterns. Ongoing analyses suggests that the biogeographic history of *A. antarcticus* and *M. furciger* may be different to that previously proposed. Divergence times indicate that the origin of *M. furciger* is older than previously thought, with the Antarctic lineage of Mesobiotus diverging approximately 50 mya from other non-Antarctic lineages of Macrobiotidae, well before the geographical isolation of Antarctica. The data also indicate that within the species' clade a significant radiation event took place approximately 10 mya, implying that *M. furciger* is in fact a species complex with at least three distinct haplotypes existing. This pattern is similarly seen within *A. antarcticus*, indicating tardigrades having ancient origins on this extreme continent.

Open Symposium

SMBE-PO-463

Head, Shoulders, Knees, and Toes: The Comparison of DNA Preservation Across Multiple Skeletal Elements From Individuals Recovered From the Abandonded Medieval Graveyard of Krakauer Berg, Germany

C. Parker [1,*], S. Friederich [2], A. B. Rohrlach [1], W. Haak [1], K. I. Bos [1], J. Krause [1]

[1]Department of Archaeogenetics, The Max Planck Institute for the Science of Human History, Jena, [2]Landesamt für Denkmalpflege und Archeologie , Landesmuseum für Vorgeschichte, Halle, Germany

Abstract: Ancient DNA (aDNA) analyses necessitate the destructive sampling of precious archaeological material. The *pars petrosa* (the portion of the temporal bone that houses the inner ear) is currently the most sought after skeletal element, as it tends to yield higher proportions of endogenous DNA, even from climatic regions that have been previously deemed unsuitable for aDNA analysis. However, the micro-structure of the inner ear is also in high demand for morphological studies or may otherwise be unavailable for sampling. In light of this, a systematic investigation that compares host DNA preservation across skeletal elements would be beneficial. Here we present a comprehensive survey of endogenous human aDNA preservation in multiple sampling sites from each of 10 skeletal elements stemming from 11 individuals excavated from the abandoned 12th century cemetery of Krakauer Berg, Germany and using high-throughput, automated, single-stranded library preparation and Illumina short read sequencing. Our analysis shows that while the dense areas around the *cochlea* of the *pars petrosa* retain the highest percentage of endogenous DNA on average, there are a range of other elements that also yield sufficient aDNA for a variety of applications. Our analyses utilize both standard bioinformatics packages and modern statistical approaches to consider several aspects of aDNA preservation including: endogenous human DNA content, contamination levels, and library complexity. This ranked listing of skeletal elements will provide investigators a better perspective on DNA preservation across the skeleton and aid in sample selection in a wide range of aDNA studies, especially for incomplete or fragmentary individuals.

Open Symposium

SMBE-PO-462

ORTHOSCOPE: an automatic web tool for estimating the origins and the functions of bilaterian protein-coding genes by comparing gene and species trees

J. Inoue [1,*], N. Satoh [2]

[1]Genomics and Evolutionary Biology, National Institute of Genetics, Mishima, [2]Marine Genomics, Okinawa Institute of Science and Technology, Okinawa, Japan

Abstract: Orthology allows us to describe the evolution of genomes and understand the function of genes. To identify orthologs among deeply diversified lineages such as bilaterians, precise estimation of a gene tree is indispensable given the complicated histories of genes over millions of years. By estimating gene trees, orthologs can be identified as members of an orthogroup, the set of genes that are descended from a single gene in the last common ancestor of all the species being considered. In addition to comparisons with a species tree, purposeful taxonomic sampling increases the accuracy of gene tree estimation and orthogroup identification. Although some major phylogenetic relationships of bilaterians are gradually being unraveled, scattering of published genome data among separate web databases is becoming a serious barrier to appropriate taxonomic sampling. By integrating more than 350 gene models with special reference to bilaterians, we developed a web tool named ORTHOSCOPE to identify orthogroups of specific protein-coding genes. Under user-specified taxon samplings, ORTHOSCOPE identifies orthogroups within estimated gene trees with reference to broadly-accepted species-nodes. Based on the delineated orthogroup including model organisms, users can estimate the origins and functions of focal genes. ORTHOSCOPE allows users to upload several sequences of a specific gene and a species tree as queries. ORTHOSCOPE is freely available at https://www.orthoscope.jp (last accessed 3 February 2019).

Open Symposium

SMBE-PO-460

Recovering a Native American lineage extinct during the Conquest Period from Admixed Genomes

M. A. C. E. Silva [1,*], K. Nunes [1], R. B. Lemes [1], À. M. Sandoval [2], C. E. Guerra Amorim [3], J. G. Mill [4], F. M. Salzano [5], M. C. Bortolini [5], A. D. C. Pereira [6], D. Comas [2], T. Hünemeier [1]

[1]Instituto de Biociências, Departamento de Genética e Biologia Evolutiva, Universidade de São Paulo, São Paulo, Brazil, [2]Institut de Biologia Evolutiva, Departament de Ciències Experimentals i de la Salut, Universitat Pompeu Fabra, Barcelona, Spain, [3]University of California, Los Angeles, United States, [4]Departamento de Fisiologia, Universidade Federal do Espírito Santo, Vitória, [5]Instituto de Biociências, Departamento de Genética, Universidade Federal do Rio Grande do Sul, Porto Alegre, [6]Instituto do Coração, Hospital das Clínicas da Faculdade de Medicina da Universidade de São Paulo, São Paulo, Brazil

Abstract: The conquest of South America by Europeans in the 15th century greatly impacted the genetic, linguistic and ethnic diversity of indigenous populations of all regions of the continent. For instance, at that time, around 900,000 Native Americans lived on the Brazilian coast and, by the end of the 18th century, all coastal native populations were declared extinct. Here we leveraged genomic data from the last remaining putative representatives of the Tupi coastal branch, a small, admixed. Self-reported Tupiniquim community, as well as date of a Guarani (Tupi) population from Southern Brazil and of three populations from Amazonian region (n=102). We demonstrated that the Tupiniquim Native American ancestry is not related to any extant Brazilian Native American population already studied (19 groups belonging to 6 linguistics stocks) and thus they could be considered the only living representative of the extinct Tupi branch that used to settle the Atlantic Coast of Brazil. Furthermore, these data show evidence of a direct migration from Amazonian to the Northeast Coast in pre-Columbian time, giving raise to the Tupi Coastal populations, and a single distinct migration southward that originated the Guarani people from Brazil and Paraguay. Brazil is the country with the greatest diversity of Native Americans, presenting more than 120 different languages. This is the first study to elucidate the population dynamics and diversification of the Brazilian Natives at genomic level, and this was only possible because we were able to recover data from the Brazilian extinct coastal population through the genomes of mestizo individuals.

Open Symposium

SMBE-PO-466

Trio deep-sequencing does not reveal unexpected off-target and on-target mutations in Cas9-edited monkeys

X. Luo [1,*], Y. He [1], B. Su [1]

[1]Chinese Academy of Sciences, Kunming Institute of Zoology, Kunming, China

Abstract: CRISPR-Cas9 is a widely-used genome editing tool, but its off-target effect and on-target complex mutations remains a concern, especially in view of future clinical applications. Non-human primates (NHPs) share close genetic and physiological similarities with humans, making them an ideal preclinical model for developing Cas9-based therapies. However, no comprehensive *in vivo* off-target and on-target assessment has been conducted in NHPs. Here, performing whole genome trio sequencing of Cas9-treated monkeys, we found they only carried a small number of *de novo* mutations that can be explained by expected spontaneous mutations, and no unexpected off-target mutations were detected. Also, using long-read sequencing, we not found large deletions or complex rearrangements at the targeted regions induced by CRISPR-Cas9 system.

Open Symposium

SMBE-PO-459

The architecture of genotype networks underlying whole-organism phenotypes in Arabidopsis thaliana

G. Schweizer [1 2,*], A. Wagner [1 2 3]

[1]Evolutionary Biology and Environmental Studies, University of Zürich, Zürich, [2]Swiss Institute of Bioinformatics, Lausanne, Switzerland, [3]Santa Fe Institute, Santa Fe, United States

Abstract: Understanding how genotypes are linked to phenotypes and ultimately to fitness is of central interest in biology, because this link provides the foundation of evolution. A common approach to study the relationship between genotype and phenotype relies on the concept of genotype-phenotype maps [1]. One way to build such maps employs the concept of genotype networks, where sets of genotypes are depicted as network. Vertices of such a network represent nucleotide or amino acid sequences, and they are connected by an edge if the underlying sequences differ by a single small mutation[2]. In this way, genotype networks highlight the effect of each small mutation on a phenotype. So far, genotype networks were used to investigate systems below the level of whole-organism phenotypes, like protein folding or binding of transcription factors to DNA. Therefore, we sought to unravel the architecture of genotype networks underlying whole-organism phenotypes. To this end, we employed the results of a previously conducted genome-wide association study in the model plant Arabidopsis thaliana [3]. In this study, genomic loci underlying phenotypes like flowering time, leaf development, and resistance to pathogen attacks or heavy metals were identified. Here, we show that the results of this study are suitable to reconstruct connected genotype networks and analyze their properties. We provide insights in the link between one-mutant neighbors in the genotype network and their phylogenetic distance, the accessibility of different phenotypes, and analyze if different neighborhoods in a genotype network bring forth different novel phenotypes. In summary, our work extends current studies of genotype networks to whole-organism phenotypes and provides further insights in the link between genotypes and phenotypes.

[1] Alberch P (1991) From genes to phenotype: dynamical systems and evolvability. *Genetica* 84, 5-11.

[2] Carter H, Hofree M, and Ideker T (2013) Genotype to phenotype via network analysis. *Curr Opin Genet Dev* 23(6), 611-621.

[3] Atwell S et al. (2010) Genome-wide association study of 107 phenotypes in Arabidopsis thaliana inbred lines. *Nature* 465(7298), 627-631.

Open Symposium
SMBE-PO-458
Evolvability of orthologous genes
H. abdalaal [1 1 1 1,*], J. Näsvall
[1]Biomedicinskt Centrum, Uppsala, Sweden

Abstract: Understanding how a specialized enzyme can evolve from a progenitor enzyme with promiscuous activity is important for explaining the evolution of novel phenotype.
The project aims to generate a library of laboratory-evolved orthologous genes for studying evolvability, the *hisA* gene from *Salmonella enterica* was taken through alternating rounds of weak selection (stimulated by random mutagenesis and screen for partial loss of function) followed by strong selection (stimulated by additional random mutagenesis and selection for restored function).
After the first few rounds of mutagenesis and screening/selection, we have found twenty deleterious and twelve compensatory mutations. Surprisingly, after performing the third round of mutagenesis and screening for loss of activity, two lineages reverted the compensating mutation rather than acquiring a second deleterious mutation. This could indicate that these compensating mutations are able to mask the effects of several deleterious mutations, and that they could be global suppressors that make the protein more robust, e.g. by making the structure more stable or by improving folding. Such mutations would be able to compensate against a wide variety of deleterious mutations, and thus make the protein more evolvable.
By combining these compensating mutations with other deleterious mutations and testing the function of the resulting HisA enzymes, we confirmed that these mutations are global suppressors. Further experiments will test the effects on evolvability, defined as the ability to accommodate additional mutations and/or acquire a new function (TrpF enzymatic activity). We have also tested the evolvability of diverging lineages (six SNPs different from wild type *hisA)*, we found out these lineages are more evolvable than the wild type.

Expanded Summary*: My projects are part of ongoing efforts to have a better understanding of how new genes diverge and evolve. We are applying experimental evolution by stimulating what happens in nature during the divergence of orthologous genes. we are simulating a “speeding up” of the molecular clock and then analyzing the divergent sequences for their “evolvability” towards a new function. That will provide us better insight into how populations adapt to a particular environment.
Also, it can be used to identify how certain enzyme can acquire drug resistance before such phenotypes appear in nature. This approach might facilitate in advance development of new drugs that target the resistant enzyme.
Furthermore, having deeper knowledge about gene evolution has great potential in the medical field. For example; it has been used to develop vaccines against viral and bacterial infections.

Open Symposium
SMBE-PO-461
Understanding X suppression: a sex- and tissue-specific mechanism of sex chromosome regulation in Drosophila
E. Argyridou [1,*], J. Parsch [1]
[1]Faculty of Biology, LMU Munich, Planegg-Martinsried, Germany

Abstract: In the XY sex chromosome system present in both mammals and *Drosophila*, males are hemizygous for the X chromosome. The disparity in ploidy between sexes exposes the X chromosome to unique selective forces that cause it to evolve special regulatory mechanisms. In the male soma of *D.melanogaster*, the mechanism of dosage compensation balances gene expression between the single X chromosome and the autosomes. Yet, in the male germline the expression of testis-specific reporter and transposed genes residing on the X chromosome is suppressed, a phenomenon known as X suppression. Here, we aim to understand the obscure mechanism behind X suppression. First, by analyzing the expression of native genes, we find that the X chromosome contains disproportionately fewer genes with high expression in testis than the autosomes, even after accounting for the lack of dosage compensation, which is consistent with the suppression of highly-expressed X-linked genes in the male germline. Second, we test a ubiquitously-expressed reporter gene and show that it is not affected by X suppression. This may be a consequence of the low expression of this reporter gene, as we also find that the extent of X suppression is positively correlated with expression level. Finally, in order to identify genes that are involved in X suppression, we perform a forward genetic screen through chemical mutagenesis and functionally test candidate genes. Identifying genes involved in this process will lay the groundwork for uncovering the molecular mechanism and the evolutionary forces that led to such specialized sex- and tissue-specific gene regulation.

Open Symposium
SMBE-PO-469
Fast and accurate statistical evolutionary alignment
N. De Maio [1,*]
[1]European Bioinformatics Institute, EMBL-EBI, Hinxton, Cambridge, United Kingdom

Abstract: Sequence alignment is essential for phylogenetic and molecular evolution inference, as well as in many other areas of bioinformatics and evolutionary biology.
Inaccurate alignments can lead to severe biases in most downstream statistical analyses.
Statistical alignment based on probabilistic models of sequence evolution addresses these issues by replacing heuristic score functions with evolutionary model-based probabilities.
However, score-based aligners and fixed-alignment phylogenetic approaches are still more prevalent than methods based on evolutionary indel models, mostly due to computational convenience.
Here, I present new techniques for improving the accuracy and speed of statistical evolutionary alignment.
The "cumulative indel model" approximates realistic evolutionary indel dynamics using differential equations.
"Adaptive banding" reduces the computational demand of most alignment algorithms without requiring prior knowledge of divergence levels or pseudo-optimal alignments.
Using simulations, I show that these methods lead to faster and more accurate parameter and pairwise alignment inference.
The cumulative indel model and adaptive banding can therefore improve the performance of alignment and phylogenetic methods.

Open Symposium
SMBE-PO-471
The demographic history of Arabidopsis thaliana: How a weedy population conquers Eurasia
C.-R. Lee [1,*]
[1]Institute of Ecology and Evolutionary Biology, National Taiwan University, Taipei, Taiwan

Abstract: The demography of most species is dynamic, with constant migration exchanging alleles and isolation events splitting populations. Working on the population genomics of *Arabidopsis thaliana*, we aim to uncover the past demography of this model species. We found that this species consists of multiple highly diverged groups, where the 'relicts' occupied post-glacial Eurasia first and were later replaced by the invading weedy 'non-relicts', which expanded through the east–west axis of Eurasia. Admixture happened along the expansion route, and in Iberia, relict introgression introduced genes locally adaptive to the abiotic environments, such as those associated with root cap development or metal ion trans-membrane transport. In China, the Yangtze River Basin population represents the most recent non-relict invasion event in Eurasia, dated to 500-1000 AD. This population also has admixture with local relicts, obtaining locally adaptive and disease-related genetic variants. Therefore in both the eastern and western ends of Eurasia, relict introgression likely introduced adaptive genetic variants into the invading non-relicts, facilitating their colonization of the whole continent in 10,000 years. We are currently investigating factors contributing to the 'weediness' of non-relicts and making them competitive over local relicts.

Open Symposium

SMBE-PO-465

Local ancestry influences gene regulation in naturally hybridizing wild baboons

T. Vilgalys [1,*], A. Fogel [1 2], J. Anderson [1], A. Lea [3 4], T. Voyles [1], M. Akinyi [5], S. Alberts [5 6], J. Tung [1 5 6 7]

[1]Evolutionary Anthropology, [2]University Program in Genetics and Genomics, Duke University, Durham, [3]Ecology and Evolutionary Biology, [4]Lewis-Sigler Institute for Integrative Genomics, Princeton University, Princeton, United States, [5]Institute of Primate Research, National Museums of Kenya, Nairobi, Kenya, [6]Biology, [7]Population Research Institute, Duke University, Durham, United States

Abstract: Changes in gene regulation are thought to play an important role in evolution and divergence. Comparative evidence clearly demonstrates that gene expression patterns differ between closely related species. However, we still know little about how gene regulation is affected when species hybridize upon secondary contact in natural hybrid zones. To address this gap, we estimated local genetic ancestry from a genome-wide resequencing data set of 179 baboons from a yellow baboon x anubis baboon hybrid zone in Kenya. We integrated this information with genome-wide gene expression and DNA methylation data from the same individuals. We detect 980 genes (16.8% of those tested) and 34,912 CpG sites (5.0%) where local genetic ancestry predicts gene expression and DNA methylation levels, respectively (10% FDR). In most cases, ancestry effects are additive, such that admixed individuals fall between the parental taxa; however, in a small minority of cases (47 genes, 947 CpG sites), we observe evidence for transgressive trait values that may point to loci involved in reproductive isolation or heterosis. Finally, we find that CpG sites affected by local genetic ancestry are strongly enriched among sites that we previously showed to be differentiated between unadmixed yellow and anubis baboons from outside the hybrid zone ($\log_2(OR)=6.06$, $p<10^{-16}$). Together, our results indicate that genetic ancestry significantly contributes to gene regulatory variation in naturally admixing wild primates. They thus highlight the value of such populations for testing how selection acts on introgressed regions, potentially as a model for admixture between humans and other archaic hominins.

Open Symposium

SMBE-PO-505

Shifts in the genetic landscape of the western Eurasian Steppe associated with the beginning and end of the Scythian dominance

M. Järve [1], L. Saag [1 2,*], C. L. Scheib [1], A. K. Pathak [1], F. Montinaro [1], L. Pagani [1], J. R. Flores [1], M. Guellil [1], L. Saag [1], K. Tambets [1], A. Kushniarevich [1], A. Solnik [1], L. Varul [3], S. Stadnikov [4], O. Petrauskas [5], M. Avramenko [5], B. Magomedov [5], S. Didenko [6], G. Toshev [7], I. Bruyako [8], D. Grechko [5], V. Okatenko [9], K. Gorbenko [10], O. Smyrnov [10], A. Heiko [11], R. Reida [5], S. Sapiehin [12], S. Sirotin [13], A. Tairov [14], A. Beisenov [15], M. Starodubtsev [16], V. Vasilev [17], A. Nechvaloda [18], B. Atabiev [19], S. Litvinov [18], N. Ekomasova [18], M. Dzhaubermezov [18], S. Voroniatov [20], O. Utevska [4], I. Shramko [4], E. Khusnutdinova [18], M. Metspalu [1], N. Savelev [18], A. Kriiska [21], T. Kivisild [1 22], R. Villems [1]

[1]University of Tartu Institute of Genomics, [2]University of Tartu Institute of Molecular and Cell Biology, Tartu, [3]Tallinn University, Tallinn, Estonia, [4]V. N. Karazin Kharkiv National University, Kharkiv, [5]National Academy of Sciences of Ukraine, [6]National Museum of History of Ukraine, Kyiv, [7]Zaporizhia National University, Zaporizhia, [8]Odessa Archaeological Museum, Odessa, [9]National Museum of History of Ukraine, Kharkiv, [10]Mykolaiv V. O. Sukhomlynskyi National University, Mykolaiv, [11]National Museum of Ukrainian Pottery in Opishne, Opishne, [12]Poltava Regional Makarenko Scientific Lyceum, Kovalivka, Ukraine, [13]Russian Academy of Sciences, Moscow, [14]South Ural State University, Chelyabinsk, Russian Federation, [15]A. H. Margulan Institute of Archaeology, Almaty, Kazakhstan, [16]Sterlitamak Krai Museum, Sterlitamak, Russian Federation, [17]LoCom Medien Akademie Europäisches Bildungsinstitut, Bonn, Germany, [18]Ufa Federal Research Centre of the Russian Academy of Sciences, Ufa, [19]Institute of the Caucasus Archaeology, Nalchik, [20]State Hermitage Museum, St. Petersburg, Russian Federation, [21]University of Tartu Institute of History and Archaeology, Tartu, Estonia, [22]KU Leuven, Leuven, Belgium

Abstract: The Early Iron Age nomadic Scythians have been described as a confederation of tribes of different origins, based on ancient DNA evidence [1-3]. It is still unclear how much of the Scythian dominance in the Eurasian Steppe was due to movements of people and how much reflected cultural diffusion and elite dominance. We present new whole-genome sequences of 31 ancient Western and Eastern Steppe individuals including Scythians as well as samples pre- and postdating them, allowing us to set the Scythians in a temporal context (in the Western/Ponto-Caspian Steppe). We detect an increase of eastern (Altaian) affinity along with a decrease in Eastern Hunter-Gatherer (EHG) ancestry in the Early Iron Age Ponto-Caspian gene pool at the start of the Scythian dominance. On the other hand, samples of the Chernyakhiv culture postdating the Scythians in Ukraine have a significantly higher proportion of Near Eastern ancestry than other samples of this study. Our results agree with the Gothic source of the Chernyakhiv culture and support the hypothesis that the Scythian dominance did involve a demic component.

1. Damgaard, P.B., Marchi, N., Rasmussen, S., Peyrot, M., Renaud, G., Korneliussen, T., Moreno-Mayar, J.V., Pedersen, M.W., Goldberg, A., Usmanova, E., et al. (2018). 137 ancient human genomes from across the Eurasian steppes. Nature *557*, 369-374.

2. Krzewinska, M., Kilinc, G.M., Juras, A., Koptekin, D., Chylenski, M., Nikitin, A.G., Shcherbakov, N., Shuteleva, I., Leonova, T., Kraeva, L., et al. (2018). Ancient genomes suggest the eastern Pontic-Caspian steppe as the source of western Iron Age nomads. Sci Adv *4*, eaat4457.

3. Unterlander, M., Palstra, F., Lazaridis, I., Pilipenko, A., Hofmanova, Z., Gross, M., Sell, C., Blocher, J., Kirsanow, K., Rohland, N., et al. (2017). Ancestry and demography and descendants of Iron Age nomads of the Eurasian Steppe. Nat Commun *8*, 14615.

Open Symposium

SMBE-PO-510

An Automated Pipeline For Detecting Recombination Patterns In Bacteria Through Tree Topology Variations

A. K. Lankapalli*, A. Herbig [1]

[1]Department of Archaeogenetics, Max Planck Institute for the Science of Human History, Jena, Germany

Abstract: Recombination plays a pivotal role in bacterial evolution. It is a process where bacteria take up and integrate DNA from the surrounding environment through transformation, transduction and conjugation. Bacteria evolve by accumulating mutations and phylogenetic trees represent these variations to elucidate evolutionary relationships. As recombination introduces new genetic material, it tends to disrupt these phylogenetic signals by inducing incongruence in the tree topologies. In this case, a single tree would be insufficient to explain evolutionary processes and thus recombinant regions should be determined. Here we present an automated pipeline to detect tree topology variations to predict and characterize recombination events.

To identify recombination events in a specific strain (target), a maximum likelihood tree (backbone tree) is generated for all the representative bacterial strains in the alignment excluding the target sequence. On this backbone tree, the target is placed on all possible positions to generate tree models. Through an overlapping sliding window of alignments, likelihood trees are constructed and our pipeline compares the tree topologies and estimates those that are consistent and inconsistent with the data. Recombination events can be indicated by genomic regions with contradicting tree topologies.

We demonstrate the application of our pipeline using the highly recombinant gut microbe *Helicobacter pylori*. It has been co-evolving with anatomically modern humans and was likely associated with them prior to their journey out of Africa. It presents phylogeographical patterns similar to those of humans and is categorized into seven lineages and sub-lineages, which show mosaic patterns of genetic exchange between each other. We sought to explore recombination events in PeCan4, a strain of *H. pylori* isolated from an individual from Peru. It is described as a recombinant of a South American and a European *H. pylori* lineage. Using representative strains from each lineage, tree topologies were reconstructed,genomic regions were assessed with respect to their phylogenetic signals and possible recombinant regions across PeCan4 genomes are visualized.Our pipeline not only highlights genomic candidate regions for recombination events but also estimates their effect on phylogenetic reconstruction.

Open Symposium

SMBE-PO-513

The fitness consequences of synonymous mutations in Escherichia coli: experimental evidence for a pleiotropic effect of translational selection

D. B. Carlini [1,*], A. Ballard, S. Bieniek

[1]Biology, American University, Washington DC, United States

Abstract: Codon usage bias (CUB) is a universal feature of genomes, and in most species CUB of protein coding genes is positively correlated with expression level and degree of evolutionary conservation. There is mounting experimental evidence that CUB is due in part to selection for translational efficiency and/or accuracy, *i.e.*, translational selection. However, there is a paucity of experimental data on whether and how CUB acts in *trans* – does the usage of preferred codons in a highly expressed gene affect the translation of other genes by freeing up more ribosomes, thereby increasing their availability to translate all mRNA transcripts in the cell? We investigated this question by creating two extreme versions of the highly expressed *Escherichia coli b-lactamase* (*bla*) gene, one comprised almost entirely of unpreferred codons, and a second comprised almost entirely of preferred codons. We monitored the fitness effects of these synonymous mutations over hundreds of generations in two selective environments that allowed us to disentangle translational effects acting in *cis* from those actin in *trans*. In a selective environment for maximizing translational efficiency in *trans* of a gene (*tetA*) encoding a tetracycline resistance protein, unpreferred synonymous mutations had a negative impact on long-term fitness, whereas preferred mutations had a positive impact on long-term fitness, providing strong experimental evidence for a pleiotropic effect of translational selection.

Open Symposium
SMBE-PO-498
Evolution of sex-determining systems in genus Silene, section Otites
R. Gogela*, V. Balounova, R. Cegan, J. Zluvova, B. Vyskot, R. Hobza, B. Janousek

Abstract: While gonochorism is common in animals, flowering plants are mostly hermaphrodites. Only 6 % of angiosperms are dioecious (male and female individuals present in population).
Plant genus *Silene* offers great opportunities to study evolution of sex-determining systems and evolution of sex chromosomes. Dioecy evolved independently in two sections of this genus and the sex chromosomes have evolved from different pairs of autosomes. Our previous research has shown a switch from female heterogamety (ZW sex-determining system) to male heterogamety (XY sex-determining system) in the section Otites. Our analyses also suggest a possibility that has so far not been considered, change in heterogamety through hybridization, in which a male-determining chromosome from one species is introgressed into another one, and over-rides its previous sex-determining system.
The genus *Silene* is rich in gynodioecious species and so the dioecy has likely repeatedly evolved via gynodioecy pathway in this genus. We have currently focused on the study in *Silene sibirica* (gynodioecious relative of section Otites). We study the region including locus for restoration of male fertility. RNA-seq analysis was used for genetic mapping of chromosome carrying restorer of male fertility in *Silene sibirica*. Identification of putative male fertility restorer proceeds both via analysis of transcriptomic data and via BAC library screening. We hope that we will find connections between genetic control in gynodioecious and dioecious systems in the genus *Silene*. This information can shed light on the evolution via gynodioecious pathway in general.
This research was supported by the Czech Science Agency grant no.: 18-06147S.

Open Symposium
SMBE-PO-494
Possible European origin of circulating Varicella-zoster virus strains
C. Pontremoli*, D. Forni, M. Clerici, R. Cagliani, M. Sironi

Abstract: Varicella-zoster virus (VZV) is the causative agent of chickenpox and shingles. The geographic distribution of VZV clades was taken as evidence that VZV migrated out-of-Africa with human populations. We show that extant VZV strains most likely originated in Europe and not in Africa. Europe was also identified as the ancestral location for most internal nodes of the VZV phylogeny, including the ancestor of clade 5 strains. We also show that strains from clades 1, 2, 3, and 5 derived a major proportion of their ancestry from each of four ancestral populations. Conversely, viruses from other clades displayed variable levels of admixture. Some low-level admixture was also observed for clade 5 genomes, but only for non-African viruses. This pattern indicates that the clade 5 VZV strains do not represent recent introductions from Africa due to migratory fluxes. These data have also relevance for the definition and classification of VZV clades.

Open Symposium

SMBE-PO-500

Confirmation that genes retained following whole genome duplication and tandem duplication show differences across a wide range of features

Z. Vance [1,*], A. McLysaght [1]

[1]Trinity College Dublin, Dublin, Ireland

Abstract: Gene duplication has long been known to be an important process in the evolution of genome structure and creation of new genetic material. Duplication may occur by one of two major mechanisms; tandem duplication or whole genome duplication. The sets of duplicates retained following these events differ greatly, forming largely non-overlapping sets. Many differences have been noted between these gene sets in function, basic gene features such as genomic length and more complex properties such as how gene expression is regulated. However, the current body of work on this topic lacks a cohesive narrative regarding what features distinguish a gene which is retained following whole genome duplication (ohnolog) from a gene which may duplicate by tandem duplication. Current studies in the literature are inconsistent in study species and definitions of the gene sets to be compared. Here we show, by comparing a large number of features in consistently defined sets of duplicable genes in human, that there is a general trend of higher constraint in ohnologs, and of lower constraint in tandem duplicated genes, relative to singletons. Ohnologs are generally more complex, slower evolving and more heavily regulated with tandem duplicates showing the opposite pattern. Additionally, we confirm previously reported trends for enriched functions in both gene sets; ohnologs are enriched for developmental, nervous system and regulatory functions while tandem duplicates show enrichment for functions associated with environmental triggers such as immune and sensory functions. As these comparisons are performed within a single species, using a consistent definition for the categories compared, they serve as a useful starting point to answer more complex questions about the nature of gene duplication. Comparisons which are consistent in this way will allow for comparison of the importance of various features in distinguishing the two categories of duplicates. Whereas up to now we have only been able to say that genes associated with the two modes of duplication are extremely different, we may now begin to determine which of these differences are important in defining the biases which shape these patterns of gene duplication. The results presented here will be highly relevant in answering such questions.

Open Symposium

SMBE-PO-503

Inferring the model and onset of natural selection under varying population size from the site frequency spectrum and haplotype structure

S. Nakagome [1,*], R. Hudson [2], A. Di Rienzo [3]

[1]School of Medicine, Trinity College Dublin, Dublin, Ireland, [2]Department of Ecology & Evolution, [3]Department of Human Genetics, University of Chicago, Chicago, United States

Abstract: A fundamental question about adaptation in a population is the time of onset of the selective pressure acting on beneficial alleles. Inferring this time, in turn, depends on the selection model. We develop a framework of approximate Bayesian computation (ABC) that enables the use of the full site frequency spectrum and haplotype structure to test the goodness-of-fit of selection models and estimate the timing of selection under varying population size scenarios. We show that our method has sufficient power to distinguish natural selection from neutrality even if relatively old selection increased the frequency of a pre-existing allele from 20% to 50% or from 40% to 80%. Our ABC can accurately estimate the time of onset of selection on a new mutation. However, estimates are prone to bias under the standing variation model, possibly due to the uncertainty in the allele frequency at the onset of selection. We further extend our approach to take advantage of ancient DNA data that provides information on the allele frequency path of the beneficial allele. Applying our ABC, including both modern and ancient human DNA data, to four pigmentation alleles in Europeans, we detected selection on standing variants that occurred after the dispersal from Africa even though models of selection on a new mutation were initially supported for two of these alleles without the ancient data.

Open Symposium
SMBE-PO-495
A new approach to infer orthologous genes directly from genome alignments
B. Kirilenko [1 2 3,*], M. Hiller [1 2 3]
[1]Max Planck Institute of Molecular Cell Biology and Genetics, [2]Center for Systems Biology Dresden, [3]Max Planck Institute for the Physics of Complex Systems, Dresden, Germany

Abstract: Annotating genes in newly sequenced genomes and distinguishing orthologous from paralogous genes is a key prerequisite for a variety of molecular evolution analyses. Existing methods infer orthologous genes mostly based on reciprocal-best BLAST analysis or computing gene trees. Here, we present a new approach to infer orthologous genes given a newly sequenced genome. Our approach makes use of pairwise whole genome alignments between a well-annotated reference species (such as human) and another query species (such as other mammals). Our approach utilizes the entire genomic context to (i) directly annotate coding genes, (ii) distinguish genomic loci containing orthologs and paralogs with a high degree of precision, and (iii) classifying orthologs in one-to-one, one-to-many, many-to-one and many-to-many groups. We show that this approach can resolve difficult cases of highly similar genes and is thus complementary to existing methods. We present results and comparisons of applying our genome alignment-based approach to annotate genes and infer orthologs and co-orthologs in the genomes of numerous mammals.

Open Symposium
SMBE-PO-499
Update of the Animal rDNA database
J. Sochorová [1,*], S. Garcia [2], F. Gálvez [3], R. Symonová [4], A. Kovařík [1]
[1]Institute of Biophysics of the CAS, v. v. i., Brno, Czech Republic, [2]Institut Botànic de Barcelona , Barcelona, [3]Bioscripts - Centro de Investigación y Desarrollo de Recursos Científicos, Sevilla, Spain, [4]University of Innsbruck, Mondsee, Austria

Abstract: Ribosomal DNA (rDNA) loci encoding 5S and 45S (18S-5.8S-26S) rRNA are important components of eukaryotic chromosomes varying both in numbers and locations. Accumulation of rDNA data caused the need of storing and analysing the information in the database. The collected data are based on in situ hybridization studies (mostly flourescence-based (FISH)) (in more than 780 papers) of metaphase chromosomes carried out in major groups of vertebrates (fish, reptiles, amphibians and mammals), invertebrates (arthropods and molluscs), minor represented groups (birds) and other groups with just few representatives. We statistically analyzed the numbers and positions of rDNA loci on different types of chromosomes in more than 2020 animal species (over 340 families). Fishes have the greatest variability in the number of 5S and 45S rDNA loci (1 to 27 5S sites/ 1C and 1 to 27 45S sites/1C). rDNA loci can occur on any chromosome including sex chromosomes (X, Y, Z and W), supernumerary B chromosomes and microchromosomes. All mammal karyotypes contain sex chromosomes, in fish karyotypes the sex chromosomes are rare. Nevertheless the proportion of rDNA localization on autosomes and sex chromosomes in these groups is the similar. In a few karyotypes (16 arthropods, two mammals and one reptile) rDNA is localized on X chromosome only (0.009% of karyotypes). Microchromosomes are present in bird (100%), reptile (93%) and a few amphibian karyotypes (4%) in the database. In birds, reptiles and amphibians rDNA is localized on microchromosomes at 89%, 36% and 4% karyotypes respectively. There was no significant correlation between the variability of rDNA loci and the age of the orders of Arthropods and between the number of 5S a 45S loci in all groups. The Animal rDNA database is accessible via a web-based interphase at http://www.animalrdnadatabase.com/.

Open Symposium
SMBE-PO-428
Comparative population genomics of transposable elements in two vertebrates
Y. Bourgeois [1,*], S. Boissinot [1]
[1]NYU Abu Dhabi, Abu Dhabi, United Arab Emirates

Abstract: Transposable elements (TEs) play important roles in shaping genomes organization and structure, and may cause dramatic changes in phenotypes. Despite the genetic load they may cause and their importance in microevolutionary processes such as adaptation and speciation, the number of population genetic studies having focused on TEs has been rather limited so far compared to single nucleotide polymorphisms (SNPs). Here, we aim at answering the following questions: i) how often do TEs escape purifying selection and provide a selective advantage; ii) what is the impact of hybridization and secondary contact on TE diversity; iii) do TE insertions facilitate the emergence of reproductive isolation between species. To address these questions, we focus on two vertebrate models, the house mouse *Mus musculus*, and the green anole *Anolis carolinensis.* These two species strongly differ in their TE content and diversity, and have diversified into multiple genetic clusters and subspecies spanning various environments. Using dozens of wild accessions sampled across the species ranges, we assess the interplay between selection and drift in determining the probability of TEs fixation. Using both TE frequencies and the information given by flanking SNPs, we detect candidate TEs for disruptive selection and resisting introgression, taking into account demographic history and intrinsic properties of genomes such as variable recombination rates. By comparing patterns of TE evolution between a mammal and a squamate, we bridge micro- and macro-evolutionary approaches to get a broader understanding of TE dynamics in vertebrates.

Open Symposium

SMBE-PO-504

Epistatic interactions in protein sites that contribute to H5N1 virulence in mammals

A. Neverov [1], K. Safina [2 3,*], G. Bazykin [2 3]

[1]Department of Molecular Diagnostics, Central Research Institute for Epidemiology, Moscow, [2]Life Sciences, Skolkovo Institute of Science and Technology, Skolkovo, [3]Institute for Information Transmission Problems RAS, Moscow, Russian Federation

Abstract: H5N1 is a highly pathogenic avian influenza virus; however, some strains of H5N1 are able to efficiently infect and cause death in mammals. Since 1997, several hundred confirmed cases of human infection were reported, with a mortality rate higher than 50%. H5N1 virulence is mainly determined by two viral proteins, HA (hemagglutinin) and PB2 (polymerase subunit). We developed a method that infers direct epistatic interactions between protein sites by analyzing the distribution of mutations on a phylogenetic tree and used it to study epistasis in H5N1 hemagglutinin. Pairs of sites involved in strong positive epistatic interactions (when a mutation in one site provokes mutation in another) are enriched with protein positions associated with H5N1 virulence in mammals. Our results imply that mutations in these pairs of sites may alleviate adaptation of the virus to a mammalian host and should be treated carefully and monitored in H5N1 population.

Open Symposium

SMBE-PO-432

Expression changes of structural protein genes that may be related to adaptive human skin characteristics

N. Arakawa [1,*], D. Utsumi [2], K. Takahashi [3], A. Matsumoto-Oda [4], A. Nyachieo [5], D. Chai [5], N. Jillani [5], H. Imai [6], Y. Satta [1], Y. Terai [1]

[1]Department of Evolutionary Studies of Biosystems, SOKENDAI (The Graduate University for Advanced Studies), Hayama, [2]Department of Dermatology, [3]University of the Ryukyus, Nishihara, Japan, [4]Graduate School of Tourism Sciences, University of the Ryukyus, Nishihara, Japan, [5]Institute of Primate Research, National Museum of Kenya, Nairobi, Kenya, [6]Primate Research Institute, Kyoto University, Inuyama, Japan

Abstract: Human skin is morphologically and physiologically different from the skin of other primates. The aim of this study is to understand genetic causes underlying human-specific skin characteristics. First, we quantitatively demonstrated that the epidermis and dermis of skin were significantly thicker in humans than in three Old World monkey species examined. We also indicated that the epidermal basement membrane (BM) zone topography was undulating in humans, which is known as a rete ridge, but was flat in the Old World monkey species examined. Second, we comprehensively compared gene expression levels between human and non-human great ape skin using next-generation cDNA sequencing (RNA-seq). We found that four genes encoding proteins that constitute the epidermal BM zone or elastic fibers in the dermis (*COL18A1*, *LAMB2*, *CD151*, and *BGN*) were expressed significantly greater in humans than in non-human great apes, suggesting that these differences may be related to the rete ridge and rich elastic fibers present in human skin. The rete ridge may enhance the intensity of adhesion between the epidermis and dermis in skin. This ridge, together with a thick epidermis and rich elastic fibers, might contribute to the physical strength of human skin with a low amount of hair. Finally, to estimate transcriptional regulatory regions for the four structural protein genes identified, we examined conserved noncoding regions with histone modifications for active regulatory regions in skin cells. Human-specific substitutions in these regions, especially those located in transcription factor binding sites, may alter the gene expression patterns and give rise to the human-specific adaptive skin characteristics.

Open Symposium

SMBE-PO-430

A Bayesian method to detect targets of selection in Evolve-and-Resequence experiments

R. Borges [1,*], C. Barata [2], C. Kosiol [2]

[1]Vetmeduni Vienna, Vienna, Austria, [2]St Andrews University, St Andrews, United Kingdom

Abstract: New approaches combining experimental evolution with high-throughput sequencing techniques, allow testing for selection on populations that are subjected to a given selective pressure over time. These approaches are referred to as Evolve-and-Resequence experiments, allowing to disentangle the architecture of the adaptive process. Nevertheless, more efficient and statistically sound methods are needed to detect the true targets of selection. Here, we present a fully Bayesian approach based on the Moran model of allele evolution to estimate selection coefficients from such experiments. The model is characterized by overlapping generations, allowing to describe alternative experimental designs. We also propose a new hypothesis testing approach to detect selected alleles that avoids the computational burden of simulating the empirical null distribution. We tested our method for several demographic and experimental conditions to assess its accuracy and precision. Our method performs well in most of these scenarios, but some care must be taken with specific allele trajectories, i.e. small effective population size (where drift largely dominates) and low starting frequencies. We compare our method with existing ones and report that ours has generally higher accuracy even for such difficult trajectories. Furthermore, our approach outperforms available software in terms of computational time, which permits its use genome-wide.

Open Symposium

SMBE-PO-515

PhyloProfile: Dynamic exploration of enriched phylogenetic profiles in functional and evolutionary studies

N.-V. Tran [1,*], I. Ebersberger [1 2 3]

[1]Dept. for Applied Bioinformatics, Goethe University Frankfurt, [2]Senckenberg Biodiversity and Climate Research Centre (BiK-F), [3]LOEWE Centre for Translational Biodiversity Genomics (TBG), Frankfurt am Main, Germany

Abstract: The presence-absence pattern of genes across species is an essential information for a wide range of comparative studies. This is even more so, since these phylogenetic profiles become increasingly enriched with accessory information, such as the domain architecture similarity between proteins. This data forms the basis for phylogenetic and phylogenomic studies, protein annotation transfer, and for studying the evolutionary history of molecular function. Yet, software for a dynamic and interactive analysis of such data is surprisingly scarce, and ad hoc scripting solutions prevail.

We present PhyloProfile, an R package that aims at closing this methodological gap. The user interface of PhyloProfile allows to interactively visualize and explore enriched phylogenetic profiles. Users can adapt the level of resolution from overviewing the full profile, where taxa can be optionally summarized in higher order systematic groups, down to the pair-wise comparison of domain architectures. For data analysis, PhyloProfile provides a comprehensive toolbox. Various filters help to remove spurious homologs, genes can be clustered according to their profiles, core gene sets can be inferred for a custom taxon selection. To extend the discovery of genetic innovation beyond the canonical gene gain/loss analysis, PhyloProfile can identify genes with a lineage-specific change in domain architecture, which indicates a change in protein function. Various data export options facilitate then a seamless integration with downstream analysis. We will demonstrate the application of PhyloProfile on the example of an evolutionary and functional characterization of the Microsporida, a group of highly reduced obligate intracellular parasites.

PhyloProfile is available at (https://github.com/BIONF/PhyloProfile).

Open Symposium

SMBE-PO-496

Plant Genus Silene as a group of models in genomic, evolutionary and developmental studies

B. Janousek [1,*], R. Gogela [1], V. Balounova [1], J. Zluvova [1], V. Hudzieczek [1], Z. Kubat [1], E. Kejnovsky [1], R. Cegan [1], R. Hobza [1]
[1]DPDG, Institute of Biophysics AS CR, v.v.i., Brno, Czech Republic

Abstract: Dioecious species of the genus Silene served as one of the first models for the studies of the sex chromosomes and sex determination. The conspicuous heteromorphy of the sex chromosomes in the section Melandrium attained the interest of researchers aiming to isolate sex linked and determining genes via methods which combined cytologenetic and molecular approaches. These studies enabled more advanced description of the particular sex determining regions of the Y chromosome of S. latifolia but the isolation of sex determining genes was hindered by large size of the Y-linked non-recombining region and big size of the nuclear genome.

Albeit the large size of non-recombining region complicates genetic analyses it also brings opportunities to the study the topics as are, e.g., the mechanisms involved in the sex specific suppression replication of retrotransposons and evolution of the repetitive sequences in context of sex determination. Comparison of the evolution of the systems with large and small sex determining regions is enabled by the existence of the section Otites of the genus Silene that contains solely dioecious species with small determining regions. The further opportunities for comparisons are brought by the fact that both ZW and XY sex determining systems occur in the section Otites. Apart of the contribution to understanding sex chromosome evolution our analyses also bring insight to topics that are also studied using species of the genus Silene as is, e.g., genome dynamics in the context of the different reproductive strategies, evolutionary aspects in heavy metal tolerance and search for sex determining genes**.** The functional studies of the candidate genes putatively involved in the sex determination have been also recently enabled as the protocol for the genetic transformation and targeted inactivation of chosen genes has been developed.

This research was supported by the Czech Science Agency grant no.: 18-06147S.

Open Symposium
SMBE-PO-493
You will never walk alone: co-dispersal of JC polyomavirus with human populations
D. Forni*, R. Cagliani, U. Pozzoli, M. Clerici, M. Sironi

Abstract: JC polyomavirus (JCPyV) is one of the most prevalent human viruses. Findings based on the geographic distribution of viral subtypes suggested that JCPyV codiverged with human populations. This view was however challenged by data reporting a much more recent origin and expansion of JCPyV. We collected information on ~1,100 worldwide strains and we show that their geographic distribution roughly corresponds to major human migratory routes. Bayesian phylogeographic analysis inferred a sub-Saharan origin for JCPyV, although with low posterior probability. High confidence inference at internal nodes provided strong support for a long-standing association between the virus and human populations. In line with these data, pairwise F_{ST} values for JCPyV and human mtDNA sampled from the same areas showed a positive and significant correlation. Likewise, a very strong relationship ($R^{2=0.99)}$ was found when nuclear-marker based F_{ST} between human populations was correlated with node ages in the JCPyV phylogeny. Reconciliation analysis detected a significant cophylogenetic signal for the human population and JCPyV trees. Notably, JCPyV also traced some relatively recent migration events such as the expansion of Lapita people (Philippines/Taiwan area) into Remote Oceania, the gene flow between North-Eastern Siberian and Ainus, and the Koryak contribution to Circum-Arctic Americans. Finally, two different molecular dating approaches dated the origin of JCPyV in a time frame that precedes human out-of-Africa migration. We thus conclude that JCPyV infected humans in Africa and accompanied our species during worldwide dispersal. JCPyV typing can provide reliable geographic information with possible application to forensic specimens or ancient biological remains.

Open Symposium

SMBE-PO-492

Rapid evolution of the telomeric repeat units and contents in hymenopteran parasitoids

X. Wang [1 2,*], Y. Zhou [1 3], Y. Wang [4], X. Xiong [1 3], A. Appel [5], C. Zhang [3]

[1]Pathobiology, Auburn University, Auburn,AL, [2]HudsonAlpha Institute for Biotechnology, Huntsville,AL, United States, [3]Life Sciences & Technology, Tongji University, [4]Anthropology & Human Genetics, Fudan University, Shanghai, China, [5]Entomology & plant pathology, Auburn University, Auburn,AL, United States

Abstract: Telomere is an important structure of the chromosome containing a region of repetitive sequences at both ends of a chromosome. Telomeres are essential to maintain the linear chromosome structure because they prevent the two chromosome ends from deterioration or fusion with other chromosomes. With a few exceptions, animal telomeric sequences are generally short tandom repeats of 5 or 6 bp G-rich repeat units. Vertebrate telomeres consist of hexanucleotide tandom repeats (TTAGGG)n, whereas 5 bp telomeric repeats (TTAGG)n were found in most insect species examined. In hymenopterans, (TTAGG)n was characterized as the telomeric repeats in ants and bees, but the telomeric sequence content of parasitoid wasps, which account for more than half of this order, is still unknown. To explore the evolutionary dynamics of telomeric repeats in insects, we developed a pipeline called RepeatMaster for fast identification of short random repeats in next-generation genome sequencing data, and applied to six parasitoid jewel wasp species we sequenced. Surprisingly, we discovered an 8 bp tandom repeats in all six species using bioinformatic methods and validated them by qPCR. FISH experiments confirmed the location of these repeats at the chromosome termini. Despite the critical function and high level of conservation, the telomeric repeat unit sequence we found in jewel wasp is longer than the other hymenopteran insects (5 bp) or vertebrate animals (6 bp) or plants (7 bp, TTTAGGG). To investigate whether the expansion of telomeric repeat unit is parasitoids specific, we examined other parasitoid groups including braconid wasps, fig wasps, jumping wasps and trichogramma wasps. Most parasitoids have the same 8 bp telomeric repeat unit as jewels, and a few species have 9 bp variation due to 1 bp addition. Free-living hymenopteran outgroups, wood wasps and sawflies, both have the classic 5 bp insect type telomeric repeats (TTAGG)n. We sequenced the genome of silverfish to represent a basal insect group and (TTAGG)n was identified as the telomeric repeats, suggesting the longer telomeric repeat units are restricted in parasitoids. We quantified the abundance of telomeric repeats in four jewel wasp species and found up to 8-fold differences among closed related species diverged 1 million years ago. After polarized with an outgroup species, the observation is consistent with lineage specific expansion of telomeric repeat contents. These finding indicate rapid evolution of telomeric repeat unit sequence and total abundance in parasitoid hymenopterans. The elongated telomeric repeat unit may play a role in the parasitoid life history, to improve the stability of their telomeres in the host by diverging from the host telomeric sequences. Further studies are needed to elucidate the biological function and fitness consequences of the longer telomeric tandem repeat unit in parasitoid hymenopterans. Our research will shed light on the telomere repeat variation in insects and the evolutionary dynamics of telomeres.

Open Symposium

SMBE-PO-509

SEXUAL SELECTION IN OUTCROSSING AND SELFING POPULATIONS OF Arabidopsis lyrata: A PHENOTYPIC AND TRANSCRIPTOMIC ANALYSIS

G. Chimetto [1,*], B. Albert [2], S. Campbell [3], R. Schmickl [1], J. Leppälä [4], A. Cornille [5], C. Lafon Placette [1]

[1]Department of Botany, Charles University, Prague, Czech Republic, [2]Laboratoire Ecologie, Systématique et Evolution, Paris-sud University, Orsay, France, [3]Department of Animal and Plant Sciences, Sheffield University, Sheffield, United Kingdom, [4]Department of Ecology and Environmental Sciences, Umea University, Umea, Sweden, [5]Génétique Quantitative et Évolution - Le Moulon, INRA - Université Paris-Sud - CNRS – AgroParisTech, Gif-sur-Yvette, France

Abstract: *Arabidopsis lyrata* is a perfect candidate for studying the influence of sexual selection in angiosperms evolution, given the existence of separate populations reproducing with different mating systems: outcrossing (self-incompatible) or selfing (self-compatible). Given the higher variability and the higher chances of pollen competition in outcrossing populations compared to selfers, the former are expected to experience a stronger level of sexual selection, while in the latter the selective pressure should be relaxed. Performing crosses within and between populations of *A. lyrata* can therefore help to draw interesting results about the importance of sexual selection in shaping angiosperms evolution of reproductive traits. We tested for differences between populations in terms of several traits that can play a role in sexual selection: pollen tube growth rate, pollen germination rate and speed, stigma receptivity, size and amount of pollen produced, number of ovules and flower traits such as pistil length and petal size. We also performed crosses within and between populations to assess fertilization ability and multiple donor crosses to test for competition in terms of paternity success. We also performed differential expression analysis of sexual specific tissues, compared with non-sexual tissues, to identify genes that may be possible candidates for the action of sexual selection and to assess whether those genes are expressed at different levels in outcrossing and selfing populations.

Open Symposium
SMBE-PO-501
RADProc: A computationally efficient de novo locus assembler for population studies using RADseq data

P. Nadukkalam Ravindran [1,*], P. Bentzen [2], I. R. Bradbury [3], R. G. Beiko [1]
[1]Computer science, [2]Biology, Dalhousie university, Halifax, [3]Department of Fisheries and Ocean, St. John's, Canada

Abstract: Restriction-site associated DNA sequencing (RADseq) is a powerful tool for genotyping of individuals, but the inference of loci and population structure can be highly sensitive to the parameters used for inference, and the best parameter combinations vary from dataset to dataset. A common practice is to run parameter sweeps and choose the best result according to some criterion, but recalculating loci many times from scratch is a computationally expensive and inefficient process. Here we introduce RADProc, a software package that implements a three-step solution, which includes identifying and filtering out comparatively less abundant sequences across all individuals in the data set at an early stage in the process, a clustering approach for faster nucleotide-distance calculation, and a graph data structure to represent all sequence reads and their similarity relationships. Storing sequence-comparison results in a graph eliminates unnecessary and redundant sequence-similarity calculations making parameter sweeps far more efficient. RADProc required 2 hours 40 minutes to infer loci from 32 different parameter settings for 20 green crab (*Carcinus maenas*) samples, as compared to 78 hours with the widely used Stacks software, while 16 brown trout (*Salmo trutta* L.) samples were processed by RADProc and Stacks in 23 hours and 263 hours respectively. Comparisons of the *de novo* loci formed and catalog built using both the methods demonstrate that the improvement in processing speeds achieved by RADProc negligible effect on the actual loci formed and the results of downstream analyses based on those loci.

Open Symposium

SMBE-PO-474

Identifying species and reconstructing the phylogeny of Botryllinae (Tunicata, Ascidiacea) using an elongated COI fragment: advantages and pitfalls

M. Salonna [1,*], F. Montesanto [2], F. Mastrototaro [2], F. Gasparini [3], C. Gissi [1 4]

[1]Dept of Biosciences, Biotechnology and Biopharmaceutics, [2]Dept of Biology and CoNISMa LRU, University of Bari, Via Orabona, 4 - 70125, Bari, [3]Dept of Biology, University of Padova, Via Ugo Bassi 58/B – 35131 , Padova, [4]IBIOM, Istituto di Biomembrane, Bioenergetica e Biotecnologie Molecolari, CNR, Via Giovanni Amendola 165/A - 70126 , Bari, Italy

Abstract: The subfamily Botryllinae comprises small colonial ascidians of the two genera *Botryllus* and *Botrylloides,* whose morphological identification at species level is very problematic due to the high phenotype variability and the few hardly-visible discriminant characters. Indeed, even in the model species *Botryllus schlosseri*, the existence of cryptic species was suggested by molecular data and not yet confirmed by morphological analyses. In addition, the Botryllinae phylogeny is also poorly resolved.

This study aims at evaluating the ability of the COI mitochondrial gene both in species identification and phylogenetic reconstructions within Botryllinae. Using two newly designed degenerate primer pairs, we set up a nested PCR strategy able to amplify an elongated COI fragment of about 860 bp in Botryllinae. Thus, we successfully analysed more than 120 *Botryllus* and *Botrylloides* worldwide-distributed colonies, mainly with uncertain or unknown morphological identification.

The phylogenetic reconstructions identified the clades of all known species as statistically significant and effectively resolved the relationships among the cryptic species of *B. schlosseri*. However, they left unresolved the basal nodes of the tree, i.e., the main relationships among Botryllinae. Moreover, species delimitation analyses showed the existence of a clear barcode gap, easily discriminating among already-described, cryptic and putative new species.

Our data indicate that the elongated COI fragment can be used to systematically complement morphological analyses in Botryllinae. Indeed, it was amplified in all analysed Botryllinae, is a DNA barcode as reliable as the common 600 bp COI fragment, and is a good marker for resolving the phylogenetic relationships among cryptic species.

Open Symposium

SMBE-PO-478

Coordinated evolution of large DNA fragments in the chromatin interaction network of Arabidopsis thaliana

Y. Yan [1,*], Z. Li [1], Y. Li [1], R. Yang [1]

[1]College of Life Sciences, Northwest A&F University, Yangling, China

Abstract: In eukaryotes, the organization of genomes is not random at both the linear (1D) and three-dimensional (3D) scales. Previous studies have shown that neighbouring genes in 1D or 3D space tend to have more similar functions and expression patterns than random gene pairs. It is known that gene function and expression level/breadth are two important determinants of evolutionary rate, thus, genes in close proximity at the 3D scale possibly evolve at similar rates, and the coordinated evolution may be extended to noncoding regions. To test this hypothesis, we construct a chromatin interaction network (CIN) in *Arabidopsis thaliana* based on high-throughput chromosome conformation capture (Hi-C) data, and demonstrate that adjacent large DNA fragments in the CIN indeed have more similar nucleotide diversity and evolutionary rate than random bin pairs. Simulations considering the linear distance between segments suggest that this pattern is not merely the consequence of chromatin organization at 1D level. Mutation rate measured as synonymous divergence and natural selection determined using various estimates also present a higher similarity between direct neighbouring Hi-C fragments than the random expectation, indicating that the coordinated evolution of 3D neighbours is a result of combined evolutionary forces. As in other biological networks, the connectivity of nodes in the CIN is negatively correlated with the overall genetic diversity of their residing DNA, suggesting that highly connected DNA fragments subject to more constraints. We further investigated the roles of 39 genomic and epigenomic features in determining the evolution of large chromatin fragments. The results show that these features can account for 54%, 68%, and 47% variation of nucleotide diversity, divergence, and mutation rate of 100kb fragments in the *Arabidopsis* genome, respectively, whereas only 17% variation of selective efficacy can be explained by them. Fragments with low mutation rate and evolutionary rate are usually associated with open and transcription active chromatin, while high rates of mutation and evolution are expected to be found in regions with condensed chromatin and low transcription activity, which suggests that epigenetic modifications play important roles not only in the maintenance of genome structure, but also in the evolution of related genetic sequences. Taken together, our results support that large DNA fragments that are close to each other in the 3D genome of *A. thaliana* are likely to evolve at similar paces, and both mutation rate and natural selection contribute to the observed coordinated evolution of chromatin neighbours. In addition, chromatin architecture possibly promotes the coordinated evolution mainly through its influence on regional mutation rate.

Open Symposium

SMBE-PO-480

Speciation and transposable element amplification in the Ethiopian frog species complex Ptychadena neumanni

I. E. Hariyani [1,*], Y. Bourgeois [1], J. Reyes-Velasco [1], S. Boissinot [1]

[1]Science Division, New York University Abu Dhabi, Abu Dhabi, United Arab Emirates

Abstract: Eukaryotic genomes contain many families of transposable elements (TEs) that exhibit high variation in diversity and copy number. The dynamics of TEs in genomes is strongly affected by the evolution and demography of their host. For instance, in populations of small size, genetic drift opposes the effect of negative selection and can result in the fixation of TE insertions that would be eliminated in large populations. A correlation between the amplification of TE families and speciation events has been proposed based on observed differences in abundance and diversity of transposable elements across species. The carrier subpopulation (CASP) hypothesis predicts that population subdivision into small subpopulations may trigger fixation of TE families by genetic drift. Following the hypothesis, we would expect TEs to accumulate in the genome during punctuated episodes of speciation, and near the nodes of phylogenies, rather than on long branches. We used the Ethiopian *Ptychadena neumanni* species complex to test this hypothesis and observe any effects of speciation on TE amplification. The *P. neumanni* species complex serves as a good model to study this effect as there are several points of comparison in the phylogeny due to its bushy radiation. Using a high-quality genome sequence for *P. neumanni 2*, we identified 2342 repeat families *de novo* using RepeatModeler and characterized the repeat landscape of the frog using RepeatMasker. Resequencing data from six other closely related frog species were used to identify mobile element insertion polymorphisms with reference to the *P. neumanni 2* genome sequence. In parallel, a variant calling workflow helped identify thousands of genome-wide SNPs, which were used to construct a phylogeny by mapping the TE polymorphisms. Our analysis revealed that the amplification of TEs is consistent with the evolutionary history of the host inferred by SNPs. We also observed substantial differences in the rate of amplification of TEs among species.

Open Symposium

SMBE-PO-483

Genetic analysis of meiotic recombination in two Drosophila species using their F1 hybrid

A. Inuyama [1], Y. Ogawa [1], M. Nozawa [1 2], K. Tamura [1 2,*]

[1]Department of Biological Sciences, [2]Research Center for Genomics and Bioinformatics, Tokyo Metropolitan University, Tokyo, Japan

Abstract: Meiotic recombination is believed to have merits for organisms with sexual reproduction. Meiotic recombination produces genetic diversity among chromosomes by exchanging parts between homologous chromosomes. Meiotic recombination also enhances the removal of deleterious mutations and supports chromosomal disjunctions. On the other hands, the rate of meiotic recombination is known to be different between sexes in many species. This phenomenon is called heterochiasmy. In the extreme case, meiotic recombination does not occur at all in one sex, which is called achiasmy. Heterochiasmy and achiasmy are expected to reduce the merits of meiotic recombination. Nevertheless, they are widely observed in many species, which indicates other merits of no recombination.

The mode of meiotic recombination has been well studied in *Drosophila* species. Recently, Satomura and Tamura (2016) found that meiotic recombination occurs only in females in *Drosophila albomicans* as in the case of many other *Drosophila* species, whereas meiotic recombination occurs in males as well as females in its sibling species, *D. nasuta*. They also suggested that the male recombination occurred in their common ancestor, analyzing the nucleotide sequences of many protein-coding genes in detail at a population level.

In this study, focusing on the biological implication of the male recombination in *D. nasuta*, we examined the rate of meiotic recombination in the F_1 hybrid of *D. albomicans* and *D. nasuta*. As the results, we found that meiotic recombination does not occur in the F_1 males as in the case of *D. albomicans*, suggesting that genes of *D. albomicans* dominantly suppress male recombination in males. According to Satomura and Tamura (2016), the female recombination rate on the second chromosome was much lower than that on the third chromosome in *D. albomicans* whereas the tendency was not clear in *D. nasuta*. In this study, we found that the female recombination rate in the F_1 hybrid was much lower on the second chromosome, which was similar to the pattern observed in *D. albomicans*. This demonstrates that the *D. albomicans* genes are dominant in regulating the female meiotic recombination. Furthermore, we observed that the male F_1 hybrid produced progenies with XXY sex chromosomal genotype via the non-disjunction of sex chromosomes at a high frequency. The mechanism of the proper disjunction of X and Y chromosomes in *D. nasuta* seemed to be disrupted by the genes from *D. albomicans* that does not have ordinary sex chromosomes but neo-sex chromosomes.

Open Symposium
SMBE-PO-491
Inference of population genetic parameters from continuously serial-sampled data, applied to human seasonal influenza A/H3N2
M. F. Croze [1,*], Y. Kim [1]
[1]EcoScience, Ewha Womans University, Seoul, Korea, Republic Of

Abstract: Understanding how influenza viruses evolve is the key to analyzing and predicting the epidemics of human seasonal influenza and the emergence of new pandemic viruses such as avian influenza. The subtype H3N2 of type A has been the most common cause of human seasonal influenza and also evolutionary genetically the best understood system due to a large number of DNA sequences sampled over 40 years. This subtype is characterized by extinction and re-colonization of viral population in a given location, over different geographic regions. While we aim to investigate this complex demography, the “cactus-like” shape of genealogy constructed for continuously serial-sampled viruses makes it difficult to apply standard methods for inferring population genetic parameters, because sequence difference for any pair of sampled viruses reflects not only the genealogical process but also sampling time difference. We therefore developed the following approaches. The mutation rate and effective population size were inferred from the line of regression of pairwise sequence differences on sampling time differences. This approach revealed that there are significant differences in mutation rates and effective population sizes among viral segments. We also transformed the shape of genealogy into mismatch distribution that takes into account the sampling time differences and the mutation rate estimated previously to correct genetic differences within pairs of viral sequences. The correlation in mismatch distribution between the viral sequence data and simulated data with varying evolutionary parameters (migration, selection) is investigated to find the best population model of human seasonal influenza viruses A/H3N2.

Open Symposium

SMBE-PO-490

Bayesian Inference of Joint Coalescence Times and the Associated Ancestral Process Distributions from Sampled Sequences

H. Simon [1,*]

[1]Research School of Biology, Australian National University, Canberra, Australia

Abstract:

Rapid advances in the availability of genomic data and in computational power have enabled exciting advances in such fields as tracing human prehistory and identifying the origins both of genetic diseases and conditions that may confer immunity. Such advances are dependent on the ability to infer the past history of a population from the pattern of genetic variation in contemporary populations, but the extent to which the resolution of inference of past events is necessarily obscured by the inherently stochastic nature of the processes involved (primarily mutation) has not been fully clarified by previously proposed methods.

The site frequency spectrum (SFS) is a commonly used statistic to summarise genetic variation in a population sample. A Bayesian model for sampling from the joint distribution of coalescence times conditional on the SFS associated with a sample of sequences is presented. A novel approach is used to express the relevant structural properties of a coalescent tree. The Bayesian approach allows the posterior distribution of tree topology to be "integrated out", hence assumptions as to the type of stochastic process generating the genealogical trees and the associated coalescence times are not required. The uncertainty in inference is indicated by the shape of the posterior distributions. The method is implemented using the general purpose Markov Chain Monte Carlo software PyMC3. Using the sampled posterior distribution of coalescence times, one can also infer related quantities such as the number of ancestors of a sample at a given time in the past (ancestral distribution) and the probability of specific relationships between branch lengths (for example, that the most recent branch is longer than all the others).

Open Symposium
SMBE-PO-497
The Impact of Cone Traits and Plate Tectonics on Podocarp Dispersal
K. V. Klaus [1,*], N. J. Matzke [2]
[1]Evolution and Biodiversity of Plants, Ruhr-University Bochum, Bochum, Germany, [2]School of Bological Sciences, University of Auckland, Auckland, New Zealand

Abstract: The ability of lineages to disperse long distances over evolutionary timescales may be influenced by the gain or loss of traits adapted to enhance local, ecological dispersal. For example, some species in the southern conifer family Podocarpaceae have fleshy cones that encourage bird dispersal, but it is unknown how this trait has influenced the clade's historical biogeography, or its importance compared to other predictors of dispersal such as the geographic distance between regions. We answer these questions quantitatively by using a dated phylogeny of 197 species of southern conifers to statistically compare standard, trait-independent biogeography models with new BioGeoBEARS models where an evolving trait can influence dispersal probability, and trait history, biogeographical history, and model parameters are jointly inferred. We validate the method with simulation-inference experiments. Comparing all models, those that include trait-dependent dispersal accrue 87.5% of the AICc model weight. Averaged across all models, lineages with non-fleshy cones had a dispersal probability multiplier of 0.49 compared to lineages with fleshy cones. Distance is included as a predictor of dispersal in all credible models (100% model weight). However, models with changing geography earned only 22.0% of the model weight, and models submerging New Caledonia/New Zealand earned only 0.01%. The importance of traits and distance suggests that long-distance dispersal over macroevolutionary timespans should not be thought of as a highly unpredictable chance event. Instead, long-distance dispersal can be modelled, allowing statistical model comparison to quantify support for different hypotheses.

Open Symposium
SMBE-PO-487
Unlocking the molecular evolution of extinct sloths: Mylodon Darwinii as a case study
M. Zicos [1 2,*], S. Brace [2], L. Frantz [1], S. Turvey [3]
[1]School of Biological and Chemical Sciences, Queen Mary University of London, [2]Department of Earth Sciences, Natural History Museum, [3]Institute of Zoology, Zoological Society of London, London, United Kingdom

Abstract: Sloths (Xenarthra, Folivora) were one of the dominant mammalian groups in Southern and Central America until the early Holocene. Although extant species of sloths are morphologically and ecologically similar, the large paleontological record of Folivora reveals a large range of body sizes, locomotion and ecology, all lost when these species became extinct.
Although new research is shedding light on the molecular evolution of sloths, this work largely focuses on modern species, which only represent a fraction of the diversity of this group. Given the large recent fossil record, ancient DNA methods can increase the number of sloth species from which genetic information is recovered, yielding new insights on their evolution.
Mylodon Cave (Ultima Esperanza, Chile) has been studied since 1899. Among the exceptionally preserved faunal record at the site, remains of Darwin's ground sloth (*Mylodon darwinii*) are found in soil layers dating from the end of the last Ice Age to their extinction in the early Holocene.
Specialised ancient DNA extraction and library preparation methods were applied on samples of bone, skin and coprolites of *M. darwinii* from Mylodon cave. Here we present preliminary findings from sequence data from multiple individuals of this species, exploring genetic diversity in this site, and the relationship of this species with extinct and extant Folivora. These findings, combined with direct radiocarbon dates, will be used to reconstruct the demographic history of the species, and to test models for their extinction, hoping to better understand why these sloths became extinct, while extant sloths persisted.

Open Symposium
SMBE-PO-488
Is temperature-induced sterility important for predicting species' responses to climate change?
S. Parratt*, T. Price [1], R. Snook [2], A. Bretman [3]
[1]University of Liverpool, Liverpool, United Kingdom, [2]Universiy of Stockholm, Stockholm, Sweden, [3]University of Leeds, Leeds, United Kingdom

Abstract: Rising global temperatures are threatening biodiversity. Studies on the impact of temperature on natural populations usually use lethal or viability thresholds, termed the "critical thermal limit" (CTL) of species. However, this overlooks important sublethal impacts of temperature that could affect species' persistence; such as the ability to produce offspring. Understanding how patterns of thermal limits to both fertility and viability correlate within and among species may be critically important for understanding species' vulnerability to rapid environmental change.
I have established standardised, novel lab methods to estimate the temperature at which species are rendered sterile; termed the thermal fertility limits (TFL). I have used this technique to estimate the TFL of multiple species of *Drosophila* that inhabit a range of natural habitats and thermal niches. I show that exposure to ecologically realistic shocks of high temperatures can cause significant drops in fertility in some, but not all, of these species. These measures of TFL do not correlate directly with CTLs in all species but do show a signal of phylogenetic constraint. I also show that in males, heat-induced fertility loss is most likely due to failures in spermatogenesis, resulting in delayed sterility following heat stress. Interestingly, the severity of the sterilising effect of heat correlates with sperm size, possibly highlighting a major cost to extreme gamete ornamentation.
These results shed light on the evolutionary trajectory and diversity of fertility. Further, they mean that we may have to revisit current predictions of how vulnerable species are to climate change.

Open Symposium

SMBE-PO-476

Ancient polyploidy in water lilies

Z. Li [1 2,*], R. Lohaus [1 2], H. Tang [3], L. Zhang [3], F. Chen [3], X. Zhang [3], Y. Van de Peer [1 2 4]

[1]Department of Plant Biotechnology and Bioinformatics, Ghent University, [2]VIB Center for Plant Systems Biology, VIB, Ghent, Belgium, [3]State Key Laboratory of Ecological Pest Control for Fujian and Taiwan Crops, Fujian Agriculture and Forestry University, Fuzhou, China, [4]Department of Biochemistry, Genetics and Microbiology, University of Pretoria, Pretoria, South Africa

Abstract: Water lilies are aquatic plants that belong to the order Nymphaeales, which is one of the three early-divergingangiosperm lineages including Amborellales, Nymphaeales, and Austrobaileyales. Transcriptomes of Nuphar species (a genus inNymphaeaceae) have uncovered an ancient whole-genome duplication (WGD) event within the lineage of Nymphaeales. Here, weanalyzed the newly-sequenced genome of Nymphaea colorata and transcriptomes of species from two of the three Nymphaealeanfamilies to investigate ancient polyploid events in the order Nymphaeales. The genome of N. colorata shows clear evidence of aWGD event. Its timing relative to the divergence of Nymphaeaceae and Cabombaceae is, however, confounded by differentsubstitution rates among Nymphaealean lineages, as well as incongruent results from phylogenomic analyses of duplicated genes oncollinear blocks and transcriptome analyses of Cabomba caroliniana (a species in Cabombaceae). In phylogenomic analyses, mostduplicated genes (lying in collinear blocks) coalesced prior to the divergence of Nymphaeaceae and Cabombaceae, supporting ashared WGD event by the two families. Transcriptome analyses of C. caroliniana show an absence of a clear KS peak and fewretained duplicates for such a shared WGD event. Our results could indicate that most duplicates were lost in the lineage to C.caroliniana after a shared WGD. Alternatively, the phylogenomic signatures for a shared WGD event could instead be interpreted asan allopolyploidy event that occurred shortly after the divergence between Nymphaeaceae and Cabombaceae ancestors, in line withthe substantial overlap of the date estimates for both the WGD and the divergence between Nymphaeaceae and Cabombaceae.

Open Symposium
SMBE-PO-484
Species distribution modelling combined with smart surveillance to estimate predictive distributions of microbes from genetic data

E. P. Grist*, M. I. Consortium

Abstract: Patterns of microbial community composition and distribution are driven by environmental factors and limitations in both spatial spread and dispersal. Understanding these associations is therefore essential for effective public health intervention and control measures as well as ongoing microbe surveillance. The strength of these associations for example, with soil habitat or climate, can provide valuable geographic information. However, a paucity of global microbiome data primarily from the urban environment currently restricts statistical inferences that can be made about these relationships. For example, in the field of forensics differentiating between potential sources of samples taken at crime scenes is most valuable. Here, we capitalize on the global microbiome dataset collected from subway and underground transportation systems by members of the MetaSUB consortium in a single day of 2018. The samples were fully sequenced and 3757 bacteria species were identified using KrakenHLL. The data included records of human pathogenic microbes such as *Klebsiella pneumonia* and *Enterobacter cloacae* established as leading causes of nosocomial infection in hospitals. We employed the species distribution model (SDM) MAXENT to estimate the most likely geospatial ranges within which a given sampled microbe species is expected to be present. The model was parameterized with georeferenced species presence in the MetaSUB database. Geographic maps were then determined for sample community compositions obtained at specific locations. We identified overlapping regions of SDM outputs to define the most probable occurrence region of a given microbiome assemblage. Model performance was evaluated by means of cross validation and ROC curve AUC. Given a microbiome assemblage found at *known* sampled location(s), our approach enables the presence of species to be estimated *elsewhere* coupled with the strength of their interdependence. Alternatively, given a microbiome assemblage found in a sample from an *unknown* location, the most and least likely sample sites of the source may be simultaneously identified.

Origins, evolution and function of novel genes

SMBE-PO-540

The rapid regenerative response of a model species Exaiptasia pallida is characterised by tissue plasticity and highly coordinated cell communication

C. A. van der Burg*, A. Pavasovic, E. Gilding, E. Pelzer, J. Surm, T. Walsh, P. Prentis

Abstract: Regeneration of a limb or tissue can be achieved through multiple pathways and mechanisms. The sea anemone *Exaiptasia pallida* has been observed to have excellent regenerative proficiency but this has not yet been described transcriptionally. In this study we examined the gene expression changes during a regenerative timecourse and report key genes involved in regeneration and wound healing. We found that the major response was an early (within the first 8 hours) upregulation of genes involved in cellular movement and cell communication, which likely contribute to a high level of tissue plasticity resulting in the rapid regeneration response observed in this species. Fifty-nine genes differentially expressed during regeneration were identified as having no orthologues in other species, indicating that regeneration in *E. pallida* may rely on the activation of species-specific novel genes. We examined the immune response during regeneration and found the immune system is only transcriptionally active in the first eight hours post-amputation. In accordance with some previous literature we conclude that the immune system and regeneration have an inverse relationship. Additionally, taxonomically-restricted novel genes, including species-specific novels, and highly conserved genes were identified throughout the regenerative timecourse, showing that both may work in concert to achieve complete regeneration.

Origins, evolution and function of novel genes
SMBE-PO-518
Widespread de novo gene emergence in the Drosophila clade
B. Heames [1,*], J. Schmitz [1], E. Bornberg-Bauer [1]
[1]University of Muenster, Muenster, Germany

Abstract: Orphan genes, which lack a detectable homologue in an outgroup species, are now recognised as consistent feature of eukaryotic genomes, typically representing 10-30% of a given organism's genes. More recently, efforts to identify how these genes are formed have indicated that de novo gene emergence from previously non-coding DNA may explain the prevalence of orphan genes. In this study, we have characterised orphan gene emergence across the Drosophila clade, and ascertained the emergence mechanism of these newly emerged genes by searching the genomes of closely related species. By comparison of the annotated proteomes of twelve Drosophila species, we are able to demonstrate a high number of taxon-specific orphan genes. Furthermore, we show that de novo gene emergence is an important driver of genetic novelty in the Drosophila clade, with around 10% of D. melanogaster orphans appearing to have emerged from non-coding genomic regions. For the majority of the remaining orphans it is not possible to find their emergence mechanism, suggesting that the true number of de novo genes could be much higher. Analysis of the sequence properties of de novo genes of different ages suggests that protein structural properties do not change significantly during their evolution, but that their nucleotide sequences gradually evolve to become more like those of ancient protein-coding genes. Finally, de novo genes consistently appear to be under selective constraint, suggesting functional roles for their protein products. Taken together, our results support a model of frequent gene emergence from non-coding DNA in the Drosophila clade, creating a pool of proteins free to acquire new functionality. While many of these proteins are rapidly lost, a proportion become fixed and go on to acquire functional cellular roles.

Origins, evolution and function of novel genes

SMBE-PO-525

Origin and evolution of thousands of small potentially coding open reading frames in primates.

D. Dowling*, J. F. Schmitz, E. Bornberg-Bauer

Abstract: Novel protein-coding genes can contribute to lineage-specific adaptations or integrate with pre-existing gene networks and acquire essential functions. Numerous mechanisms have been proposed to explain the origin of such novel genes including highly divergent paralogs, domestication of transposable elements, or *de novo* emergence from ancestrally non-coding DNA. A number of recent studies suggest that almost the entire human genome is transcribed into RNA and can serve as the basis for the formation of novel *de novo* genes. Small transcribed open reading frames (ORFs) have also been shown to be associated with ribosomes indicating that genomes may harbour vast quantities of small novel proteins. Most of these translated ORFs are likely non-functional but a small fraction may benefit the organism and become acquire functions and become genuine protein-coding genes. The structural properties of these novel ORFs and how gain functions and evolve over time are little understood. Here, using transcriptomic data from human and other primates we trace structural evolution of novel transcribed ORFs over 90 million years of evolution. Our results suggest that many short ORFs have arisen in the human lineage. The detection of homologous transcribed ORFs in other primate species indicates that these ORFs may have arisen millions of years ago. Moreover, many of these conserved ORFs are under purifying selection suggesting that they code for short functional polypeptides. However, we do not find evidence that these ORFs change over time in terms of physiochemical properties such as intrinsic structural disorder or aggregation propensity.

Origins, evolution and function of novel genes

SMBE-PO-522

Phylogenetic analysis of the CDGSH iron-sulfur binding domain reveals its ancient origin.

S. Sengupta [1,*], R. Nechushtai [2], P. A. Jennings [3], J. N. Onuchic [4], P. A. Padilla [1], R. K. Azad [1], R. Mittler [5]

[1]Department of Biological Sciences, University of North Texas, Denton, United States, [2]The Alexander Silberman Institute of Life Science, Hebrew University of Jerusalem, Jerusalem , Israel, [3]Department of Chemistry & Biochemistry, University of California at San Diego, La Jolla, [4]Center for Theoretical Biological Physics and Department of Physics, Rice University, Houston, [5]Division of Plant Sciences, Christopher S. Bond Life Sciences Center University of Missouri, Missouri, United States

Abstract: The iron-sulfur (2Fe-2S) binding motif CDGSH appears in many important plant and animal proteins that regulate iron and reactive oxygen metabolism. In human it is found in CISD1-3 proteins involved in diabetes, obesity, cancer, aging, cardiovascular disease and neurodegeneration. Despite the important biological role of the CDGSH domain, its origin, evolution and diversification, are largely unknown. Our main findings are: (1) the CDGSH domain appeared early in evolution, perhaps linked to the heavy use of iron-sulfur driven metabolism by early organisms; (2) a CISD3-like protein with two CDGSH domains on the same polypeptide appears to represent the ancient archetype of CDGSH proteins; (3) the origin of the human CISD3 protein is linked to the mitochondrial endosymbiotic event; (4) the CISD1/2 type proteins that contain only one CDGSH domain, but function as homodimers, originated after the divergence of bacteria and archaea/eukaryotes from their common ancestor; and (5) the human CISD1 and CISD2 proteins diverged about 650–720 million years ago, and CISD3 and CISD1/2 share their descent from an ancestral CISD about 1–1.1 billion years ago. Our findings reveal that the CDGSH domain is ancient in its origin and shed light on the complex evolutionary path of modern CDGSH proteins.

Origins, evolution and function of novel genes

SMBE-PO-524

Gene duplication and subfunctionalization underlies the evolution of opsins in the Lepidoptera

K. Kunte [1,*]

[1]NATIONAL CENTER FOR BIOLOGICAL SCIENCES, Bangalore, India

Abstract: Light plays an integral role in the life of organisms. The properties of ambient light have shaped the diversity of habitats and behaviors of animals. Animals use opsin genes to sense light, and this gene family has undergone tremendous evolutionary change. This is most prominent in the Lepidoptera (butterflies and moths) where diversity of opsins is mirrored by the diversity of habitat, and by wing coloration and patterning used for both intra- and inter-specific signaling. Opsins have a seven transmembrane receptor domain containing seven alpha-helices separating four extracellular and intracellular regions. Based on the wavelength sensitivity or their homology with *Drosophila* sequences, there are four opsin genes characterized in the Lepidoptera: UV opsin, Blue opsin, Long Wavelength (LW) and Rh7 opsin. To explore the molecular evolution within the opsin gene family, we analyzed the gene sequences and protein structures of 506 opsin genes from 132 species of Lepidoptera. We found that opsin gene family has undergone massive expansion in Lepidoptera with multiple duplication events unique to different families. We observed that these duplication events were more common than previously reported. This has given rise to multiple instances of subfunctionalization within the opsin gene family. Even though this gene family has undergone massive expansion, we found that the pattern of conservation and variation within the protein sequences was similar. The cytoplasmic regions of the proteins were relatively well-conserved compared to the extracellular and transmembrane regions. This conservation indicates possible signaling or recognition components within the opsin proteins. We also found that habitat had a significant influence on the evolution of opsin genes, with rates of evolution being different in nocturnal and diurnal species. By mapping the changes within the genes to the structure of proteins, we found mutation events that have influenced interactions of the protein with the chromophore. This has led to changes in the spectral sensitivity of the genes enabling lepidopterans to explore new light environments, which has had a profound influence on the evolution of wing color patterns and intra-specific communication. Thus, using an elaborate dataset, we show that gene duplication, and altered spectral sensitivity resulting from subfunctionalization, underlies the evolution of opsin genes in the Lepidoptera.

Origins, evolution and function of novel genes
SMBE-PO-519
A cryptic source of new genes in the E. coli pangenome
A. K. Watson*

Abstract: A range of different processes contributing to the origin of new genes in genomes have been described, including subfunctionalization and neofunctionalization following gene duplication events, acquisition of genes from external sources in HGT or EGT, and the association of existing protein coding sequences in new ways in domain fusion/fission events. Gene families can also originate from a combination of these different processes, often complicating the process of tracing their evolutionary history. As an example, our recent work has identified chimeric fusion genes made up of components from different sources, e.g. following EGT or HGT events. These included contributions of EGT or HGT to genomes that had not been observed with traditional screens focusing on full sized genes. Another way in which new genes can form is by "overprinting", where a new open reading frame originates from a *de novo* point mutations inside an existing gene. These "overlapping genes" do not necessarily stay together over time, but can be subject to subfunctionalization and separation. Overprinting is an established source of new genes in viruses, and has been reported in prokaryotes where they are hypothesised to contribute to the formation of ORFans, genes that have no detectable homologs in other lineages. However, few prokaryotic genes originating from overprinting have been positively identified. Here, we use sequence similarity searches to screen all complete genomes in the *E. coli* pangenome for gene families that may have originated by overprinting. Further, we explore the extent to which parts of "overprinted" genes may have contributed to the origin of new fusion genes, which may have previously hindered their detection. Finally, we explore the putative functions of their genes and the proportion of their contribution to the genes considered as "ORFans" in the *E. coli* pangenome.

Origins, evolution and function of novel genes
SMBE-PO-520
The Biographies of Young Overlapping Genes in Bacteria - Phylostratigraphy and Gene Properties
Z. Ardern[*], K. Neuhaus[1], S. Scherer[1]
[1]Technical University of Munich, Freising, Germany

Abstract: Many ORFs which exhibit evidence for translation in bacteria and are evolutionarily young are found to significantly overlap annotated genes in alternate reading frames; either in sense or antisense orientation. Surprisingly, there is evidence that hundreds of young ORFs in Escherichia coli and other bacterial species are translated and are significantly overlapping in antisense to annotated genes. This phenomenon is made possible by the triplet genetic code, which allows for six possible translational reading frames at each DNA locus. In recent years several of these overlapping genes in bacteria have been phenotypically characterized, and the existence of many others has been inferred based on publicly available data including different methods for ribosomal profiling.

Here we report gene ages as determined by phylostratigraphic analysis and comparative transcriptomics across species, as well as details of sequence evolution. Gene properties considered include the relative reading frame of the overlapping genes, which affects codon and amino acid usage. These remarkable overlapping regions may be a significant contributor to novelty in the bacterial proteome, and are striking examples of de novo gene origin. The mechanism of the origin of such genes is presumably through overprinting events, where ORFs in alternative reading frames are translated and subsequently conserved due to some selected benefit. The details of this evolutionary process in bacteria, which we have begun to elucidate, deserve significant further attention.

Origins, evolution and function of novel genes
SMBE-PO-528
The draft genome of Actinia tenebrosa reveals insights into the evolution of venom innovations

J. M. Surm [1,*], Z. K. Stewart [2 3], A. Papanicolaou [4], A. Pavasovic [5], P. J. Prentis [2 3]
[1]Ecology, Evolution and Behavior, Alexander Silberman Institute of Life Sciences The Hebrew University of Jerusalem, Jerusalem, Israel, [2]School of Earth, Environmental and Biological Sciences, [3]Institute for Future Environments, Queensland University of Technology, Brisbane, [4]Hawkesbury Institute for the Environment, University of Western Sydney, Penrith, [5]School of Biomedical Sciences, Queensland University of Technology, Brisbane, Australia

Abstract: Venomous animals have evolved a wide array of toxic compounds with highly diverse and specialised pharmacological and biochemical properties. Toxin peptides found in venoms are often novel, lacking homology to other taxa. Cnidarians are an ancient venomous lineage that delivers toxins using novel stinging cells, called cnidocytes. These cnidarian-specific cells can be found across multiple morphological structures, some of which are unique to lineages within this phylum. Investigating venom and its delivery in cnidarians provides insights into the evolution and function of novel structures, cells, and genes. This project generated a draft genome for the sea anemone *Actinia tenebrosa* to examine the evolution of novel genes in cnidarian species using a comparative phylogenetic approach. Additionally, multiple quantitative RNA-seq analyses were performed to elucidate patterns of spatiotemporal gene expression. My analyses identified a highly conserved cnidarian core-gene set, but also a large proportion of novel gene families restricted to specific cnidarian orders. Overall, gene duplication dominated the evolution of novel genes in cnidarians, with many novel gene families undergoing pronounced expansion events in at least one taxa. Taxa-specific toxin genes contributed to a significant proportion of the novel and expanded gene families in sea anemone species. Many of these novel toxin genes showed distinct patterns of spatiotemporal gene expression and were upregulated in lineage-specific envenomating structures. Our results are an unfolding story demonstrating that cnidarian evolution is underpinned by the neofunctionalisation of novel toxin genes that can confer an advantage across their complex life cycle and ecological niche.

Origins, evolution and function of novel genes

SMBE-PO-544

The ancient salicoid genome duplication event: a platform for intra-genome, inter-species and inter-genera reconstruction of de novo gene evolution in Populus trichocarpa

T. Yates [1,*], K. Feng [2], J. Zhang [2], P. Ranjan [2], P. Abraham [2]

[1]Bresdesen Center for Interdisciplinary Research, University of Tennessee, Knoxville, [2]Biosciences Division, Oak Ridge National Laboratory, Oak Ridge, United States

Abstract: Orphan genes lack sequence similarity to genes in other species and represent an important part of genome evolution as a novel source of genetic material. We identify 435 genes specific to P. trichocarpa with 5 of the 435 showing evidence of de novo gene evolution. Populus and its sister genera Salix are particularly well suited for the study of orphan gene evolution as a result of the Salicoid whole genome duplication (WGD) which resulted in highly syntenic sister chromosomal segments across the Salicaceae. We leverage this genomic feature to reconstruct de novo gene evolution from inter-genera, inter-species, and intra-genomic perspectives, by comparing the syntenic regions within P. trichocarpa, then P. deltoides, and finally Salix purpurea. Additionally, we also utilize the Populus Genome-wide association mapping panel (GWAS) population, a collection of 1,084 undomesticated genotypes to further understand the population genetics of orphan gene evolution. Furthermore, we use transcriptomics and proteomics to provide evidence of function for a large cohort of orphan genes. Overall, we provide new insights into the processes of de novo gene evolution in the context of a long-lived perennial tree.

Origins, evolution and function of novel genes

SMBE-PO-527

A well timed snapshot of the human regulome reveals positions and function of conserved regulatory enhancers and identifies putative novel promoters.

D. T. Gerrard [1,*], A. Berry [1], N. Bobola [1], K. Piper-Hanley [1], I. Donaldson [1], R. Jennings [1], N. Hanley

[1]Faculty of Biology, Medicine and Health, University of Manchester, Manchester, United Kingdom

Abstract: We have generated genome-wide regulatory maps from eleven human tissues during organogenesis. The maps reveal promoter activation, repression and novel candidate enhancer-to-gene interactions. We validate conserved enhancer function using zebrafish and show, for the first time in human development, the extent of enhancer re-use across tissues. Additionally, add RNA-seq to locate previously un-annotated promoter-like elements active in early human development.

Origins, evolution and function of novel genes

SMBE-PO-526

Evolution of the DUF26-containing genes

A. Vaattovaara [1,*], B. Brandt [2], S. Rajaraman [1], O. Safronov [1], A. Veidenberg [1], M. Luklova [1], J. Kangasjärvi [1], A. Löytynoja [1], M. Hothorn [2], J. Salojärvi [3], M. Wrzaczek [1]

[1]University of Helsinki, Helsinki, Finland, [2]University of Geneva, Geneva, Switzerland, [3]Nanyang Technological University, Singapore, Singapore

Abstract: Linking gene family expansions to the functional evolution of proteins is an interesting challenge in evolutionary and plant biology alike. Among their many large gene families, plants contain a large number of receptor-like protein kinases (RLKs) to be able to respond to changes in their environment. One subgroup of RLKs, the cysteine-rich receptor-like protein kinases (CRKs) are distinguished by their extracellular region which typically contains two DUF26 (domain of unknown function 26; stress-antifung domain, PF01657) domains. The DUF26 domain harbours the conserved cysteine motif C-8X-C-2X-C in its core and is also found in two groups of receptor-like proteins, the plasmodesmata-localized proteins (PDLPs) and the cysteine-rich receptor-like secreted proteins (CRRSPs).

DUF26-containing proteins have experienced dramatic lineage-specific expansions during the evolution of land plants. Thus, in order to understand the evolution of DUF26 genes, we identified gene models containing the DUF26 domain from 32 algal and plant genomes. Manually curated DUF26 protein sequences were used for construction of phylogenetic trees and analysis of domain composition rearrangements and duplication mechanisms. To complement the sequence-based analyses, crystal structures of two PDLPs were determined and analysed.

The DUF26 domain was identified only from land plants based on the protein sequence motif. Our data suggests that genes containing one DUF26 domain appeared in liverworts and the tandem DUF26 arrangement appeared in lycophytes. In Angiosperms, in particular CRKs have expanded as the result of tandem duplications in different lineages. This might be linked to the roles of CRKs in stress adaptation. While CRKs, PDLPs and CRRSPs share a common ancestor they underwent different domain rearrangements and expansions including exchange of kinase domains and secondary evolution of CRRSPs from CRKs. Intriguingly, in genes with two DUF26 domains, the first and the second DUF26 have differentiated into specific forms with unique sequence context surrounding the conserved cysteines. There is also considerable variation within DUF26 domains between different phylogenetic subgroups. This variation is likely functionally important as it results in vastly different surface charge distribution of the DUF26 domain. Interestingly structural analyses revealed a strong similarity to fungal lectins. This suggests that the DUF26 domain could function as lectin.

Origins, evolution and function of novel genes

SMBE-PO-523

Extensive translatome data illuminates the evolution of gene expression and new gene origination in mammals

E. Leushkin*, J. Knopf, F. Murat, M. Sepp, S. Anders, M. Lemberg, H. Kaessmann

Abstract: Gene expression programs define shared and species-specific phenotypes. However, previous comparative expression studies were largely restricted to the transcriptome. We generated ribosome profiling and matched RNA sequencing data for three organs (brain, liver, testis) across five representative mammals (human, macaque, mouse, opossum, platypus) and a bird (chicken), with the aim to scrutinize the dynamics of translatome and transcriptome evolution. Specifically, we sought to contrast the rate of translatome (protein synthesis) versus transcriptome evolution for old genes, shared across mammals, and to assess in detail the formation of new protein-coding genes through various mechanisms. I will present highlights of this work, with a focus on the birth of new genes from scratch (de novo gene origination) and the rebirth of new coding sequences following the decay of ancestral precursors.

Origins, evolution and function of novel genes

SMBE-PO-535

The double role of enhancers in the formation and perseverance of novel genes

P. Majic [1,*], J. L. Payne [1]

[1]ETH Zürich, Zürich, Switzerland

Abstract: Gene regulatory networks control the spatiotemporal expression patterns that give rise to and define the individual cell types of multicellular organisms. In Metazoa, distal regulatory elements called enhancers play a key role in determining the structure of such networks, particularly the wiring diagram of 'who regulates whom.' Mutations that affect enhancer activity can therefore rewire regulatory networks, causing changes in gene expression that may be adaptive. Here, we use single-cell transcriptomic and chromatin accessibility data from mouse to show that enhancers play an additional role in the evolution of regulatory networks: They facilitate network growth by creating transcriptionally active regions of open chromatin that are conducive to *de novo* gene evolution. Specifically, our comparative transcriptomic analysis with three other mammalian species shows that young, mouse-specific transcribed open reading frames are preferentially located near enhancers, whereas older open reading frames are not. Interactions with enhancers are then gained incrementally over macro-evolutionary timescales, helping to integrate new genes into existing regulatory networks. Taken together, our results highlight the dual role of enhancers in expanding and rewiring gene regulatory networks.

Origins, evolution and function of novel genes
SMBE-PO-532
Evolution of a new function by chimeric fusion between phage DNA and a bacterial gene

O. M. Warsi [1,*], M. Knopp [1], S. Surkov [1], J. J. Hultqvist [2], D. Andersson [1]
[1]Department of Medical Biochemistry and Microbiology, Uppsala University, Uppsala, Sweden, [2]Department of Biochemistry and Molecular Biology, Dalhousie University, Halifax, Canada

Abstract: Mobile genetic elements, in the form of plasmids, phages and transposons, are important sources for evolution of novel functions. In this study we perform large-scale screening of metagenomic phage libraries in temperature-sensitive Salmonella enterica mutants to elaborate on the mechanistic basis for this route of evolutionary novelty. We identified a 23 amino acid insert from an extracellular mobile genetic element that when fused with a host DNA-binding repressor (LacI) results in the formation of a chimeric protein, which localizes to the outer-membrane. This relocalization of the chimeric protein results in a six-fold increase in membrane vesicle formation, suppressing the temperature sensitivity in the host. Both the host gene and the extracellular 23 amino acid stretch are specific and necessary for the generation of the novel phenotype. Furthermore, mutation analysis of the chimeric protein shows that although the native repressor function of the protein is maintained in this chimeric structure, it is not necessary for the new function. Thus our study demonstrates how a gene fusion between extracellular DNA and bacterial DNA can generate novelty without compromising the native function of a given gene.

Origins, evolution and function of novel genes

SMBE-PO-545

From organismal evolution to genic evolution-The neverending story of new genes in the Red Queen landscape

G.-A. Lu [1,*], Y. Zhao [1], H. Yang [1], J. Xu [2], T. Tang [1], C.-I. Wu [3]

[1]State Key Laboratory of Biocontrol School of Life Sciences, Sun Yat-Sen University , Guangzhou, China, [2] Center for Personal Dynamic Regulomes, Stanford University, Stanford, California, [3] Department of Ecology and Evolution, University of Chicago, Chicago, Illinois, United States

Abstract: The Red-Queen hypothesis depicts an ever-changing adaptive landscape. The hypothesis, commonly invoked to explain organismal evolution, can also be applied to genic evolution. In this study, we test the Red-Queen hypothesis by studying the evolution of new genes, which may be compared to the younger taxa in organismal evolution. If the adaptive landscape shifts rapidly as posited by the Red-Queen hypothesis, new genes, especially those originated from non-genic sequences, may be under strong pressure to evolve continually to adapt to the changing word. We survey the evolution dynamic of six *Drosophila* de novo miRNAs by generating their mutants and analyzing their contribution to fitness. In *D. melanogaster*, these miRNAs' fitness advantage seems to be neutral even though their sequences subject to past positive selection, revealing ongoing period-specific adaptations. When survey the function divergence of two miRNAs in *Drosophila* sibling species, we find their fitness contribution is neutral in one species, but positive in the other, suggesting rapid functional evolution or death of de novo genes. Furthermore, gene expression profiles for miRNAs mutants reflect extensive transcriptome divergence in fly sibling species, illustrating great re-writing of miRNA mediated gene regulation networks during evolution. Collectively, our results demonstrate de novo genes either evolve rapidly or face elimination, reminiscent of the metaphor of the "Red Queen" effect. In the Red Queen landscape, adaptive evolution does not always resemble the large-scale and stable adaptive changes. Instead, adaptations may often be transient, small-scale, and local.

Origins, evolution and function of novel genes

SMBE-PO-529

Life sustaining activities of random DNA sequences: Rescuing a lethal E. coli RNase P mutant

A. M. Babina [1,*], D. I. Andersson [1]

[1]Department of Medical Biochemistry and Microbiology, Uppsala University, Uppsala, Sweden

Abstract: The origination of new genes is one of the primary processes driving evolutionary innovation, as new genes often provide the starting material necessary for the emergence of adaptive novelties. While the mechanisms governing the evolution of novel genes from pre-existing genes are relatively well understood (i.e. horizontal gene transfer, duplication-divergence, gene fusion/fission), how a new gene is born *de novo* from nonfunctional sequence remains enigmatic. To address this question, we introduced highly diverse plasmid libraries encoding randomly generated small ORFs of varying lengths into a temperature-sensitive *Escherichia coli* RNase P mutant and screened for sequences that enable strain growth at the non-permissive temperature. We recovered seven unique sequences from our screen and preliminary characterization suggests that these sequences do not encode peptides, but rather act at the RNA level. We are currently investigating whether the selected constructs rescue the temperature-sensitive phenotype by altering RNase P expression or directly interacting with the mutant RNase P complex. Our findings demonstrate that the evolution of new genes from nonfunctional and/or random sequence is more probable than previously believed. Similar studies have selected for new functions using libraries limited to small stretches of randomization within pre-existing structural and/or functional scaffolds. Our approach allows for the *in vivo* selection of true *de novo* genes from completely random nucleotide sequences, without structural constraints or functional progenitor sequences. Our experimental setup also provides a convenient framework for examining how novel gene products evolve and integrate into existing pathways or regulatory networks within an organism.

Origins, evolution and function of novel genes

SMBE-PO-531

Meta-analyses of an integrated resource, GenTree, detected hundreds of primate-specific coding genes implicated in fast-evolving biological processes of human

Y. E. Zhang [1,*], Y. Shao [1], C. Chen [1], H. Shen [2], B. Z. He [3], D. Yu [1], S. Jiang [4], S. Zhao [5], Z. Gao [6], H. Yuan [1], Z. Zhu [7], X. Chen [8], Y. Fu [6], H. Chen [5], G. Gao [4], M. Long [9]

[1]Key Laboratory of Zoological Systematics and Evolution and State Key Laboratory of Integrated Management of Pest Insects and Rodents, Institute Of Zoology, Chinese Academy of Sciences, Bejing, [2] College of Computer,, Hunan University of Technology, Hunan, China, [3]Department of Biology, University of Iowa, Iowa, United States, [4]School of Life Sciences, Peking University, [5]CAS Key Laboratory of Genomic and Precision Medicine, Beijing Institute of Genomics, Chinese Academy of Sciences, [6]National Center for Mathematics and Interdisciplinary Sciences, Academy of Mathematics and Systems Science, Chinese Academy of Sciences, Bejing, [7]School of Life Sciences, Chongqing University, Chongqing, [8]Wuhan Institute of Biotechnology, Wuhan, China, [9]Department of Ecology and Evolution, The University of Chicago, Chicago, United States

Abstract: The origination of new genes contributes to phenotypic evolution in humans. Two challenges in the functional study of new genes are the inference of gene ages and annotation of their protein-coding potential. To tackle them, we developed GenTree, a database (http://gentree.ioz.ac.cn) that compiles age inferences data and functional genomic data. Meta-analyses revealed that the age data of synteny-based pipeline (SBP) are largely consistent with that of the protein-family–based methods for *de novo* genes. However, SBP appears suitable for recently duplicated genes, whereas protein-family–based methods are useful for ancient genes. For SBP-dated primate-specific genes (PSGs), we performed manual evaluation based on published PSG lists and showed that SBP tended to be conservative by masking unreliable syntenic regions. We then curated 254 PSGs with protein evidence including 41 pseudogenes and found that they are preferentially recruited into spermatogenesis, immune response, mother-fetus interaction and brain development. For brain development, primate-specific KRAB zinc-finger proteins are up-regulated in the mid-fetal stage, which may contribute to the evolution of this critical period. Altogether, hundreds of PSGs are recruited to fast-evolving processes. We are now updating GenTree as: 1) to infer synteny by taking advantage of outgroup assemblies based on PacBio long reads; 2) to correct exon-intron models based on PacBio transcriptomes; 3) to estimate coding potential based on lineage-specific inferences of selection force; and 4) to incorporate single-cell RNA-sequencing data. In this way, GenTree should be able to provide more accurate age or gene model information and to motivate more specific functional hypothesis.

Origins, evolution and function of novel genes

SMBE-PO-538

Reassessing the count of human protein-coding genes

J. Mudge [1,*], J. M. G. Martinez [1], I. Jungreis [2], M. Tress [3], P. Flicek [1], A. Frankish [1]

[1]European Bioinformatics Institute, European Molecular Biology Laboratory, Hinxton, United Kingdom, [2]Broad Institute of MIT and Harvard, Cambridge, United States, [3]Bioinformatics Unit, Spanish National Cancer Research Centre, Madrid, Spain

Abstract: While the question '*how many human protein-coding genes are there*?' may seems like a relic of the late 1990s, it remains hotly debated. Indeed, the counts of the major gene and protein annotation sets – e.g. Ensembl / GENCODE, RefSeq and Uniprot – are fluid; novel proteins are still being discovered, while existing proteins are frequently re-appraised as incorrect. Here, we will discuss the current drive of the Ensembl / GENCODE project to finalize gene counts in human and mouse, and also to explicitly 'itemize' the relationships between all protein-coding genes, as well as pseudogenes, in terms of homology and evolutionary provenance. These efforts are centered on a detailed workflow incorporating comparative manual annotation and evolutionary analysis, combined with high-throughput transcriptomics and proteomics datasets. Nonetheless, there are currently several thousand prospective coding genes with ambiguous functionality in both species. These broadly fall into two categories: genes with experimental support for translation that lack a signature of protein-coding evolution, i.e. potential *de novo* genes, and – conversely - genes that exhibit protein homology but lack experimental support, i.e. potential pseudogenes. Uncertainties in the former partially reflect interpretive questions on datasets that indirectly support protein function, including ribosome profiling assays. However, we will also discuss the possibility that 'aberrant' proteins exist, and that the link between translation and genuine protein functionality is not straightforward. Meanwhile, it is often hard to distinguish protein-coding genes from pseudogenes; some genes are known to produce viable proteins in spite of nonsense mutations, while genes with intact coding sequences can become pseudogenised by other means. Although such cases are ultimately resolved through experimental methods, we shall also discuss the utility of human variation datasets in examining the selective pressure acting on ambiguous coding sequences.

Origins, evolution and function of novel genes

SMBE-PO-537

De novo evolved genes are essential for Drosophila male fertility and act at multiple stages of spermatogenesis

G. D. Findlay [1,*], E. L. Rivard [1], P. H. Patel [1], J. F. Schmitz [2], B. J. Kelly [1], G. C. Mascha [1], E. R. Scott [1], P. H. Rumde [1], E. Bornberg-Bauer [2]

[1]Department of Biology, College Of The Holy Cross, Worcester, United States, [2]Institute for Evolution and Biodiversity, Muenster, Germany

Abstract: *De novo* evolved genes arise from non-protein coding genomic material and have potential to develop essential functions within a short evolutionary time frame. Across species, such genes are often expressed specifically in male reproductive structures, suggesting that their retention after gene birth may be driven by sexual selection. We conducted a bioinformatic screen in *Drosophila melanogaster* that identified 96 putative *de novo* genes enriched for expression in the testes, and we have used testis-specific RNA interference and CRISPR-mediated gene mutation to evaluate their functions. To date, we have identified three putative *de novo* genes that have become required for male reproduction, as both knockdown and knockout of each gene result in almost complete infertility. Cytological analyses revealed that these genes affect at least four steps in spermatogenesis: the individualization of mature sperm cells after meiosis; the condensation of late-stage spermatid nuclei; the coiling of individualized sperm bundles prior to their entry into the seminal vesicle; and, the ability of transferred sperm to localize properly within the female reproductive tract after mating. Molecular evolutionary analyses found that these genes have evolved under different selective regimes since their origins. For example, one gene has evolved under positive selection and undergone a complex pattern of gene duplication and loss across *Drosophila* species, while another gene migrated from the X chromosome to an autosome, possibly to escape meiotic sex chromosome inactivation. Taken together, our results highlight the myriad ways in which novel genes may lead to male reproductive adaptations and suggest that *de novo* genes may rapidly integrate into, and become essential parts of, existing cellular networks.

Origins, evolution and function of novel genes
SMBE-PO-541
Furthest From Singleton: a new way of dating paralogs
M. Reardon [1,*], A. Martin-Geary [1], D. L. Robertson [2]
[1]Faculty of Biology, Medicine and Health, University of Manchester, Manchester, [2]University of Glasgow, Glasgow, United Kingdom

Abstract: Various methods have been previously used to assign approximate ages to genes, based on their evolutionary history as represented by phylogenetic trees constructed from putative speciation and duplication events. The ages assigned to genes in a family of paralogs vary widely according to the method used. For example, the Last Common Ancestor method assigns the oldest gene age to all family members whereas the Most Recent Duplication method assigns the age of the youngest duplication event in each gene's history. These contrasting approaches, while having valid uses, result in characteristically skewed age distributions and fail to utilise all the available information in the phylogenetic tree. Here we present a novel method, Furthest From Singleton, that uses all the information contained in the phylogenetic tree of each paralog family for the first time. We demonstrate the utility of the approach for dating genes linked to heritable disease. The results indicate that the strong association between paralogs, specifically ohnologs, and dominant disorders is often a consequence of a mechanism through which pre-existing dosage-sensitive or haploinsufficient genes are successfully duplicated and retained. Heritable disease is thus as much a consequence of the fragility of evolutionarily more ancient genes as compensatory mechanisms.

Origins, evolution and function of novel genes

SMBE-PO-543

Using biological features for new gene predictions

G. Goldstein [1 2,*], F. Monfardini [1], M. D. Vibranovski [1]

[1]Instituto de Biologia, Universidade de São Paulo, São Paulo, Brazil, [2]Institute of Zoology, UCAS, Beijing, China

Abstract: New genes can appear in a genome through several genetic mechanisms like DNA duplication and retrotransposition, and are classified as such because they are recent in a taxon's evolutionary history. Studies in the area usually start with the identification of new genes through a genome dependent method, such as those using synteny information. After identification and annotation, new genes have their features characterized, such as expression profile and evolutionary traits. Thanks to past studies in the area, several patterns emerged: new genes are usually expressed in testis, evolve in a faster rate than older genes and are preferentially expressed in the post-meiosis phase of spermatogenesis. Empowered with this information, we asked the question of how well these features and characteristics are able to predict gene age. In other words, are we able to infer if a gene is new by looking at its expression profile and evolutionary signatures. For this purpose, we worked with *Drosophila melanogaster*, which has extensive genomic and evolutionary data as well as an extremely reliable new gene list, allowing our investigation to be done in an optimal environment. We have performed relative risk calculations using the results of differential expression data obtained using EBSeq and found that new genes are 40% more likely to be testis biased. That percentage increases to 260% when considering genes that have three-fold change difference between testis and ovary, while old genes are five times more likely to be ovary biased than new genes. With this approach, we hope to give more methodological options to the identification of new genes, particularly in non-model species.

Origins, evolution and function of novel genes

SMBE-PO-536

Structure prediction and experimental validation of a putative evolved de novo gene from D. melanogaster

A. Lange [1,*], P. H. Patel [2], G. Findlay [3], E. Bornberg-Bauer [1]

[1]Institute for Evolution and Biodiversity, WWU Muenster, Muenster, Germany, [2]Department Biology, [3]Department of Biology, College of the Holy Cross, Worcester, United States

Abstract: In recent years it has become apparent that protein-coding genes emerge from non-coding regions of the genome at a measurable rate. Such *de novo* genes have now been identified across the Tree of Life, and appear to form a crucial source of innovation for protein evolution (Zhang W et al., 2015; Guerzoni et al., 2016). The question arises: what are the structural properties of proteins that emerge from 'random' sequence space - and how does this affect their evolution? Relatedly, what did the first proteins look like, and how did their ancestral sequence composition influence their structure? Characterisation of random-sequence proteins reveals abundant secondary structure and apparent *in vivo* toleration (Tretyachenko et al., 2017). However, identified *de novo* proteins are often disordered, with reduced aggregation propensity. A complete picture of the structure and dynamics of random-sequence proteins will eventually come from experimental case studies (Bungard et al., 2017; Vakirlis et al., 2017).

One of the greatest obstacles for the field of *de novo* genes is the lack of high-quality genomic data with very short divergence times which could help to precisely pin down their location. As a result, it remains unclear how these properties are shaped by evolution, depend on genetic mechanisms and influence gene survival. Taken together, it seems that *de novo* genes survived initial purging, likely because they are selectively neutral or beneficial for the organism.

To date, some *de novo* genes have been identified in different species, however, their further characterisation (structure, function, etc.) is often missing. Here we present a putative *de novo* gene from *D. melanogaster* named Goddard. We used a combination of structural prediction, expression in *E. coli* cells and CD and NMR measurements to investigate its possible folding pattern. Goddard is small with only 112aa. It forms two helices as secondary structure motifs which we could also see in CD and NMR measurements. Additionally these two helices, one at the N-terminus and a bigger one in the middle tend to form a coiled-coil pattern. The C-terminus, however, remains disordered. Last we compared Goddard from Dmel with other *Drosophilas* (*ana*, *vir*, *moij*, and *gri*) to find similarities. Only the folding pattern of *gri* seems to be different. This could mean that our target protein likely adopted a certain structure and function early on after its birth, and has been largely conserved since then.

Origins, evolution and function of novel genes

SMBE-PO-534

EXTENSIVE NON-REDUNDANCY IN A RECENTLY DUPLICATED DEVELOPMENTAL GENE FAMILY

E. A. Baker [1,*], A. Woollard [1], S. P. R. Gilbert [1], S. Shimeld [2]

[1]Biochemistry, [2]Zoology, University of Oxford, Oxford, United Kingdom

Abstract: It is widely accepted that recently duplicated genes are more likely to be redundant with one another compared to ancient paralogues. The evolutionary logic underpinning this idea is simple, as the assumption is that recently derived paralogous genes will be more similar in sequence compared to ancient gene families. On this basis, it is expected that these novel or taxon-restricted duplicates rarely perform essential roles in development. We set out to interrogate these assumptions by using molecular phylogenetics and exploiting the genetic tractability of the nematode *Caenorhabditis elegans* in studying the nematode-specific family of Hedgehog-related genes, the Warthogs. Hedgehog is one of a handful of signal transduction pathways that underpins the development of bilaterian animals. While having lost a bona fide Hedgehog gene, nematodes have evolved an expanded repertoire of Hedgehog-related genes, 10 of which reside within the Warthog family. We have characterised their evolutionary history and their roles in *C. elegans* and found that these genes function in aspects of post-embryonic development, including left-right asymmetry and cell fate determination. Analysis of various double and triple mutants of the Warthog family reveals that more recently derived paralogues are not redundant with one another whilst more divergent Warthogs are. Therefore contrary to expectation, we find that novel gene classes can assume important roles in development. Moreover despite their recent derivation, members of taxon-restricted gene families are not always functionally redundant which is considered paradoxical according to the current framework in gene evolution.

Origins, evolution and function of novel genes

SMBE-PO-539

Evolution of the intelectin gene family through gene gain, gene loss and adaptive sequence evolution

F. Almalki [1], L. Patel [1 2], E. B. Nonnecke [3], P. A. Castillo [3], D. T. Akahoshi [3], L. Odenthal-Hesse [4], L. R. Machado [1 2], C. L. Bevins [3], E. J. Hollox [1,*]

[1]Department of Genetics and Genome Biology, University of Leicester, Leicester, [2]School of Health, University of Northampton, Northampton, United Kingdom, [3]Department of Microbiology and Immunology, University of California Davis, Davis, United States, [4]Max Planck Institute for Evolutionary Biology, Plon, Germany

Abstract: The intelectin (*inte*stinal *lectin*) gene family encodes secreted glycoproteins involved in the specific recognition of microbial glycans. In mammals, intelectins are primarily produced in the lung and intestine and there is increasing evidence for an important role in innate immunity. Intelectins show extensive lineage-specific duplication events, and recent gene loss in particular mouse laboratory strains.

Here we analyse patterns of evolution of intelectins, with a particular focus on mice. Analysis of the intelectin region in mice is problematic, as the reference genome strain C57/BL6 has only one intelectin gene, Itln1. Other intelectin genes will be invisible to any analysis in mouse that has relied on the reference genome for either detection of variation or for development of a genome of a different strain. We used a BAC contig spanning the six known intelectin genes in mouse strain 129S7 to remap short Illumina sequence reads and infer intelectin copy number and sequence variation in laboratory mouse strains and wild-caught mice. We validated a subset of the variation using long PCR and Sanger sequencing.

We can use this analysis in the context of a phylogenetic tree of vertebrate intelectins to infer a pattern of gene gain and gene loss in rodents, and to identify amino acid sites that have undergone positive selection across the vertebrate tree using maximum-likelihood-based dN/dS statistics, and within rodents using McDonald-Kreitman tests. Positively-selected sites are mapped on to a known intelectin crystal structure to relate the evolutionary history of sequence change to the function of intelectins.

Outreach in Molecular Biology and Evolution: good practices and challenges

SMBE-PO-546

Old Fossils, New Tricks: Using techniques learned in Palaeontological Outreach to better communicate Molecular Biology

J. Fleming*

Abstract: Few scientific fields can boast the 'wow' factor that palaeontology can immediately evoke in a crowded room, particularly in the classroom. A large amount of academic palaeontological research is government-funded, and is reliant on keeping the field in the public eye and recipient of public goodwill. In addition, palaeontology cultivates and nurtures a large community of valuable, enthusiastic amateur fossil hunters and taxonomists, without whom many specimens would never be found or identified. Whilst charismatic extinct creatures can often provide a useful hook, initiatives such as the Bristol Dinosaur Project draw on palaeontology's long history of successful outreach and incorporate cleverly designed tactile and kinetic learning experiences to greatly expand public understanding of the entire discipline.

During my time as a palaeontology MSc and PhD student, I was able to work alongside the Bristol Dinosaur Project, as well as a number of other outreach initiatives across a range of audiences. After switching disciplines to molecular biology, I found that the techniques I had learned to express abstract scientific concepts such as geological history and taphonomy could be easily applied to discuss issues of phylogeny and gene duplication.

In this talk I will discuss some of the skills that I have learnt whilst working in palaeontology that I have been able to use to help establish and maintain effective communication in molecular biology, in both classroom and public settings. Particularly, I will focus on the design of props and kinetic learning activities, and how to avoid some of the common pitfalls they present.

Outreach in Molecular Biology and Evolution: good practices and challenges
SMBE-PO-547
Melanogaster Catch The Fly: a citizen science project on adaptation genomics
J. González [1,*], R. Torres [2]
[1]Comparative and Computational Genomics, Institute of Evolutionary Biology (CSIC-UPF), [2]La Ciència Al Teu Mòn, Barcelona, Spain

Abstract: Melanogaster: Catch the Fly!" (MCTF) is the first European network of citizen science in adaptation genomics. In MCTF, high school teachers and students from rural areas of Spain, participate collecting and classifying Drosophila melanogaster (fruit flies). The collections are send to the Evolutionary and Functional Genomics Laboratory (IBE, CSIC-UPF) in order to study how organisms adapt to the environment. To promote the interest of the public in Evolutionary Biology and the importance of their contributions to scientific research, prior to collecting samples, participants learn about Biology and Genetics concepts as well as the importance of generating knowledge through basic research. We provide the schools with educational materials that help teachers convey the importance of the scientific questions behind the citizen science project. All Spanish schools that are part of MCTF will continue their collaboration with scientists for a five-year period. They will also act as role models for other schools in Europe, as we want to extend MCTF to the countries that are part of the European Drosophila Population Genomics Consortium, and beyond.
MCTF, promotes a deep engagement of the participants and their communities, because it allows them to directly contribute to research in Evolutionary biology. MCTF also allows the public to be part of science advances, no matters how far away they are from a research center.

Phylogenomics under the multispecies coalescent

SMBE-PO-560

Characterising and visualising gene families within Galaxy using GeneSeqToFamily and Aequatus

A. Thanki [1,*], N. Soranzo [1], W. Haerty [1], J. Herrero [2], R. P. Davey [1]

[1]Earlham Institute, Norwich, [2]UCL, London, United Kingdom

Abstract: The phylogenetic information inferred from the study of homologous genes helps us to understand the evolution of gene families and plays a vital role in finding ancestral gene duplication events as well as identifying genes that are under positive selection within species. Various tools exist to identify gene families, but they typically do not provide information about structural changes within a gene. Similarly, collating and configuring the many software to discover gene families often requires solving many dependencies making the pipeline dependent on a single computing environment. Here, we present a complete Galaxy workflow for annotating and characterising gene families using GeneSeqToFamily (Thanki et al. GigaScience 2018) and visualising their relationships using Aequatus (Thanki et al. GigaScience 2018).

GeneSeqToFamily is a Galaxy workflow based on the Ensembl GeneTrees pipeline, and generates gene families, providing sequence alignments, associated phylogenetic trees, as well as details about exon conservation. The workflow helps users to run large-scale gene family analyses without requiring command-line usage while still allowing flexibility in parameter configuration, and tools used.

Aequatus is a standalone web-based tool that provides an in-depth view of gene structure across gene families, including gene order and protein domains. It is now also available as a visualisation plugin within the Galaxy web platform to visualise gene trees generated by GeneSeqToFamily.

All tools used in the GeneSeqToFamily workflow and the full workflow itself are available within the Galaxy ToolShed. GeneSeqToFamily is also available for use from the usegalaxy.eu public instance. Source code for all the tools is available on GitHub.

Phylogenomics under the multispecies coalescent
SMBE-PO-559
Phylogenomics of biotrophy in Oomycetes.
K. Fletcher [1,*], R. Michelmore [1]
[1]Genome Center, University of California, Davis, United States

Abstract: Oomycetes are fungal-like organisms, which are phylogenetically and biochemically distinct from true fungi. Biotrophy, the dependence of a pathogen on a living host, has arisen multiple times within the order Peronosporales, class Oomycota. Downy mildews are important biotrophic plant pathogens and distinct from the monophyletic, biotrophic family Albuginaceae. Downy mildews have recently been reported as being polyphyletic, sharing common ancestors with hemibiotrophic, paraphyletic *Phytophthora* species that switch from biotrophy to necrotrophy during the course of infection. Other studies, however, have suggested a monophyletic origin for the downy mildews. We tested the hypothesis of polyphyly of downy mildews using comparative phylogenomics. Both maximum likelihood analysis of concatenated alignments and multispecies coalescence support polyphyly of the downy mildews and paraphyly of *Phytophthora* species. Adaptation to biotrophy has been reported to correlate with loss of genes associated with transport, carbohydrate binding, and pathogenicity. We are determining which genes are significantly depleted in downy mildews in order to identify oomycete orthologs that have been lost in parallel adaptations to biotrophy.

Phylogenomics under the multispecies coalescent

SMBE-PO-548

Bayesian model comparison of population histories under the multi-species coalescent

I. Gronau [1,*], R. Visbord [1], D. E. Chavez [2], R. K. Wayne [2]

[1]Computer Science, Herzliya Interdisciplinary Center, Herzliya, Israel, [2]Ecology and Evolutionary Biology, University of California, Los Angeles, Los Angeles, United States

Abstract: The past decade has shown a rise in the use of sophisticated computational demography inference methods based on the multi-species coalescent. Efficient MCMC sampling algorithms enable Bayesian estimation of parameter values in complex demographic models for data sets with up to a dozen populations. A typical model in such a scenario may consist of dozens of parameters capturing divergence times, effective population sizes, and rates of post-divergence gene flow. Despite the progress made in Bayesian demography inference methods, little progress has been made in the fundamental task of comparing different hypotheses regarding the assumed structure of the population phylogeny and scenario of gene flow. This is unfortunate, because often times the most interesting questions regarding the population history of a newly studied species involve the structure of the phylogeny and migration edges and not the particular values of the demographic parameters. We present here a new Bayesian method for comparing demographic models, and we use it to resolve open questions regarding phylogenetic relationships between diverged canid species.

Bayesian demography inference methods based on the multi-species coalescent typically assume a hypothesized structure for a demographic model, M_{hyp}, which is represented by a population phylogeny augmented with migration edges. These methods use MCMC to sample parameter values from an approximate posterior distribution, but this sampling procedure does not directly provide a reliable measure for model fit. Our new Bayesian method for model comparison addresses this challenge by utilizing importance sampling to estimate the Bayes factor of M_{hyp} relative to some reference model M_{ref}. The use of **relative Bayes factors (RBFs)** allows us to improve the accuracy of the comparison by modifying the selection of M_{ref}. We implemented a method for computing RBFs in the Generalized Phylogenetic Coalescent Sampler (*G-PhoCS;* Gronau *et al.* 2011) and used it to examine open questions in canid evolution.

Large ancestral population sizes and high rates of gene flow have led to unresolved phylogenetic relationships between several diverged canids. We examine some of these relationships using RBFs computed from public whole genome sequence data from nine canid species. Our analysis shows that, unlike claimed in recent studies, the Golden Jackal is an outgroup to all wolves, including the deeply diverged Golden Wolf. We also show that multiple episodes of gene flow contribute to the high level of genomic similarity observed between the Golden Jackal and Gray Wolf. We use this analysis to demonstrate the challenges scientists face when reconstructing complex demographic histories for newly studied species.

Phylogenomics under the multispecies coalescent

SMBE-PO-549

Exact likelihoods from the generalized Ewens Sampling Formula

M. Uyenoyama [1,*], N. Takebayashi [2], S. Kumagai [1]

[1]Biology, Duke University, Durham, [2]Institute of Arctic Biology, University of Alaska, Fairbanks, United States

Abstract: The celebrated Ewens Sampling Formula (ESF) provides the exact probability of a sample, the likelihood of a model. In its original form, the ESF addressed unstructured populations under the infinite-alleles model. We have developed an inductive method that extends the ESF to accommodate population structure and other forms of mutation. Our method uses a labeled coalescent argument, which differs from the usual coalescent in its use of information contained in the present state (allele and location) of a lineage. Among the forms of population structure we have addressed are regular inbreeding and subdivision of the population into demes, allowing for migration among demes. We also discuss generalized models of base substitution, including non-reversible models.

Phylogenomics under the multispecies coalescent

SMBE-PO-557

IQ-TREE-POMO: Polymorphism-aware tree estimation

D. Schrempf [1 2], B. Q. Minh [3 4], R. Borges [5], C. Kosiol [5 6,*]

[1]Department of Biological Physics, Eötvös Loránd University, Budapest, Hungary, [2]Center for Integrative Bioinformatics Vienna, Max F. Perutz Laboratories, Vienna, Austria, [3]Ecology and Evolution, [4]Research School of Computer Science, Australian National University, Canberra, Australia, [5]Institut für Populationsgenetik, Vetmeduni Vienna, Vienna, Austria, [6]Centre for Biological Diversity, University of St Andrews, St Andrews, United Kingdom

Abstract: Molecular phylogenetics has neglected polymorphisms within present and ancestral populations for a long time. Recently, multispecies coalescent (MSC) based methods have increased in popularity, however, their application is limited to a small number of species and individuals. We have introduced an alternative approach called polymorphism-aware phylogenetic model (PoMo), which overcomes this limitation and scales well with the increasing amount of sequence data. PoMo circumvents handling of gene trees and directly infers species trees from allele frequency data. PoMo extends any DNA substitution model and additionally accounts for polymorphisms in the present and in the ancestral population by expanding the state space to include polymorphic states. It is a selection-mutation model which separates the mutation process from the fixation process. PoMo naturally accounts for incomplete lineage sorting because ancestral populations can be in a polymorphic state. Our method can accurately and time-efficiently estimate the parameters describing evolutionary patterns for phylogenetic trees of any shape (species trees, population trees, or any combination of those).

We have implemented our PoMo approach as software package IQ-TREE-POMO with several new features: (i) a search for the statistically best-fit mutation model (ModelFinder), (ii) the ability to allow mutation rate variation across sites (e.g., gamma distribution), assessment of branch support values (bootstrapping and jackknifing), (iv) simulator of sequences evolving under PoMo (bmm-simulate), and (v) inference of allelic selection.

Applications comparing the MSC and PoMo approaches on simulated and on real great ape data sets will be presented. In particular, the new genome-wide data set of seven baboon populations (genus *Papio*) present a unique opportunity to apply our method to a primate clade that involves more complex processes than those usually assumed by phylogenetic models. The history of *Papio* includes episodes of introgression or admixture among genetically distinct lineages. We will discuss the effect of this complex history on genome-wide phylogenetic inference with PoMo as well as other approaches.

Phylogenomics under the multispecies coalescent

SMBE-PO-551

Phylogenomics on universal nuclear targets under the MSC decode contentious relationships across all angiosperm families

L. Pokorny [1,*], S. Dodsworth [2], M. G. Johnson [3], O. Maurin [1], N. Epitawalage [1], S. Bellot [1], L. R. Botigué [4], G. Brewer [1], M. W. Chase [5], R. S. Cowan [5], W. L. Eiserhardt [6], M. de la Estrellla [7], O. Hidalgo [8], P. Kersey [5], J. T. Kim [9], I. Larridon [10], B. Murphy [1], J. Pellicer [1], J. M. de Vos [11], N. J. Wickett [12], I. J. Leitch [1], F. Forest [5], W. J. Baker [1]

[1]Comparative Plant and Fungal Biology, Royal Botanic Gardens, Kew, Richmond, [2]School of Life Sciences, University of Bedfordshire, Luton, United Kingdom, [3]Biological Sciences, Texas Tech University, Lubbock, United States, [4]Plant and Animal Genomics, Center for Research in Agricultural Genomics, Cerdanyola, Spain, [5]Royal Botanic Gardens, Kew, Richmond, United Kingdom, [6]Bioscience, Aarhus University, Aarhus, Denmark, [7]Botany, Ecology, and Plant Physiology, University of Córdoba, Córdoba, [8]Pharmacy, University of Barcelona, Barcelona, Spain, [9]Biodiversity Informatics and Spatial Analysis, [10]Identification and Naming, Royal Botanic Gardens, Kew, Richmond, United Kingdom, [11]Environmental Sciences, University of Basel, Basel, Switzerland, [12]Plant Science and Conservation, Chicago Botanic Garden, Glencoe, United States

Abstract: In here we describe the first (order- and) family-complete phylogenomic analysis of angiosperms. We demonstrate that the angiosperms 353 kit we developed works to enrich 353 orthologs for all angiosperm families. Additionally, we recover organellar off-target coding regions as well as non-coding introns, the latter used to test our kit at shallow levels, i.e., genera and species. Our phylogenomic pipeline explores effects of alignment, trimming, partitioning, gene tree inference, species tree inference, on our reconstruction. Recalcitrant nodes are explored in light of recent support measures to discuss robustness of alternative contentious relationships and present a backbone hypothesis for the flowering plant tree of life.

Phylogenomics under the multispecies coalescent

SMBE-PO-552

Conservation genomic analysis of Madagascar's hibernating dwarf lemurs reveals unexpected patterns of ancient introgression

R. C. Williams*, M. Blanco, J. Poelstra, K. Hunnicutt, A. Comeault, A. Yoder

Abstract: Madagascar's biodiversity is notoriously threatened by human-mediated deforestation and climate change. And though we would ideally utilize standard practices for measuring genetic diversity within Malagasy species and populations, many of these organisms are rare, cryptic, and severely threatened, making population-level sampling unrealistic. Such is the case with Madagascar's dwarf lemurs (genus *Cheirogaleus*), the world's only obligate hibernator within the primates. Given that genetic sampling is limited to only a few or even single individuals, we have turned to comparative genomic approaches for understanding the potential impacts of inbreeding and low levels of heterozygosity in these endangered primates. We generated a reference genome for the dwarf lemur, *Cheirogaleus medius,* and used this to facilitate analyses of high-coverage (~30x) genomes for wild-caught individuals representing four dwarf lemur species - *Cheirogaleus sp. cf. medius, C. major, C. crossleyi* and *C. sibreei*- rendering this study the largest contribution to date of novel genomic resources for Madagascar's lemurs.We show that the four lineages are genomically distinct as would be expected of their relatively ancient divergence time, estimated to have been roughly 6 - 23 Ma, though surprisingly, ancient admixture has occurred between species during early stages of lineage divergence. Introgressed regions contain genes associated with hibernation, though most significantly, show over-represented gene ontology categories relating to transcription. We tested contemporary levels of heterozygosity finding that the lowest levels are seen in an individual sampled from a disjunct population that we refer to as *C. sp. cf. medius*. Our study highlights the power of comparative genomic analysis for identifying species of conservation concern, as well as for illuminating possible mechanisms of adaptive phenotypic evolution - in this case, hibernation.

Phylogenomics under the multispecies coalescent

SMBE-PO-554

Phylogenomics reveals an ancient hybrid origin of the Persian walnut

B.-W. Zhang [1], K. Lin [1], S. S. Renner[*], D.-Y. Zhang [2], W.-N. Bai [1]

[1]College of Life Sciences, [2]College of Life Science, Beijing Normal University, Beijing, China

Abstract: Persian walnut (*Juglans regia*) is cultivated worldwide for its high-quality wood and nuts, but its origin has remained mysterious because in phylogenies it occupies an unresolved position between American black walnuts and Asian butternuts. Equally unclear is the origin of the only American butternut, *J. cinerea*. Using whole-genome sequencing and re-sequencing of 80 individuals from 19 of the 22 species of *Juglans*and its relatives *Pterocarya stenoptera*and *Platycarya strobilacea*,and phylogenetic-network analysis of single-copy nuclear genes, genome-wide site pattern probabilities, and Approximate Bayesian Computation (ABC) we discovered that *J. regia*(and its landrace *J. sigillata*) arose as a hybrid between the American and the Asian lineages and that *J. cinerea* resultedfrom massive introgression from an immigrating Asian butternut into the genome of an American black walnut.ABCmodelling placed the hybrid origin in thelate Pliocene,~3.45 Ma, with both parental lineages since having gone extinct in Europe.

Phylogenomics under the multispecies coalescent

SMBE-PO-561

High levels of gene tree discordance partially explain accelerated rates of molecular evolution in tarsiers

G. P. Tiley [1,*], C. Oglesby [1], A. D. Yoder [1]

[1]Department of Biology, Duke University, Durham, United States

Abstract: Well-assembled genomes have supported the placement of tarsiers as sister to monkeys and apes, the Haplorhini hypothesis, with certainty. However, the position of tarsiers was a contentious node in early molecular phylogenetic literature. We used genomic data to interrogate the both the biological processes and technical artifacts underlying this contention. With approximately 7000 nuclear protein-coding alignments, we showed that both supermatrix and coalescent methods are consistent with the Haplorhini hypothesis. Phylogenomic methods recover well-resolved species trees despite high levels of gene-tree discordance. Gene tree discordance was evident in the form of ambiguous gene trees that do not clearly favor any one hypothesized topological position of tarsiers, and strongly conflicting gene trees that rejected Haplorhini. However, simulations under the multispecies coalescent model suggested that observed levels of gene tree discordance were consistent with expectations of incomplete lineage sorting. Additionally, accounting for gene tree discordance under the multispecies coalescent model reduced the disparity in absolute rates of molecular evolution in tarsiers relative to anthropoids in comparison to assuming a single shared topology across genes. Through sampling concordant and discordant gene trees, we showed that increased rates of molecular evolution in the tarsier lineage are due, at least in part, to additional substitutions implied by discordant gene trees. Our results demonstrate the importance of considering incomplete lineage sorting for accurate substitution rate estimation in cases of short internal branches, even when divergences are relatively ancient.

Phylogenomics under the multispecies coalescent

SMBE-PO-555

Inference of gene flow in the process of speciation: an efficient maximum-likelihood implementation of a "generalised isolation-with-migration model".

R. J. Costa [1], H. M. Wilkinson-Herbots [2,*]

[1]European Bioinformatics Institute (EMBL-EBI), Hinxton, [2]Department of Statistical Science, University College London, London, United Kingdom

Abstract: The "isolation with migration" (IM) model has been extensively used in the literature to detect gene flow during the process of speciation. In this model, an ancestral population split into two or more descendant populations which subsequently exchanged migrants at a constant rate until the present. Of course, the assumption of constant gene flow until the present is often over-simplistic in the context of speciation. In this paper, we consider a "generalised IM" (GIM) model: a two-population IM model in which migration rates and population sizes are allowed to change at some point in the past. By developing a maximum-likelihood implementation of this model, we enable inference on both historical and contemporary rates of gene flow between two closely related populations or species. The GIM model encompasses both the standard two-population IM model and the "isolation-with-initial-migration" (IIM) model as special cases, as well as a model of secondary contact. Our method makes it possible to distinguish between such different variants of the model (representing alternative evolutionary scenarios) by means of likelihood ratio tests or AIC scores. Our ML method is based on the coalescent and is suitable for data sets consisting of the number of nucleotide differences between one pair of DNA sequences at each of a large number of independent loci. As our method relies on an explicit expression for the likelihood, it is very fast – fitting a GIM model to a data set consisting of thousands of loci typically takes just a few minutes on a personal computer.

Phylogenomics under the multispecies coalescent

SMBE-PO-558

Performance of a priori and a posteriori calibration strategies in divergence time estimation

A. J. S. Beavan [1,*], P. C. J. Donogue [2], M. A. Beaumont [1], D. Pisani [1 2]

[1]School of Biological Science, [2]School of Earth Science, University of Bristol, Bristol, United Kingdom

Abstract: The molecular clock is an extension of the multispecies coalescent that allows us to estimate the divergence times of genes, populations, species or larger phylogenetic groups. One challenge faced by biologists that use molecular clocks is estimating divergence times when the rate of evolution is not constant. To circumvent this problem, models of lineage specific rate variation are often implemented in a Bayesian framework. These models, accompanied with prior information about the age of nodes, usually gleaned from the fossil record, allow the clock to be relaxed, and the detection of different evolutionary rates across lineages. Alternatively, calibrations can be used *a posteriori,* to transform previously estimated relative divergence times that were inferred without considering fossil information, into absolute divergence times, such as in the non-Bayesian method, RelTime. However, as branch length is the product of the rate of evolution and the duration in time of the considered branch, the extent to which *a posteriori* calibration can disambiguate time and rate, is unclear. Here, we compare different molecular clock methods and models, and *a priori* and *a posteriori* calibration strategies, using forward simulations of evolution. Specifically, we compare three Bayesian methods, the strict clock, uncorrelated clock and autocorrelated clock, and the non-Bayesian method RelTime. Our simulations reveal that relative divergence times and *a posteriori* calibration strategies almost invariably inferred incorrect rate changes, and divergence times. The a *priori* integration of fossil calibrations is fundamental to improve the accuracy of the estimated divergence times. Relative divergence times, and absolute timescales derived by calibrating relative timescales to geological time *a posteriori*, appear to be less reliable than standard, *a priori* calibrated, timescales that should therefore be preferred.

Phylogenomics under the multispecies coalescent

SMBE-PO-556

Linked-Reads Provide New Reference-Genomes for Five Species of Falcons

J. J. S. Wilcox*, B. Arca Ruibal, N. Drou, V. D. Mateuta, J. H. Samour, Y. Idaghdour, S. Boissinot [1]

[1]Biological Sciences, New York University Abu Dhabi, Abu Dhabi, United Arab Emirates

Abstract: Falcons (Genus *Falco*) represent a diverse group of birds of substantial economic, cultural, and cconservation concern. Despite their importance, these birds present an evolutionary puzzle: the large falcons of falconry have emerged and diversified from a single common ancestor to approximately two-dozen ecologically and morphologically distinct lineages, in the absence of reproductive isolation and in the span several hundred-thousand years. Humans have further contributed to the complexity of these events, both through intentional and unintentional practices emerging from falconry, and through conservation efforts in response to the past decimation of some falcon populations by DDT and other pesticides. Previous genome-wide approaches have also highlighted several intrinsic peculiarities of falcon genomes, including chromosome-fusion events that are uncharacteristic of birds, extremely low levels of microdeletions (<30 bp), record levels of mitochondrial DNA insertions into the nuclear genomes of a bird, and duplications within the mitogenome. New genomic tools now provide an opportunity explore the genomic architecture of falcons in unprecedented detail: illuminating the processes underlying the evolution of falcons and providing broad insights into the developments that drive distinctions between species at the molecular level. Here, we provide a first look at new reference genomes for gyr (*F. rusticolus*), peregrine (*Falco peregrinus*) and saker (*F. cherrug*) falcons, and first-ever reference genomes for two other falcon species—the barbary falcon (*F. peligrinoides*), the lanner falcon (*F. biarmicus*). Through the use of 10X Genomics Chromium linked-reads these reference genomes drastically improve on previous genome assemblies from falcons and provide new insights into the genetic architecture of speciation in these birds. Ultimately, these reference genomes will lay the groundwork for future studies on speciation and genome evolution in falcons, and help to inform applications relating to falconry and falcon conservation.

Repeats and mobile elements as drivers of innovations in protein coding genes

SMBE-PO-571B

Who is in the driver's seat? The impact of poly-A microsatellite heterologies in meiotic recombination hotspots

A. Heissl [1], A. Betancourt [2], P. Hermann [1], G. Povysil [1], A. Futschik [1], T. Ebner [3], I. Tiemann-Boege [1,*]

[1]Johannes Kepler University Linz, Linz, Austria, [2]University of Liverpool, Liverpool, United Kingdom, [3]Kepler University Clinic, Johannes Kepler University Linz, Linz, Austria

Abstract: Meiotic recombination has strong, but poorly understood effects on short tandem repeat (STR) instability. Here, we screened thousands of single recombinant products with sperm typing to characterize the role of polymorphic poly-A repeats at a human recombination hotspot in terms of hotspot activity and STR evolution. We show that the length asymmetry between heterozygous poly-As strongly influences the recombination outcome: a heterology of 10 As (9A/19A) reduces the number of crossovers and elevates the frequency of non-crossovers, complex recombination products, and long conversion tracts. Moreover, the length of the heterology also influences the STR transmission during meiotic repair with a strong and significant insertion bias for the short heterology (6A/7A) and a deletion bias for the long heterology (9A/19A). In spite of this opposing insertion/deletion biased gene conversion, we find that poly-As are enriched at human recombination hotspots that could have important consequences in hotspot activation.

Repeats and mobile elements as drivers of innovations in protein coding genes
SMBE-PO-571A
new genes generated and amplified by transposons in animals
S. Tan*

Abstract: Transposable elements are mobile genetic units ubiquitous in various organisms and a major force to influence the genomic architectures. A key contribution of TEs to evolution is their capacity to mediate the formation of new genes. According to their transposition mechanisms, TEs can be classified into two groups: retrotransposons (class I) and DNA transposons (class II). Our first work （Genome Research, 2016） shows that LTR retrotransposons can mediate the generation of retrogenes in animals. The recently originated retrocopies have a similar chimeric structure: the internal retrocopies are flanked by discontinuous LTR retrotransposons. At the fusion points we identified shared short similar sequences, suggesting the involvement of microsimilarity-dependent template switches at RNA level. Our second work (unpublished) shows DNA transposons can also mediate the generation of functional genes in animals. We identified genes with the similar chimeric structure in various animals. Deep analysis on one gene in *Drosophila melanogaster* shows that it was successively amplified to multiple copies in populations. This gene is preserved by selection sweep and highly expressed in midgut, which suggests it acquires certain function. Combined with the HiC-seq data, we proposed a model involving template switches during DNA replication process to explain the formation of these new genes. Both mechanisms at RNA and DNA level are conserved across a wide range of animal taxa, which represents ancient and ongoing mechanisms mediated by transposons continuously shaping gene content evolution in animals.

Repeats and mobile elements as drivers of innovations in protein coding genes
SMBE-PO-566

Transposable elements as drivers of intra-specific evolution of genome architecture

T. Badet [1,*], U. Oggenfuss [1], L. Abraham [1], B. A. McDonald [2], D. Croll [1]
[1]Laboratory of Evolutionary Genetics, University of Neuchâtel, 2000, [2]Institute of Integrative Biology, ETH Zurich , 8092, Switzerland

Abstract: Intra-specific genetic diversity is a major driver of adaptive evolution. In pathogens, rapid host adaptation often emerges from standing genetic variation. The recent advent of whole genome sequencing techniques has highlighted the primary role of structural variation in shaping within-species genetic diversity. Structural variation involves any rearrangement that leads to deletion, insertion, duplication, inversion and/or translocation of a chromosomal sequence. The consequences of genomic structural rearrangements can be deleterious for coding sequences, but can also lead to evolutionary novelty. However, while single nucleotide polymorphisms have been extensively studied for their contributions to adaptation, the contribution of structural variation remains largely unknown. In addition, little is known about the impact of genome sequence composition on structural variation.

In this study, we generated a map of genome-wide structural variation in a highly polymorphic fungal pathogen of wheat. Zymoseptoria tritici has a global distribution and its emergence is tied to the domestication of wheat. By analyzing chromosome-level assemblies of 20 individuals from six different continents, we first show that substantial genomic rearrangements occurred during the evolutionary history of this fungal pathogen. We then constructed the pangenome of core and accessory genes and show that more than a third of all orthologous gene sets segregate variation in presence and absence within the species. In particular, candidate genes for promoting disease also called effector genes are enriched in the accessory genome. Finally, we analysed proximate mechanisms driving the evolution of the accessory genome. We found that transposable elements are over-represented in regions harbouring singleton genes and small insertion/deletion events.

Histone methylation is known to control genome structure and stability. Here we show that the H3K4 methylation pattern associated with euchromatin is negatively correlated with structural variation events and the presence of transposable elements. In contrast, H3K9 and H3K27, both associated with heterochromatin, are positively correlated with the presence of structural variation. We find that different types of structural variation show associations with transposable element families differing in expression and evolutionary age. Altogether, our study identifies the mechanistic basis for functionally relevant structural variation in an eukaryotic plant pathogen.

Repeats and mobile elements as drivers of innovations in protein coding genes

SMBE-PO-562

Variation in the genomic landscape of repetitive elements across Caenorhabditis nematodes

G. Woodruff [1,*]

[1]University of Oregon, Eugene, OR, United States

Abstract: The abundance, diversity, and genomic distribution of repetitive elements is highly variable among species. In addition to being influenced by population size and reproductive mode, these patterns are thought to be driven by the interaction of selection and recombination, as recombination rates typically vary by chromosomal position. In the nematode *C. elegans*, repetitive elements are enriched at chromosome arms and depleted on centers, and this mirrors the chromosomal distributions of other genomic features. How conserved is this genomic structure of repeats, and what evolutionary forces maintain it? To address this, I compared the genomic structure of repetitive elements across five *Caenorhabditis* species with chromosome-level assemblies. Repeat content is enriched in regions of high recombination in most *Caenorhabditis* species. In contrast, the fig-associated *Caenorhabditis inopinata* has experienced rampant repetitive element expansion and reveals no association of global repeat content with chromosome position. Patterns of repeat family-specific distributions reveal this global pattern is driven largely by a few repeat families that in *C. inopinata* have either expanded in number, have weak associations with chromosome position, or both. Additionally, ~15% of predicted protein-coding genes in *C. inopinata* align to transposon proteins. Population genomics reveals genetic diversity is about five times higher in the gonochoristic *C. inopinata* than in the hermaphroditic *C. elegans*; the genomic distribution of diversity in multiple species is also consistent with conserved recombination rate variation along chromosomes in this group. Thus, population size and reproductive mode alone cannot be driving patterns of repeat abundance in this group. Repeat family-specific variation in insertion rate, deletion rate, and/or fitness effects are likely influencing the evolution of the genomic repeat landscape in *Caenorhabditis*, although demography and recombination rate variation cannot be ruled out entirely. Taken together, these results highlight the power of comparative genomics in testing hypotheses regarding the causes of genome architecture.

Repeats and mobile elements as drivers of innovations in protein coding genes

SMBE-PO-568

Genome assemblies of 25 Antarctic Notothenioid fish species illuminate causes and consequences of adaptive radiation, and provide insights into transposon regulation within this iconic fish group.

I. Bista [1 2 3,*], S. McCarthy [2], M. Clark [4], T. Desvignes [5], J. Postlethwait [5], C. -.-H. C. Cheng [6], H. W. Detrich [7], W. Salzburger [8], Z. Ning [1], W. Chow [1], J. Wood [1], K. Howe [1], E. Miska [1 3], R. Durbin [1 2]

[1]Wellcome Sanger Institute, Sanger Institute, [2]Department of Genetics, University of Cambridge, [3]Wellcome Trust/Cancer Research UK , Gurdon Institute, [4]British Antarctic Survey, Cambridge, United Kingdom, [5]Institute of Neuroscience, University of Oregon, Eugene, [6]Department of Animal Biology, University of Illinois at Urbana - Champaign, Urbana - Champaign, [7]Department of Marine and Environmental Sciences, Northeastern University, Nahant, United States, [8]Zoological Institute, University of Basel, Basel, Switzerland

Abstract: The Notothenioid radiation of Antarctic fish presents one of the most dramatic examples of marine adaptive radiations, with over 120 species dominating the Southern Ocean in fish species richness and biomass. At the Wellcome Sanger Institute and in collaboration with the Vertebrate Genomes Project (VGP), we are sequencing 25 species representing 5 Notothenioid families, using long read sequencing technologies. Assembling using long read sequencing will allow study of highly repetitive gene families (such as the antifreeze glycoproteins or AFGPs) and detection of structural variation, whilst providing deep understanding of genome evolution in this iconic marine radiation.

We have generated a high-quality reference assembly for the channel bull blenny, *Cottoperca gobio* from a combination of PacBio, 10X Chromium, BioNano optical mapping and Hi-C technologies. This assembly has been assigned to 24 chromosomes, achieving VGP quality standards (assembly size 609Mb, with contig N50 5Mb and scaffold N50 14.74 Mb), and is currently submitted for gene annotation in Ensembl. Furthermore, we have generated draft reference assemblies for 4 additional species: *Trematomus bernachii, Harpagifer antarcticus, Gymnodraco acuticeps,* and *Pseudochaenichtys georgianus* (PacBio at >50X coverage and 10X Chromium), as well as 12 Supernova (10X Chromium), and 8 Soap-denovo HiSeqX Illumina assemblies from species across the radiation.

Comparative analysis demonstrated a notable increase in genome size from 609Mb in basal Bovichtidae species to over 1Gb in members of the derived white-blooded icefish Channichthidae family. This increase in genome size is linked to a substantial expansion of repetitive elements, mainly DNA transposons and LINE1/Gypsy retrotransposons. Germline transposition is regulated through the Piwi/piRNA small regulatory RNA pathway, so we have generated smallRNA transcriptome data for multiple species and tissues, in order to study the evolution of transposon control across the radiation. Comparison of expression levels of piRNAs across 16 Notothenioid species, in relation to specific transposons and their abundance, will enable investigation of mechanisms controlling the genome size expansion in this group. Additionally, we are investigating the evolutionary relationships of the highly repetitive antifreeze and globin gene families between different species. Overall this study will provide a deep genomic characterization of this iconic fish group, and an important platform to investigate the mechanisms of fish genome evolution.

Repeats and mobile elements as drivers of innovations in protein coding genes
SMBE-PO-569
Tandem repeats contribute to coding sequence variation in bumblebees (Hymenoptera: Apidae)
C. Sun*, S. Schaack, X. Zhao, L. Su, W. Xu

Abstract: Tandem repeats (TRs) are highly dynamic regions of the genome. Mutations at these loci represent a significant source of genetic variation and can facilitate rapid adaptation. Bumblebees are important pollinating insects occupying a wide range of habitats.However, to date, molecular mechanisms underlying the potential adaptation of bumblebees to diverse habitats are largely unknown.In the present study, we investigate how TRs contribute to genetic variation in bumblebees, thus potentially facilitating adaptation.We identified 26,595 TRs from the assembled 18 chromosome sequences of the buff-tailed bumblebee (*Bombus terrestris*), 66.7% of which reside in genic regions. We also compared TRs found in *B. terrestris* with those present in the assembled genome sequence of a congener, *B. impatiens*. We found that a total of 1,137 TRs were variable in length between the two sequenced bumblebee species, and further analysis reveals that 101 of them are located within coding regions. These 101 TRs are responsible for coding sequence variation and correspond to protein sequence length variation between the two bumblebee species. The variability of identified TRs in coding regions between bumblebees was confirmed by PCR amplification of a subset of loci. Functional classification of bumblebee genes where coding sequences include variable-length TRs suggests that a majority of genes (87%) that could be assigned to a protein classare related to transcriptional regulation. To understand if length variations in those protein sequences could lead to functional consequences, we made antibodies for three of those proteins, which are transcription factors, and used Chip-seq technique to detect their bindings sites in couple of bumblebee species. Results indicated that length variation in the coding regions of genes encoding transcription factor could lead to different binding sites. Our results show that TRs contribute to coding sequence variation in bumblebees, and thus may facilitate the adaptation of bumblebees through diversifying proteins involved in controlling gene expression.

Repeats and mobile elements as drivers of innovations in protein coding genes

SMBE-PO-567

Tandem repeats are selfish elements which mark the level of hidden recombination in animal mitochondrial genomes

A. A. Mikhailova [1,*], K. Ushakova [1], A. G. Milhaylova [1], V. Lobanova [1], P. Kravchenko [2], I. Mazunin [1], D. Knorre [3], A. Reymond [4], K. Gunbin [5], K. Popadin [1 4 6]

[1]Center for Mitochondrial Functional Genomics, Immanuel Kant Baltic Federal University, Kaliningrad, [2]faculty of bioengineering and bioinformatics, [3]Belozersky Institute Of Physico-Chemical Biology, Lomonosov Moscow State University, Moscow, Russian Federation, [4]Center for Integrative Genomics, University of Lausanne, Lausanne, Switzerland, [5]The Institute of Cytology and Genetics of the SB RAS, The Institute of Cytology and Genetics of the SB RAS, Novosibirsk, Russian Federation, [6]Swiss Institute of Bioinformatics, Lausanne, Switzerland

Abstract: Despite the fact that mitochondrial genome is streamlined, its size still varies significantly across all vertebrate taxa (from 15kb in Mammalia to 25kb in Reptilia approximately). Analysing 4000 complete mitochondrial genomes of chordata species we observed that: (i) genome length variation is explained mainly by variation in control region, which is driven by the abundance of tandem repeats; (ii) tandem repeats are more common in short-lived species; (iii) long, GC-rich, low-degraded and high-copy-number motifs of tandem repeats correlate better with generation time. There are three non-mutually exclusive potential explanations of the observed negative correlation between the tandem repeats abundance and generation time: (i) strong negative selection against tandem repeats in mtDNA of long-lived species (within long-lived dormant oocytes), (ii) selfish drive of tandem repeats (H. Ma et al., 2016) and (iii) neutral propagation of tandem repeats in mtDNA of species with non-zero level of recombination. According to our additional results the third explanation is the most plausible one. Tandem repeats (especially with long motifs) may propagate better in the recombinational environment and thus they might be a neutral marker of the hidden recombination. It is commonly accepted that mtDNA of chordata are mainly non-recombining, maternally inherited genome. However, numerous well-proofed exceptions are known from this rule. We hypothesize, that tandem repeats can be a marker of the low level of recombination, undetectable by many other methods.

Resistance evolution in real-time

SMBE-PO-572

Striking back at Beta-lactams with a new drug combination

S. Garsevanyan [1,*], M. Barlow [1]

[1]School of Natural Sciences , University of California, Merced, Merced, United States

Abstract: We lack fundamental understanding of the relationship between gene expression and fitness. Antibiotic resistance system is an ideal model for studying this relationship because there are many known resistance genes with strong fitness effect when antibiotics are present. However, as simple as this relationship may seem, identifying the presence of resistance genes is not a reliable way of predicting fitness. I am examining the relationship between resistance and fitness. Working with the community based hospital in Merced, Dignity Health Mercy Medical Center, we have collected over 900 patient microbial isolates. Using these samples, I am investigating genes and their fitness, to identify their resistance genes. Growth rates have been a very valuable in obtaining the fitness of these strains. The latest and newest inhibitor for antibiotics is Avibactam, which is a non-beta-lactam used with Ceftazidime. This new drug combination works against multidrug-resistant gram-negative bacteria. Gram negative organisms have been identified to be among the most serious threats. I have found multiple strains in this collection with elevated resistance to the Avibactam/Ceftazidime drug combination. Minimum inhibitory concentrations have shown 30 isolates at 0.000975, 11 isolates at 0.0039, 2 isolates at 0.0156, 1 isolate at 0.03125, and 3 isolates at 0.0625. This gives us a better understanding of how this drug combination is working against highly resistant bacteria strains.

Resistance evolution in real-time

SMBE-PO-574

Challenging the convention wisdom on HIV multi-drug resistance evolution

A. F. Feder [1,*], P. S. Pennings [2]

[1]Department of Integrative Biology, University of California, Berkeley, Berkeley, [2]Department of Biology, San Fransisco State University, San Francisco, United States

Abstract: The first drugs used to treat HIV failed quickly and predictably due to drug resistance evolution, somtimes in a matter of weeks. Combination therapies of three simultaneous drugs drastically reduced HIV's ability to evolve drug resistance to treatment. The conventional reasoning of why these combinations work so well - that a single viral genome should need three drug resistance mutations in order to replicate - fails to account for patterns present in ongoing resistance evolution. We can better understand why combination therapies work by studying the instances in which they fail. Here, we present evidence for an alternative explanation for how drug resistance continues to evolve in response to combination therapies. Drug penetrance into tissues is heterogeneous and spaces with poor drug coverage act as refugia where resistance to single drugs can be selected. In support of this hypothesis, we show that HIV populations treated with combination therapies acquire mutations to specific drugs in a semi-predictable order. This ordering often (but not always) matches our expectations from drug penetrance profiles into different tissues. This suggests that HIV drug resistance prevention requires not only combining orthogonal drugs, but additional synergy among the profiles of such drugs in time and space.

Resistance evolution in real-time

SMBE-PO-586

From vaccine to pathogen: modeling the evolutionary epidemiology of circulating vaccine-derived poliovirus

W. Wong [1,*], M. Famulare [1]

[1]Institute for Disease Modeling, Bellevue, United States

Abstract: Mass vaccination campaigns with oral polio vaccine (OPV) have been a key factor for reducing the burden of paralytic poliomyelitis and led to the eradication of wild poliovirus (WPV) type 2. While the Sabin strains in OPV exhibit highly-attenuated virulence and reduced infectivity relative to WPV,[MF1] they remain transmissible in human populations. During prolonged transmission, Sabin-derived polioviruses can regain wild phenotype and cause circulating vaccine-derived poliovirus outbreaks (cVDPV) with similar disease burden to those caused by WP. Understanding the emergence and spread of virulent cVDPV is crucial for preventing future vaccine-derived paralytic poliomyelitis, but requires the integration of both epidemiology and population genetic models. Here, we infer the expected transmission fitness of seven adaptive gateway mutations (A481G, U398C, U209C, U2523C, C2006A, U1376A, U3320A) under a modified Wright Fisher framework that allows for expansion within and transmission between human hosts. These mutations are involved in a mutational pathway for attenuation reversal, whose order and fixation times are similar across a wide variety of epidemiological settings. We also estimate the expected genetic load of maladaptive mutations accumulated during circulation. We then jointly model transmission epidemiology and evolution to quantify how advantageous and deleterious mutations interact to determine the fixation dynamics of reverted WPV-like strains following Sabin OPV vaccination. Our study combines observations collected from extensive disease surveillance and genomic epidemiology to develop a comprehensive evolutionary framework for vaccine-derived poliovirus. Although our study focuses on attenuation reversion, our methods are broadly applicable to other systems where highly adaptive mutations can occur, such as those involved with drug resistance.

Resistance evolution in real-time

SMBE-PO-582

Investigating the Fitness Costs of Mobile Colistin Resistance Genes

L. Ogunlana [1,*], C. MacLean [1], S. Uphoff [2]

[1]Zoology, [2]Biochemistry , The University of Oxford, OXFORD, United Kingdom

Abstract: The antibiotic resistance crisis is a worldwide problem and one of the major scientific focuses currently. Recently, plasmid mediated mobile resistance to colistin (a last-resort antibiotic against gram-negative bacteria) via the mcr-1 gene has been discovered in E. coli and K. pneumoniae. However, expression of this gene imposes a major fitness cost on bacteria. Multiple mobile colistin resistance (MCR) gene variants and homologs have since been identified globally. My project aims to assess the fitness costs of these homologs and variants in order to determine if these costs are being alleviated. To achieve this goal, we have been competing different MCR-1 variant genes in bacterium of identical genetic backgrounds in order to probe the individual effects of the gene on bacterial fitness. Preliminary results show that certain variants achieve higher levels of colistin resistance than the consensus mcr-1 gene with the caveat of increased fitness costs. Other variants show lower colistin resistances but have increased fitness in the absence of the antibiotic. Results also demonstrate the non-linearity between MCR expression and colistin resistance.

Resistance evolution in real-time

SMBE-PO-583

Complete assembly of Escherichia coli ST131 genomes using long reads demonstrates antibiotic resistance gene variation within diverse plasmid and chromosomal contexts

A. Decano [1], C. Ludden [2 3], T. Feltwell [2], K. Judge [2], J. Parkhill [2], T. Downing [1,*]

[1]School of Biotechnology, Dublin City University, Dublin, Ireland, [2]Wellcome Trust Sanger Institute, Hinxton, [3]London School of Hygiene & Tropical Medicine, London, United Kingdom

Abstract: Background: Bacterial antimicrobial resistance (AMR) genes such as are often rearranged, amplified and translocated by mobile genetic elements (MGEs) and show extensive rearrangements within and between genomes. Short DNA reads do not fully resolve the architecture of repetitive elements on bacterial plasmids to allow MGE structures encoding AMR genes to be fully determined. *Escherichia coli* ST131 has variable plasmid composition and an array of genes enabling AMR including *bla*$_{\text{CTX-M-14/15/27}}$. ST131 is a major clonal group among extraintestinal *E. coli* (ExPEC), which are a major public health concern due to their increasing infection rates worldwide. ExPEC strains resistant to antimicrobials are associated with excess mortality, prolonged hospital stays and higher healthcare costs. Consequently, resolving the genomic architecture and evolution of AMR genes has direct application to better ExPEC treatments.
Methods: We deciphered the genome structures of six *E. coli* ST131 isolated from six patients using long read sequencing with a Nanopore GridION platform. Most long read assemblies generated entire chromosomes and plasmids as single contigs, contrasting with more fragmented assemblies created with short reads alone.
Results: The bacterial long read assemblies highlighted diverse accessory genomes with *bla*$_{\text{CTX-M-15}}$, *bla*$_{\text{CTX-M-14}}$ and *bla*$_{\text{CTX-M-27}}$ genes identified in three, one and one isolates, respectively. One sample had no *bla*$_{\text{CTX-M}}$ gene. Two samples had chromosomal *bla*$_{\text{CTX-M-14}}$ and *bla*$_{\text{CTX-M-15}}$ genes, and the latter was at three distinct locations, likely transposed by the adjacent MGEs: IS*Ecp1*, IS*903B* and Tn*2*. These patterns corresponded to results in a collection of 794 ST131 short read libraries that had similar evolutionary rearrangements at *bla*$_{\text{CTX-M}}$ genes.
Conclusion: We found that GridION X5 platforms can potentially reconstruct bacterial genome structures more accurately than using short reads alone. We showed that AMR genes exist in multiple different chromosomal and plasmid contexts in a clonal group of closely-related *E. coli* ST131 isolates. We applied this to decode plasmid and AMR gene transmission events in larger collections of *E. coli* ST131.
See pre-print at https://biorxiv.org/content/10.1101/558635v1

Resistance evolution in real-time

SMBE-PO-580

Inference of demography and selection in organisms characterized by skewed offspring distributions

A. Sackman [1,*], R. Harris [1], J. Jensen [1]

[1]Arizona State University, Tempe, United States

Abstract: The recent increase in the availability of population genomic data from experimental, natural, and ancient populations has been accompanied by a promising growth in methodologies for inferring demographic and selective parameters from such data. However, these methods have largely presumed that the populations of interest are well-described by the Kingman coalescent. In reality, many groups of organisms, including viruses and many clonal pathogens, typified by high variance in progeny number, may be best characterized by multiple-merger coalescent models. Estimation of population genetic parameters under Wright-Fisher assumptions for these organisms may thus be prone to serious mis-inference. We propose novel methods for the inference of population demography and selection under the Ψ-coalescent model. We first demonstrate mis-inference under the Kingman of common methods for estimating demographic history and the strength of selection, and then exhibit the superior performance of our new methods which properly account for skewed progeny distributions. We apply our approaches to identify mutations responsible for the evolution of drug resistance in influenza A virus and human cytomegalovirus.

Resistance evolution in real-time

SMBE-PO-581

The population genetics of within-host drug resistance in Mycobacterium tuberculosis

A. Y. Morales-Arce [1,*], R. Harris [1], A. Stone [1], J. Jensen [2]

[1]Center for Evolution and Medicine, Arizona State University, Tempe, AZ, [2]Center for Evolution and Medicine, Arizona State University, Tempe, United States

Abstract: The within-host evolutionary dynamics in*Mycobacterium tuberculosis*(TB) populations, and consequent development of resistance to antibiotics, remain unclear. This partly owes to the fact that the underlying biological characteristics of TB (*e.g*., clonality, compact genomes, and lack of recombination) render standard population genetic approaches based on Wright-Fisher (WF) assumptions largely inappropriate. Strong purifying selection, widespread background selection effects, and highly skewed progeny distributions all act to strongly reduce within-patient diversity and skew observed genomic variation. Regardless, it is empirically observed that resistance rapidly and independently evolves across the globe. This rapid adaptation, in spite of lacking variation, represents something of a paradox to date. However, our analyses suggest that previous estimates of mutation rates may in fact be strongly under-estimated owing to the inappropriate use of WF assumptions. Utilizing an approximate Bayesian (ABC) approach based upon within-host site frequency information, we demonstrate an ability to disentangle the relative contributions of these evolutionary processes by accounting for these WF violations within the framework of so-called multiple merger coalescent models. This work thus provides novel population genetic insight into this important human pathogen, uniquely provides unbiased estimates of key evolutionary parameters via the development of an organism-appropriate null model, and serves to reconcile this long-standing paradox.

Resistance evolution in real-time

SMBE-PO-573

The evolution and natural occurrence of agrochemical resistance

L. Parts [1,*], A. Flemming [2], A. Woollard [1]

[1]Department of Biochemistry, University of Oxford, Oxford, [2]Syngenta Ltd., Jeallott's Hill, United Kingdom

Abstract: Resistance to pesticides is a global food security problem and a pressing issue for the agrochemical sector, analogous to the global health concerns caused by widespread antibacterial resistance. A number of compound families, both broad spectrum pesticides and more targeted nematicides, are available for agricultural use to combat the damage caused by pests that results in about 15% of global crop loss annually. Most of these compounds were introduced decades ago and reports of resistance exist for each class, yet little is known about the molecular mechanism and evolutionary biology of resistance.

In order to understand the relationship between mode of action of compounds and evolution of pesticide resistance, we employed an experimental evolution setup where a *Caenorhabditis elegans* strain was exposed to pesticide environments for 15 generations. This timeframe enabled us to observe the emergence of resistance to a commonly used worm control agent. We have also assessed natural variation in pesticide resistance by investigating the development of 25 highly divergent *C. elegans* wild isolates upon exposure to 29 bioactive agrochemicals. We have identified wild isolates sensitive and resistant to different pesticides, and agrochemicals to which the wild isolates display high variation in susceptibility. This information, alongside the findings of our experimental evolution work, enable us to begin to understand the complex relationship between evolution of pesticide resistance and compound chemistry as well as naturally occurring resistance and emergence of resistance in the wild. Understanding these elements of the biology of pesticide resistance will aid in selecting new agrochemical compounds overall less likely to see the emergence of resistant species in the wild.

SMBE Editors symposium
SMBE-PO-589
Selection and adaptation of the splicing machinery in the budding yeast, Saccharomyces cerevisiae
X. Xia*

Abstract: Introns in different genes, or even different introns within the same gene, often have different splice sites and differ in splicing efficiency (SE) which depends on key splice sites (SS) such as 5'SS, 3'SS and branchpoint site (BPS). Introns in highly transcribed genes (HTGs) are expected to be more efficiently spliced than lowly transcribed genes (LTGs). Consequently, introns in HTGs are expected to have stronger signal strength in their splice sites than LTGs. I quantified SE for each of the 304 introns in yeast (*Saccharomyces cerevisiae*) genes, including 24 in the 5'UTR, by an RNA-Seq approach. I measured 1) number of transcriptomic reads mapped to exon-exon junctions (N_{EE}) as a proxy for the abundance of spliced form, and 2) number of reads mapped to exon-intron junction (N_{EI5} and N_{EI3} at 5' and 3' ends of intron) as a proxy for the abundance of unspliced form. The total mRNA is $N_{Total} = N_{EE} + p*N_{EI5} + (1-p)*N_{EI3}$, with the simplest p = 0.5 but statistical methods were presented to estimate p from data. An estimated p is needed because N_{EI5} is expected to be smaller than N_{EI3} due to 1) step 1 splicing occurs before step 2 so EI5 is broken before EI3, 2) enrichment of poly(A) mRNA by oligo-dT, and 3) 5' degradation. SE is defined as the proportion (N_{EE}/N_{Total}). I quantified the signal strength of 5'SS, 3'SS and BPS by position weight matrix scores. The experimentally quantified SE from RNS-Seq data can be well modelled as a function of gene expression and signal strength at 5'SS, 3'SS, and BPS. Introns in HTGs have higher SE, as well as stronger signal strength at key splice sites, than LTGs. A highly significant compensation effect is detected among splice sites, e.g., a strong BPS compensates for a weak 3'SS, partly because the decoding of 3'SS by yeast spliceosome depends strongly on the recognition of BPS. Optimal distance between BPS and 3' end of intron is within the range of 29-41 in yeast. Introns within the same gene have nearly identical intron length which may lead to similar splicing rates among introns of the same gene, consistent with previous findings that intron length is a key determinant of splicing efficiency.

SMBE Editors symposium

SMBE-PO-587

Partial protection from cyclical selection generates a high level of polymorphism at multiple non-neutral sites

Y. Park [1], Y. Kim [*]

[1]EcoScience, Ewha Womans University, Seoul, Korea, Republic Of

Abstract: Temporally varying selection was known to maintain genetic polymorphism under restrictive conditions. However, if a part of population can escape from selective pressure, condition for polymorphism becomes greatly broadened, producing a type of negative frequency dependent selection termed “storage effect”. We investigate whether seasonally fluctuating selection can maintain polymorphism at multiple loci, if cyclically fluctuating selection is not acting in a subpopulation called “refuge”. A phenotype with oscillating seasonal optimum is determined by alleles at multiple sites, across which the phenotypic effects of mutations are distributed randomly. This model resulted in long-term polymorphism at multiple sites that greatly increase the level of non-neutral polymorphism. The level of polymorphism at linked neutral sites however is either similar to or lower than expected for unlinked neutral loci. Overall, these results suggest that, for a protein-coding sequence, nonsynonymous-to-synonymous ratio of polymorphism can exceed one. In addition, under randomly disturbed fitness oscillation, different sets of sites may take turns in harboring long-term polymorphism, thus making it harder to observe trans-species polymorphism. Therefore, cyclically fluctuating selection and storage effect can greatly increase non-neutral genetic variation in a population without leaving classical signatures of balancing selection.

SMBE Editors symposium
SMBE-PO-591
Caught in the act? Speciation, divergence and admixture in wild tomatoes.
L. E. Rose*, T. Kloesges, I. Beddows

Abstract: Hybridization between closely related plant species is widespread, but the long-term outcome of hybridization is not always clear. We have investigated the phylogenetic relationships and the history of hybridization in the wild tomato clade (Solanum sect. Lycopersicon). We sequenced RNA from individuals of 38 different populations and, by combining this with published data, built a comprehensive genomic data set for the entire clade. The data indicate that many taxa are not monophyletic and many individuals are admixed due to repeated hybridization. The most polymorphic species, *Solanum peruvianum*, has two genetic and geographical subpopulations, while its sister species, *S. chilense*, shows reduced heterozygosity and much less substructure. Furthermore, we discovered a new set of populations (currently recognized as *S. chilense*) which are genetically intermediate between *S. chilense* and *S. peruvianum*. Based upon molecular, morphological, and crossing data, we tested the hypothesis that these disjunct "*S. chilense*" populations are an example of recombinational speciation. Recombinational speciation is rarely reported and presents many challenges, both in its unequivocal recognition and difficulty in distinguishing it from other modes of speciation and population history. The discovery of these new cryptic hybrid populations opens new avenues to investigate the genomic outcome of hybridization in plants.

SMBE Editors symposium
SMBE-PO-594
Divergence pattern at the base of teleost fish
N. Takezaki*

Abstract: The divergence of three groups, Osteoglossomorpha (bonytongues and others), Elopomorpha (eels and relatives), Clupeocephala (the remaining teleost fish), at the base of teleost fish is controversial. Previous studies generated all three possible relationships: (1) ((Clupeocephala, Elopomorpha), Osteoglossomorpha), (2) ((Clupeocephala, Osteoglossomorpha), Elopomorpha), (3) (Osteoglossomorpha, Elopomorpha), Clupeocephala). This study analyzed two genome-scale datasets from two recent studies. A dataset of 412 genes from 12 species revealed that Elopomopha is likely the first group that splits from the other two groups by use of closely related outgroup and the inclusion of multiple deep-diverging lineages in the groups, whereas another dataset (1105 genes each of which consists of various number of species (average 240.6 ± 44.5) from total of 304 species) did not provide sufficient information regarding the branching pattern of the three groups after excluding identical sequences within individual genes in contrast to the original study which supported the relationship (2).

SMBE Editors symposium

SMBE-PO-596

Computational Approaches for Characterizing Microbiome and Transposable Element Diversity and Evolutionary Dynamics

K. A. Crandall [1,*]

[1]Biostatistics & Bioinformatics, George Washington University, Washington, DC, United States

Abstract: Microbiome characterization has become an integral component to the study of a wide variety of disease and treatment approaches. Through the collection of metagenomic sequence data from DNA and/or RNA samples isolated from host individuals, effective microbiome characterization can identify pathogens, link diversity to disease state, characterize treatment effects, and identify drug resistant variants. I present a computational platform, PathoScope, for metagenomic sequence analysis to characterize microbiome diversity and test hypotheses about diversity associates with disease and diversity dynamics over time. I then describe a second software package, TeleScope, that characterizes transposable elements in genomic data, maps those elements back to reference genomes, and identifies active mobile elements. I present results from both empirical studies and simulation studies characterizing the utility of our computational approaches with metagenomic data and compare our approach to other leading packages. I then demonstrate our computational tools with applications in endangered species conservation, agriculture, and a variety of aspects of human health. Specifically, I will demonstrate the use of microbiome characterization related to black rhino health, human health related to Konzo disease, and human fecal transplant diversity over time. Finally, I will demonstrate the characterization of human endogenous retroviral elements (HERVs) in relation to mental disorders and characterize the evolutionary history of these elements from specific HERV families.

SMBE Editors symposium
SMBE-PO-598A
Using machine learning to predict long non-protein coding RNAs from plant transcriptomes
B. Golding[*], C. Simopoulos[1], E. Weretilnyk[1]
[1]Biology, McMaster University, Hamilton, Canada

Abstract: Background: In plants, long non-protein coding RNAs are believed to have essential roles in development and stress responses. However, relative to advances on discerning biological roles for long non-protein coding RNAs in animal systems, this RNA class in plants is largely understudied. With comparatively few validated plant long non-coding RNAs, research on this potentially critical class of RNA is hindered by a lack of appropriate prediction tools and databases. Supervised learning models trained on data sets of mostly non-validated, non-coding transcripts have been previously used to identify this enigmatic RNA class with applications largely focused on animal systems. Our approach uses a training set comprised only of empirically validated long non-protein coding RNAs from plant, animal, and viral sources to predict and rank candidate long non-protein coding gene products for future functional validation.
Results: Individual stochastic gradient boosting and random forest classifiers trained on only empirically validated long non-protein coding RNAs were constructed. In order to use the strengths of multiple classifiers, we combined multiple models into a single stacking meta-learner. This ensemble approach bene fits from the diversity of several learners to effectively identify putative plant long non-coding RNAs from transcript sequence features. When the predicted genes identified by the ensemble classifier were compared to those listed in GreeNC, an established plant long non-coding RNA database, overlap for predicted genes from *Arabidopsis thaliana*, *Oryza sativa* and *Eutrema salsugineum* ranged from 51 to 83% with the highest agreement in *Eutrema salsugineum*. Most of the highest ranking predictions from *Arabidopsis thaliana* were annotated as potential natural antisense genes, pseudogenes, transposable elements, or simply computationally predicted hypothetical protein. Due to the nature of this tool, the model can be updated as new long non-protein coding transcripts are identified and functionally verified.
Conclusions: This ensemble classifier is an accurate tool that can be used to rank long non-protein coding RNA predictions for use in conjunction with gene expression studies. Selection of plant transcripts with a high potential for regulatory roles as long non-protein coding RNAs will advance research in the elucidation of long non-protein coding RNA function.

SMBE Editors symposium

SMBE-PO-593

Transgenic rhesus monkeys carrying the human MCPH1 gene copies show human-like neoteny of brain development

L. Shi [1], J. Jiang [2], Y. Chen [3], C. Liu [4], T. Hu [1], M. Li [1], Q. Lin [2], Y. Li [2], H. Wang [3], Y. Niu [3], Y. Shi [5], M. Styner [5], J. Wang [2], Y. Lu [6], X. Sun [6], H. Yu [6], W. Ji [3], B. Su [2 7 8,*]

[1]Kunming Institute of Zoology, Chinese Academy of Sciences, [2]Kunming Institute of Zoology, Chinese Academy of Sciences, [3]Kunming University of Science and Technology, Kunming, China, [4]National Institute of Neurological Disorders and Stroke, National Institutes of Health, Bethesda, [5]University of North Carolina, Chapel Hill, United States, [6]the First Affiliated Hospital of Kunming Medical University, [7]Primate Research Center, Kunming Institute of Zoology, Chinese Academy of Sciences, [8]Center for Excellence in Animal Evolution and Genetics, Chinese Academy of Sciences, Kunming, China

Abstract: Brain size and cognitive skills are the most dramatically changed traits in humans during evolution, and yet the genetic mechanisms underlying these human-specific changes remain elusive. Here, we successfully generated 11 transgenic rhesus monkeys (8 first-generation and 3 second-generation) carrying human copies of *MCPH1*, an important gene for brain development and brain evolution. Brain image and tissue section analyses indicated an altered pattern of neural cell differentiation, resulting in a delayed neuronal maturation and neural fiber myelination of the transgenic monkeys, similar to the known evolutionary change of developmental delay (neoteny) in humans. Further brain transcriptome and tissue section analyses of major developmental stages showed a marked human-like expression delay of neuron-differentiation and synaptic signaling genes, providing a molecular explanation to the observed brain developmental delay of the transgenic monkeys. More importantly, the transgenic monkeys exhibited better short-term memory and shorter reaction time compared to the wild type controls in the delayed matching to sample task.The presented data represents the first attempt to experimentally interrogate the genetic basis of human brain origin using a transgenic monkey model, and it values the use of nonhuman primates in understanding human unique traits.

SMBE Editors symposium
SMBE-PO-598

Six amino acid substitutions in the Smad binding domain underlie neofunctionalization of mCORL1 from mCORL2/dCORL

M. Stinchfield [1], K. Miyazawa [2], S. Newfeld [1,*]
[1]School of Life Sciences, Arizona State University, Tempe, AZ, United States, [2]Dept. Biochemistry, University of Yamanishi, Chuo, Yamanishi, Japan

Abstract: Uncovering how new members of multigene families acquire new functions is an important topic in evolutionary and developmental genetics. CORL proteins (SKOR in mice, Fussel in humans and fussel in Flybase) are a family of CNS specific proteins related to Sno/Ski oncogenes. dCORL participates in TGF-beta and insulin signaling during development and adults but roles for the two mCORL proteins are largely unknown. A series of transgenic studies in parallel of mCORL1, mCORL2 and dCORL in adult wings and the mushroom body of the larval brain as well as biochemical analyses of mCORL2 were conducted to address both issues. We show that mCORL1 has a distinct function from mCORL2 and dCORL and that the latter two presumably share the ancestral function. The data supports the hypothesis originally proposed to explain human Smad2 and Smad3 transgene data, that a small number of amino acid differences in a highly conserved functional domain can confer distinct activities for recently duplicated proteins. The data also suggests testable new hypotheses for mCORL2 function in mammals. Overall, the study reiterates the value of transgenic methods in Drosophila to provide new information about mechanisms of evolution and the function of family members in other species.

The application of evolutionary approaches to the study of microbial communities and microbiomes
SMBE-PO-602

Scales of differentiation – phylogeny, abundance and 'core' microbiomes

G. Singh [1], A. Brass [2], S. Cruickshank [1], C. Knight [3,*]
[1]Lydia Becker Institute of Immunology and Inflammation, [2]Division of Informatics, Imaging and Data Sciences, [3]Faculty of Science and Engineering, The University of Manchester, Manchester, United Kingdom

Abstract: Major efforts in amplicon sequencing have described differences in microbial community members across diverse environments. However, understanding the sorts of evolutionary changes in composition and community differences we expect to find associated with particular environmental changes remains less clear. Firstly, biological variation needs to be disentangled from technological issues: in a re-analysis of independent studies of soil microbiomes, many apparent community differences, particularly those among abundant high-level taxa (domains and phyla) were confounded with the precise amplicon sequenced and sequencing technology used. Nonetheless, particular, rarer, intermediate-level taxa (class to genus) do distinguish community structures [1]. To examine explicitly the phylogenetic scales of microbiome changes associated with different forms of environmental change, we conducted a focused 16S rRNA sequencing study of mouse gut bacterial communities. The experiment was designed, using co-housed siblings with different genotypes, to be able to distinguish effects of niche within host (gut mucus or stool microbiome), host genotype (either wildtype or colitis-prone *mdr1a*$^{-/-}$), age and the particular cage in which the host mice were housed. We bypass many of the issues of bacterial taxonomy by direct phylogenetic reconstruction of the 16S amplicons sequenced across the full study. We used the relative abundance of each clade in this tree as a predictor in a machine learning models to distinguish microbiomes. Strikingly we find *no*consistent differences in microbial communities associated with host genotype. In contrast, individual cages comprise readily distinguishable microbial communities, differing in the proportions of intermediate abundance, low-level clades. As expected, microbiomes from young and older hosts and stool and mucus niches are readily distinguished. Both age-associated evolutionary changes and niche differences are primarily among clades at intermediate phylogenetic scales. However, the most abundant clades only differ among niches. We discuss these findings in the terms of whether it is possible to define evolutionarily stable 'core' microbial communities associated with any particular environment.

[1] Ramirez, K.S., Knight, C.G., de Hollander, M., Brearley, F.Q., Constantinides, B., Cotton, A., Creer, S., Crowther, T.W., Davison, J., Delgado-Baquerizo, M., Dorrepaal, E., Elliott, D.R., Fox, G., Griffiths, R.I., Hale, C., Hartman, K., Houlden, A., Jones, D.L., Krab, E.J., Maestre, F.T., McGuire, K.L., Monteux, S., Orr, C.H., van der Putten, W.H., Roberts, I.S., Robinson, D.A., Rocca, J.D., Rowntree, J., Schlaeppi, K., Shepherd, M., Singh, B.K., Straathof, A.L., Bhatnagar, J.M., Thion, C., van der Heijden, M.G.A. and de Vries, F.T. (2018) **Detecting macroecological patterns in bacterial communities across independent studies of global soils**. *Nature Microbiology*, 3, 189-196. doi:10.1038/s41564-017-0062-x

The application of evolutionary approaches to the study of microbial communities and microbiomes
SMBE-PO-622

phyloRECOMB: Testing recombination events by phylogenetic congruence in whole-genome sequences.

B. Beamud [1,*], F. González-Candelas [1]
[1]Universitat de València-I2SysBio, Valencia, Spain

Abstract: INTRODUCTION
Recombination is one of the main processes shaping the evolution of microorganisms by generating new genetic variants that will be unlikely to appear only by mutation. These variants can, for instance, confer resistance to hosts' immune systems and drug therapies or lead to epidemic outbreaks. Many analysis tools have been developed to analyze recombination. However, none of these has emerged as the best under all possible analytical conditions. Given that the hallmark of recombination is the co-existence of genome fragments with different ancestries, the ultimate test for recombination should be based on revealing a statistically significant lack of phylogenetic congruence among genome portions. In this work, we have developed a pipeline, phyloRECOMB, to rigorously contrast and characterize recombination events by phylogenetic congruence. This pipeline is compatible with high-throughput analysis of whole-genome sequences.
METHODOLOGY
phyloRECOMB is an open pipeline based on phylogenetic placement complemented with topological congruence using the ELW test. It assigns the most likely clade to a given genome fragment after considering its phylogenetic signal. It includes an optional prior step of recombination detection by jpHMM. The reliability of this methodology has been assessed for the assignment of HIV-1 subtypes considering different background alignments, recombination event sizes, donors, and genome locations. The assessment was performed using a sliding-window approach with datasets of "pure" subtypes. Next, we applied phyloRECOMB to previously reported HIV-1 recombinant forms to assess the consistency of their defining recombination events.
RESULTS
phyloRECOMB showed an accuracy over 90% for HIV-1 subtype assignment. Most incongruences were due not to stochastic errors of the methodology, but they indeed reflected additional evolutionary processes or lack of resolution in the case of small recombination fragments. Regarding the consistency of previously reported HIV-1 recombinant forms, we found that only one third (116 out of 379) of the reported recombination events are compatible with detection by both jpHMM and phyloRECOMB. In addition, the recombination events detected by jpHMM seem to be more robust than the ones accepted at LANL HIV database, with phyloRECOMB verification percentages of 86% and 73%, respectively.
DISCUSSION
We have developed a pipeline, called phyloRECOMB, that allows to test recombination in a rigorous manner. Although it has been developed within the scope of HIV-1 recombination, it can be easily adapted to other microorganisms or to the detection of other processes such as Horizontal Gene Transfer (HGT) because their conceptual basis are the same, i.e., lack of phylogenetic congruence among regions with enough phylogenetic signal. We have found many inconsistencies in previously described HIV-1 recombination events, especially those that involve small genome fragments. This observation lead us to propose that additional guidelines, using rigorous statistical and phylogenetic tests of congruence, should be considered to detect and properly characterize recombination events.

The application of evolutionary approaches to the study of microbial communities and microbiomes
SMBE-PO-609

Taxonomic vote, a new method to assign taxonomy after similarity searches

N. García-González [1,*], B. Wrobel [2], F. González-Candelas [1 3]
[1]I2SysBio (Institute for Integrative Systems Biology), Paterna, Spain, [2]Adam Mickiewicz University, Poznan, Poland, [3]CIBER in Epidemiology and Public Health, Madrid, Spain

Abstract: One of the most popular methods to identify the taxon of a nucleotide sequence relies on obtaining the best hit in a BLAST search. However, this method has many disadvantages and limitations. Normally, only the first hit is selected because in most cases the best hit corresponds to the source organism or its most close relative present in the database. However, in certain situations the first hit can be misleading in that respect. Moreover, if BLAST cannot identify a gene clearly, the first hit might be an artifact . BLAST does not identify orthologous sequences nor close evolutionary relatives but only sequences sharing segments of high similarity which may arise through additional processes to shared inheritance or phylogenetic history.

In light of the limitations of this popular approach, we propose an alternative method, denoted Taxonomic Vote (TV). TV is based on the results obtained by BLAST+, although it can be easily applied to other programs yielding scored results of similarity searches against predefined databases. Our goal is to gain insight about the overall taxonomy of a sample given the taxonomy of each of its genes. For this, a BLAST+ search is made for each query gene/sequence. The first 500 BLAST+ hits for each query are considered.

TV allows obtaining the most probable taxonomy for each query and to evaluate how many queries in a sample are well identified. Also, it gives information about the taxonomy of the sample for each taxonomic level, contributing to create a very approximate idea of the closest relative organism to that in the sample of interest.

There are several differences between the TV procedure and traditional identifications based on BLAST searches. Firstly, in most traditional projects based on BLAST results, only the first hit is retained. In TV, all the hits with a significant score (90% of the best-hit score) are evaluated. Secondly, it is assumed that traditional BLAST accuracy reaches the species level for each best hit. The TV algorithm performs a vote (counting the number of occurrences of each taxon) in each taxonomic level and estimates at what taxonomic level a robust taxonomic identification has been produced. This is indicated with different proposed Taxonomic Vote classes dependent on how well a gene is identified at each taxon level.

Here, we present an application of TV to obtain an overall identification of 10 bacterial isolates obtained from marine sediments collected in the Artic ocean. This objective was not reached at the species level, but all the samples were identified at the genus level. Naturally, this is dependent on the available databases. The results of the TV method are similar to other genomic approaches used, as the estimation of ANI values. The method with the most positive results at the species level was the 16S rRNA sequence comparison. Nevertheless, 16S rRNA comparisons and ANI values did not reflect the complex situation of two samples, whereas the TV results improved their identification.

Although Taxonomic Vote was developed to understand the taxonomic content of the samples, it can be applied to many other situations, such as identification of contamination in high-throughput sequencing studies, characterizing horizontal gene transfer events, or identifying the taxonomic contents of metagenomic samples.

The application of evolutionary approaches to the study of microbial communities and microbiomes
SMBE-PO-605

Paleogenomic assessment of ancient pathogens in Pre-Hispanic and Colonial individuals from Central Mexico

M. Bravo-Lopez [1,*], V. Villa-Islas [1], A. Guzman-Solis [1], E. Mejia-Perez [2], A. Herrera-Muñoz [3], G. Zepeda-Garcia [4], K. Sandoval-Mendoza [5], M. A. Nieves-Colon [5 6], M. Moreno-Cabrera [2], A. Meraz-Moreno [2], J. Gomez-Valdes [7], J. Wesp [8], M. C. Avila-Arcos [1]

[1]International Laboratory for Human Genome Research, National Autonomous University of Mexico, Santiago de Queretaro, [2]National Institute of Anthropology and History, Mexico, [3]National Institute of Anthropology and History, Santiago de Queretaro, [4]National Institute of Anthropology and History, [5]National Laboratory of Genomics for Biodiversity (LANGEBIO-UGA), Guanajuato, Mexico, [6] School of Human Evolution and Social Change, Arizona State University, Tempe, AZ, Arizona, United States, [7]Postgraduate Studies Division, National School of Anthropology and History, UNAM, Mexico, Mexico, [8]Department of Sociology and Anthropology, North Carolina State University, North Carolina, United States

Abstract: Ancient DNA–based approaches have enabled health reconstruction in past populations from genomes of pathogenic microorganisms. To gain insights into human infectious diseases in individuals from Mexico, we analyzed Pre-Hispanic and Colonial teeth samples from Central Mexico. As a first step, we generated low-depth shotgun sequencing data from both sample sets and used KRAKEN to search for traces of ancient pathogen DNA, we compared the sequences with a database including complete bacterial, archaeal, and viral genomes from RefSeq. The taxonomic assignments specific to known pathogenic species were evaluated in more detail. Interestingly, we were able to detect two human pathogens present in important amounts, *Tannerella forsythia* and *Salmonella enterica* serovar *Paratyphi C. T. forsythia,* which was identified in both pre-Hispanic and Colonial individuals*,* has been considered as the prime candidate for the progression of periodontitis and to be involved in the etiology of chronic systemic diseases, including cardiovascular disease. *Salmonella enterica serovar Paratyphi C*, a gram-negative bacteria that causes enteric fever, was found only in an individual in the colonial dataset. The two pathogens found in Pre-Hispanic and Colonial individuals revealed the characteristic damage patterns expected for ancient DNA, thus supporting their authenticity as ancient pathogens. The reconstruction of *T. forsythia* and *S. enterica* ancient genomes through a capture-enrichment approach will provide a unique opportunity to better understand past lifestyle, health, diet and the genetic makeup of pathogens at key transitional periods in Central Mexico.

The application of evolutionary approaches to the study of microbial communities and microbiomes
SMBE-PO-608

An eradicated European Plasmodium vivax strain retrieved from antique medical slides sheds light on its dispersal

L. Van Dorp [1,*], P. Gelabert [2], A. Rieux [3], M. de Manuel [2], T. de-Dios [2], S. Gopalakrishnan [4], C. Caroe [4], M. Sandoval-Velasco [4], R. Fregel [5], I. Olalde [6], R. Escosa [7], C. Aranda [8], S. Huijben [9], I. Mueller [10], T. Marques-Bonet [2], F. Balloux [1], T. Gilbert [4], C. Lalueza-Fox [2]

[1]UCL Genetics Institute, London, United Kingdom, [2]Institute of Evolutionary Biology (CSIC-UPF), Barcelona, Spain, [3]CIRAD, La Reunion, France, [4]Section for EvoGenomics, Copenhagen, Denmark, [5]Department of Genetics, Standford University, Standford, [6]Department of Genetics, Harvard Medical School, Boston, United States, [7]8Consorci de Polítiques Ambientals de les Terres de l'Ebre (COPATE), Deltebre, [8]Servei de Control de Mosquits, Sant Feliu de Llobregat, Spain, [9]Center for Evolution and Medicine, School of Life Sciences, Arizona State, Tempe, United States, [10]ISGlobal Barcelona Institute for Global Health, Barcelona, Spain

Abstract: Although several studies have tried to reconstruct the phylogeography of *Plasmodium vivax* alongside human population movements, the lack of genomic information from extinct European strains has prevented a clear understanding of its spread. We used medical microscope slides prepared in 1944 from malaria-affected patients in Spain's Ebro Delta to generate the first complete genome from an eradicated European sample. Population genetics and phylogenetic analyses placed this strain basal to a cluster including samples from the Americas. This genome allowed us to calibrate a genomic mutation rate for *P. vivax*, and estimate the age of the last common ancestor between European and American strains to the colonial period of the 17^{th}-19^{th} century. In addition, we found that some known variants for resistance to anti-malaria drugs were already present in the European strain, predating their use, which has important implications for modelling the emergence and spread of future resistance mutations.

The application of evolutionary approaches to the study of microbial communities and microbiomes
SMBE-PO-624

Differential patterns of evolutionary dynamics of genomic compartments in co-occurring Prochlorococcus populations

H. Gardon [1,*], C. Petit [1], I. Jouan-Dufournel [1], G. Bronner [1]
[1]Université Clermont Auvergne, UMR CNRS 6023, Aubière, France

Abstract: Bacterial co-occuring populations differentiate over time and space to form distinct genetic units. The mechanisms governing this diversification are presumed to result from the ecological context of living units. A model assuming the exchange of advantageous genes among populations rather than whole genome sweeps has emerged to explain differentiation of populations. However the characteristics of these genes and whether their evolution is driven by adaptative or neutral processes remain controversial. By analysing single-amplified genomes of the cyanobacterium *Prochlorococcus* HLII ecotype co-occuring populations from Kashtan et al. (2014), we highlight that genomic compartments rather than populations are characterized by differences in their evolutionary dynamics. Although genome phylogeny, average nucleotide identity and the content in flexible genes sustain the delineation of these populations, no differences were found with respect to their evolutionnary rate. On the other hand, genomic islands show contrasted profiles of gene evolution. Whatever the populations, flexible genes are enriched with functional categories ensuring the interface of the bacteria with its environment. Associated genes occur in some but not all genomic islands, may have no cyanobacterial origin and are subject to negative selection. Our results suggest that high genomic variability among co-occurring populations does not reflect differences in their selective constraints, but patchy processes of evolution along genomes, shared by all populations. Further investigation should include quantification of homologous recombination and horizontal transfer with regard to the nature and fate of acquired genes.
Kashtan et al., 2014. “Single-Cell Genomics Reveals Hundreds of Coexisting Subpopulations in Wild Prochlorococcus.” Science 344 (6182):416–20.

The application of evolutionary approaches to the study of microbial communities and microbiomes
SMBE-PO-623

Carbon fixation by marine ultra-small prokaryotes

R. Lannes [1,*], K. Olsson-Francis [2], P. Lopez [1], E. Bapteste [3]
[1]Université Sorbonne, Paris VI, Paris, France, [2]The Open University, School of Environment, Earth and Ecosystems, Milton Keynes, [3]Université Sorbonne, Paris VI, Paris, United Kingdom

Abstract: Autotrophic carbon fixation is a crucial process for sustaining life on Earth. To date, six pathways, the Calvin-Benson-Bassham cycle, the reductive tricarboxylic acid cycle, the 3-hydroxypropionate bi-cycle, the Wood-Ljungdahl pathway, the dicarboxylate/4-hydroxybutyrate cycle, and the 4-hydroxybutyrate cycle have been described. Nanoorganisms, such as members of the Candidate Phyla Radiation (CPR) bacterial superphylum and the Diapherotrites, Parvarchaeota, Aenigmarchaeota, Nanoarchaeota, Nanohalorchaeota (DPANN) archaeal superphylum, could deeply impact carbon cycling and carbon fixation in ways that are still to be determined. CPR and DPANN are ubiquitous in the environment but understudied; their gene contents are not exhaustively described, and their metabolisms are not yet fully understood. Here, the completeness of each of the above pathways were quantified and tested for the presence of all key enzymes in a diversity of nanoorganisms across the World Ocean. The novel marine ultra-small prokaryotes was demonstrated to collectively harbor the genes required for carbon fixation, in particular the 'energetically efficient' DH pathway, and HBC pathways. This contrasted with the known carbon metabolic pathways associated with CPR memebers in aquifers, where they are described as degraders (Castelle 2015 et al., 2015, Castelle et al., 2018, Anantharaman et al., 2016). Our findings offer the possibility that nanoorganisms have a broader contribution to carbon fixation and cycling than currently assumed. Furthermore, CPR and DPANN are possibly not the only nanosized prokaryotes; therefore, the discovery of new autotrophic marine nanoorganisms, by future single cell genomics is anticipated.

The application of evolutionary approaches to the study of microbial communities and microbiomes
SMBE-PO-604

The evolution of Arabidopsis thaliana-associated Pseudomonas

H. Ashkenazy [1,*], T. Karasov [1], O. Shalev Skriptchak [1], M. Neumann [1], D. Weigel [1]
[1]Department of Molecular Biology, Max Planck Institute for Developmental Biology, Tuebingen, Germany

Abstract: Members of the *Pseudomonas* genus of gram-negative bacteria are often highly abundant in metagenomic samples collected from plants. To better understand the process of *Pseudomonas* colonization, we study the evolution of a local collection of plant-associated *Pseudomonas* strains via comparative genomics.
Specifically, we analyzed 1,524 genomes of *Arabidopsis thaliana*-associated *Pseudomonas* isolates, collected across several years and sites near Tübingen, Germany. We inferred the pan-genome for all isolates based on their orthologous groups. Out of the 72,397 orthologous groups specified, only 1.3% belonged to the bacterial 'core genome', while 36.3% were unique to individual isolates. Analyzing the dynamics of gain-and-loss events within and between orthologous groups with an evolutionary framework allowed us to infer a co-evolutionary network composed of orthology groups gained-and-loss together. Overall, 9.81% of the orthologous groups co-evolve with at least one partner. Further dividing this co-evolutionary network into 1,014 highly connected modules, we identified modules that are associated with specific cellular functions. By comparing closely-related isolates we also discovered genomic islands and further characterized their evolutionary dynamics. Finally, by focusing on a subgroup of phylogenetically close isolates differing in their pathogenicity levels, we could associate specific orthologous groups and co-evolving modules with strain pathogenicity.

The application of evolutionary approaches to the study of microbial communities and microbiomes
SMBE-PO-610
Microbiome predicts geography on a global scale
L. Mccarthy [1,*], E. Elhaik [2], C. Mason [3]
[1]University of Sheffield, Sheffield, United Kingdom, [2]Animal and plant sciences , University of Sheffield , Sheffield , United Kingdom, [3]Weill Cornell Medical College, New York, United States

Abstract: Microbiome predicts geography on a global scale

L. McCarthy[1], C.E. Mason[2,3,4], MetaSUB International Consortium, and E. Elhaik[1],

[1] Department of Animal and Plant Sciences, University of Sheffield, Sheffield, UK
[2] Department of Physiology and Biophysics, Weill Cornell Medicine, New York, NY, 10021, USA
[3] The HRH Prince Alwaleed Bin Talal Bin Abdulaziz Alsaud Institute for Computational Biomedicine, New York, NY, 10021, USA
[4]The Feil Family Brain and Mind Research Institute, Weill Cornell Medicine, New York, NY, 10021, USA

Recent advancements in genomic sequencing have allowed for a greater understanding of the makeup of bacterial communities present around us and the development of microbiome profiles for a range of different environments. The relationships between the microbiome and geography are poorly understood, mainly due to the lack of global datasets. Here, we capitalize on two global microbiome datasets: the soil microbiome (Delgado-Baquerizo et al. 2018; Science), which contains the relative abundance of 511 dominant bacterial taxa collected from 237 soil surfaces worldwide and the MetaSUB datasets collected from subway and underground transportation systems in a single day of 2017 from 4116 sites in cities across the world. The samples were fully sequenced and 3757 bacteria species were identified using KrakenHLL. We developed a Random Forest-based approach to predict the geographical coordinates of the sampling sites for either dataset based solely on the relative bacterial abundances. For this, we applied supervised learning algorithms to two sperate datasets. For the soil microbiome dataset, we predicted 54% of the microbiome samples to be within 100km of their country of origin and 68% to within 1500km. For the MetaSUB dataset, we predicted 66% of samples to be within 500km of their sampling city and 76% within 1000km. Our results demonstrate the potential of our algorithm and data for forensics application and mapping microbiome data to the environment. We anticipate that the predictive accuracy of our models would increase with time as more data are accumulate.

The application of evolutionary approaches to the study of microbial communities and microbiomes
SMBE-PO-612

Impact of environmental perturbations on adaptive eco-evo processes in ammonia oxidisers

C. Gubry-Rangin [1,*], A. Aigle [1], C. Thion [2], J. Pett-Ridge [3], G. Nicol [2], M. Firestone [3], J. Prosser [1]
[1]University of Aberdeen, Aberdeen, United Kingdom, [2]Ecole Centrale Lyon, Lyon, France, [3]University of Berkeley, Berkeley, United States

Abstract: Niche specialisation and differentiation are crucial concepts for understanding environmental adaptation of microbial communities. However, they are often difficult to demonstrate in natural environments because of the complexity of microbial communities, lack of information on physiological characteristics of environmental significance and functional redundancy. Alternative approach is to use knowledge gained from culture ecophysiological studies in controlled microcosms.

Information is available on the ecophysiology of archaeal and bacterial ammonia oxidisers (AOA and AOB), which perform the same ecosystem function and control rates of soil nitrification, the oxidation of ammonia to nitrate via nitrite; ammonia oxidisers therefore present a valuable model for studying terrestrial adaptation. This study adopted a hypothesis-based approach to investigate niche specialisation of AOA and AOB and their differential responses to perturbations in two potentially important environmental characteristics, soil pH and temperature.

Hypothesis predictions of responses of AOA and AOB abundance, activities and community structure to perturbations in soil pH and temperature were tested experimentally in soil microcosms in which both groups were present and potentially active by measuring temporal changes in gene and transcript abundance and levels of stable isotope assimilation. These measures of growth and activity were complemented by analysis of the community composition at a very fine phylogenetic scale using a newly developed Illumina MiSeq sequencing approach, employing a bioinformatics pipeline adaptable to any functional gene. Fine-scale diversity of active microbes was also matched to a comprehensive database of pH preferences of soil AOA.

The findings provide evidence for niche specialisation of AOA and AOB following environmental perturbations and demonstrate the importance of considering prokaryote richness and phylogenetic resolution for understanding ecosystem function and assessment of functional redundancy. Environmental perturbation also led to preferential selection and growth of generalist or specialist AOA, depending on the intensity of pH perturbation. The study therefore demonstrates the need for well-defined hypotheses for understanding microbial adaptation in complex ecosystems and provides novel experimental and conceptual approaches for studying richness-function relationships, functional redundancy and niche specialisation concepts.

The application of evolutionary approaches to the study of microbial communities and microbiomes
SMBE-PO-611

Evolutionary contingency drives metabolic interactions and the emergence of ecological structure in microbial communities

J. Meijer*, B. Dijk, van [1], T. Cuypers [1], P. Hogeweg [1]
[1]Theoretical Biology & Bioinformatics, Utrecht University, Utrecht, Netherlands

Abstract: Metabolite exchange is widespread in natural microbial communities and an important driver of ecosystem structure and diversity. It can evolve under laboratory conditions in initially clonal populations, leading to the emergence of simple ecologies of crossfeeding microbes, and is predicted from genome-scale and kinetic metabolic models. However, metabolic generalists individually capable of performing all necessary metabolic functions also evolve in these experiments, and are common in nature. This raises the question what determines whether evolution leads to an 'ecosystem-based' solution of specialized crossfeeding species, or an 'individual-based' solution of self-sufficient generalists.

Here, we investigate the role of random mutations in this process by performing a parallel evolution experiment. We evolved a community of digital organisms under the exact same conditions in a constant, one-niche environment. Specifically, we use *Virtual Microbes*, a multilevel, agent-based model of microbial eco-evolutionary dynamics that features an evolvable genotype-to-phenotype map, toy biochemistry, structured genomes, metabolic and gene regulatory networks, and structured environment. We do not a priori define metabolic and ecological strategies, trade-offs, or an explicit fitness function. Instead, we let these evolve via emergent interactions, and characterize the evolved solutions.

We show that initially identical microbial communities follow different evolutionary trajectories, with half of the population replicates diversifying in crossfeeding lineages, and half evolving self-sufficient generalist lifestyles. Which type of community evolves is determined by prior metabolic adaptations that fixate in the initial community. While the requirements for reproduction are identical for all microbes, the large degree of freedom allows evolution to find different metabolic solutions, and the evolved solution sets limitations on future ecological roles. This suggests that evolution of microbial communities does not reflect global optimization of resource utilization and cannot be predicted from the biochemical constraints or first principles, but instead is contingent on evolved properties of the cell.

The application of evolutionary approaches to the study of microbial communities and microbiomes
SMBE-PO-614

Swords to ploughshares: discovering antibiotics honed by evolution in microbial communities and using them as next generation human therapeutics.

B. J. Thomas [1 2], J. Gani [3], K. Hilpert [3], S. Huws [4], C. J. Creevey [1,*]
[1]Institute of Global Food Security, Queen's University, Belfast, [2]IBERS, Aberystwyth University, Aberystwyth, [3]St. George's University Hospital, London, [4]School of Medical Biology, Queen's University, Belfast, United Kingdom

Abstract: Microbial antibiotic resistance is widely regarded to be one of the most pressing threats facing humanity. Finding new antibiotics is a vital research area and can now be supported by a vast reservoir of readily available 'omic data on the back of the explosion of low cost sequencing technologies.
Antimicrobial peptides (AMPs) are evolutionarily ancient weapons and their ubiquity across all kingdoms of life supports the hypothesis that they have played a key role in the successful evolution of complex multicellular organisms. Despite their ancient lineage, antimicrobial peptides have remained effective, bringing into question the inevitability of the fact that bacteria, fungi and viruses have the potential to develop resistance.
The synthesis of AMPs from scratch in volume is now possible. A well organised screening program can screen 100s of prospects a day against model bacterial organisms to test for activity and is one of the few areas of biological science that can scale to meet the data output from computational prediction toolkits.
AMPLY, an in-house tool designed at Aberystwyth University supported by Life Science Wales and working in collaboration with Tika Diagnostics at St. George's Hospital (London) and Queen's University (Belfast) is part of a next wave of computational drug discovery platforms and is already uncovering a treasure trove of novel, natural AMPs, shaped by evolutionary pressure, in diverse microbial environments.
We highlight the significant benefits of forming a feedback loop between the understanding of microbial community dynamics, predictive modelling and confirmatory lab screening to appropriate nature's naturally honed antimicrobial weapons for medical use.

The application of evolutionary approaches to the study of microbial communities and microbiomes
SMBE-PO-606

Co-evolution of termites and gut microbiomes?

X. Xie [1,*], A. Anderson [1], L. Wran [1]
[1]Department of Biology, Virginia State University, Petersburg, United States

Abstract: Under what conditions the host species and their associated microbiota can co-evolve remains an unanswered question. The termites and their gut microbiomes have been proposed to have evolved together, but other studies have suggested otherwise. To better answer this question, we studied termite populations/species from different geographic areas and their associated microbiomes, using both data generated in our own lab and publically available datasets. Detailed analyses have revealed several interesting findings, which will be presented at the meeting.

The application of evolutionary approaches to the study of microbial communities and microbiomes
SMBE-PO-619

Coincident relationships between evolving objects in prokaryotes

F. J. Whelan [1,*], M. Rusilowicz [2], J. O. McInerney [1]
[1]University of Nottingham, Nottingham, [2]University of Manchester, Manchester, United Kingdom

Abstract: Throughout evolution, evolving objects (domains, genes, operons etc.) have continuously combined, forming new proteins, gene clusters, and functional pathways. Horizontal gene transfer, particularly among prokaryotes, has facilitated this combinatorial process. As such, we expect that positively interacting evolving objects will co-occur with each other more often than by chance; conversely, evolving objects which have antagonistic functions may exclude each other. In this work, we explore these coincident relationships using a novel method reliant on graph theory. We have implemented networks in which each node represents an evolving object which are connected by an edge to another node if there is a coincident (co-occurring or exclusatory) relationship between them. Our method incorporates the phylogenetic distribution and synthenic distances of these objects, and we demonstrate how these concepts can be used to identify conserved clusters of vertical and horizontally inherited units of selection. We apply this methodology to various prokaryotic datasets including pangenomes of common human commensal (e.g. Streptococcus sp.) and pathogenic (e.g. Pseudomonas sp.) organisms, as well as metagenomic sequencing of human-associated microbial communities. We find evidence for evolving objects that significantly co-occur with each other, including genes from characterized biological pathways and genes with unknown functions. Further, we identify objects that exclude each other, indicating antagonistic or redundant biological functional relationships. This work represents a different approach to understanding the evolution of prokaryotes and allows us to draw novel hypotheses as to the potential role of these genetic clusters in host-adapted communities.

The application of evolutionary approaches to the study of microbial communities and microbiomes
SMBE-PO-618

Recovering an authentic oral microbiome signature from historical dental calculus of non-human mammals

H. G. Leitão [1,*], J. C. Brealey [1], K. Guschanski [1]
[1]Department of Ecology and Genetics/Animal Ecology, University of Uppsala, Uppsala, Sweden

Abstract:
Hosts and their associated microbial commensals have a long-standing co-evolutionary relationship, which has been inferred through phylogenetic studies of host-associated microbiomes across millions of years of host radiation. However, we have little understanding of how microbial communities change on a shorter temporal scale of millennia in response to environmental and host demographic factors. Here we explore the suitability of dental calculus to study the co-evolution of host and its microbiome over these timescales. Dental calculus is a calcified microbial biofilm that forms on mammalian teeth, remains unchanged over millennia, and can be readily recovered from historical and archaeological specimens. Although it has been characterized from historical humans, its use in non-human mammals has been limited. To establish dental calculus as a tool for the study of host-microbiome coevolution, we have developed a rigorous laboratory and bioinformatics workflow to produce and analyse shotgun metagenomic data from historical museum specimens of non-human mammals. We devised an analytical pipeline that detects and removes modern microbial contamination and taxonomic artefacts based on fragment length, taxa relative abundances, and presence in negative controls. Taxa identified as contaminants were consistently present in all samples, supporting their likely exogenous nature. Our processing steps improve our inferences and increase the proportion of authentic oral microbial taxa, as identified through SourceTracker analyses. Contamination filtering is thus a critical step in metagenomic analyses of historical microbiomes and in our case allows the successful identification of a true oral signal.

The application of evolutionary approaches to the study of microbial communities and microbiomes
SMBE-PO-617

Rapid loss of CRISPR-mediated herd immunity from bacterial populations

S. Meaden [1,*], L. Capria, E. Opsal [1], L. Lenzi [2], A. Buckling [1], S. Van Houte [1], E. Westra [1]
[1]University of Exeter, Penryn, [2]University of Liverpool, Liverpool, United Kingdom

Abstract: The effectiveness of CRISPR-Cas immunity depends on the diversity of spacers present within a bacterial population, with higher levels of herd immunity conferred from greater spacer diversity. This is due to diverse spacers constraining phages from evolving to overcome CRISPR immunity. Despite the importance of spacer diversity for CRISPR herd immunity levels, the mechanisms that maintain this diversity over time are unexplored. We performed experimental evolution using the opportunistic human pathogen *Pseudomonas aeruginosa* strain PA14, and its phage DMS3vir. Using deep-sequencing and traditional assays we explored population dynamics. We found that upon infection bacterial populations initially generate high population-level spacer diversity, causing rapid phage extinction. However, levels of spacer diversity rapidly decline after phage extinction and sensitive bacteria and receptor mutants invade the population. The consequence of this process is that immunized bacterial populations that were initially effective in driving phage extinct, rapidly lose this ability upon re-infection. This suggests in intrinsic cost of CRISPR immunity which we explore via a number of mechanisms. These data have implications for the maintenance of genetic diversity and coexistence of hosts and parasites in microbial populations.

The application of evolutionary approaches to the study of microbial communities and microbiomes
SMBE-PO-621

On the influence of urbanization on the human gut and oral microbiome – biological and methodological perspectives

A. Lokmer [1,*], A. Sophie [1], S. Lafosse [1], F. T. Ekwin [2], A. Froment [1], L. Ségurel [1]
[1]Laboratory of ecoanthropology and ethnobiology, UMR7206 - CNRS/MNHN/Paris Diderot, Paris, France, [2]Faculty of Medicine and Biomedical Sciences, University Yaounde 1, Yaounde, Cameroon

Abstract:
Industrialization has been associated with a loss of gut microbiota diversity in humans. As a decreased gut microbiome diversity accompanies a number of modern diseases, understanding which factors drive this loss is vital for public health. It is also of great evolutionary interest to understand how gut bacteria adapt to rapidly changing environments. However, industrialized and non-industrialized populations differ in many ways, making it practically impossible to disentangle the effects of diet, sanitary conditions, medical practices, genetic factors or geography. Moreover, gut protozoa, who have likely shaped the human-gut microbiota interactions throughout their coevolutionary history but are virtually absent from industrialized populations, are rarely taken into account. Finally, even less is known about the effects of industrialization on other microbiomes, including the oral microbiome, another important health-associated microbial community of humans. To address some of these limitations, we examined oral and gut microbiomes of 140 individuals from Cameroon along a small-scale urbanization gradient. Apart from metagenetic and metagenomic data, we collected a number of ethnological, medical, sanitary and parasitological parameters in order to identify factors that influence microbiome diversity and variation.
In addition, given the complexity of these microbial communities, we examined how our conclusions changed depending on the taxonomic and phylogenetic resolution scale used, as well as on the way we calculate diversity estimates. Such a comprehensive approach is more likely to identify the processes generating microbiome variation within and among individuals.
Overall, our results shed light on the link between various aspects of urbanization and human microbiome variation in a non-industrialized setting and highlight the importance of exploring various methodologies for inference and hypothesis generation in microbiome research.

The application of evolutionary approaches to the study of microbial communities and microbiomes
SMBE-PO-616

Host interactions may drive Streptococcus mitis genetic diversity.

C. Davison [1,*], Y. L. Tam [1], S. Tallman [1], B. Kwambana [2], E. Sambou [2], E. Foster-Nyarko [2], S. Beleza [1]
[1]Genetics and Genome Biology, University of Leicester, Leicester, United Kingdom, [2]West Africa Partnerships and Strategy, Medical Research Council Unit The Gambia at The London School of Hygiene and Tropical Medicine, Fajara, Gambia

Abstract: The commensal bacterium *Streptococcus mitis* is a major coloniser of the oral cavity. This close cousin of *Streptococcus pneumoniae* has been identified to occupy a range of oral niches independent of pH, epithelial surface and redox state. We have previously observed a high level of genetic diversity in *S. mitis* with an open pan-genome and only 46% of the genome identified as core. We aim to identify the driving force behind the high levels of genetic diversity identified in *S. mitis* and how this varies within hosts and between host populations to better understand the co-evolution of microbes with their human hosts. As new research emerges highlighting the importance of our microbiome in forensics, anthropology and health and disease, understanding how microbes diversify according to their host becomes an increasingly pressing question. We have collected 99 whole genome sequences from the buccal mucosa and tongue dorsum of individuals of European (31), Asian (36) and African (32) descent. We have found that, In contrast to *S. pneumoniae*, genetic diversity in *S. mitis* is predominantly driven by mutation with recombination playing only a minor role. This suggests *S. mitis* is under different selective pressures to other commensal and pathogenic bacteria. The diversity observed within isolates from a single host points towards this pressure being from interactions with the host, perhaps the immune system. This highly mutational nature obliterates evidence of geographical stratification and thus, vertical transmission. Contrasting the genetic variation and evolution of *S. mitis* with close relative *S. pneumoniae* will shed light on the acquisition of the trait(s) that conferred pathogenicity to *S. pneumoniae*, making it the most lethal transmissible agent to humans.

The application of evolutionary approaches to the study of microbial communities and microbiomes
SMBE-PO-625

Modelling microbiome recovery after antibiotics using a stability landscape framework

L. Shaw [1,*], H. Bassam [2], C. P. Barnes [3], A. S. Walker [1], N. Klein [4], F. Balloux [5]
[1]Nuffield Department of Clinical Medicine, University of Oxford, Oxford, [2]CoMPLEX, [3]Department of Cell and Developmental Biology, [4]UCL Institute of Child Health, [5]UCL Genetics Institute, UCL, London, United Kingdom

Abstract: Treatment with antibiotics is one of the most extreme perturbations to the human microbiome. Even standard courses of antibiotics dramatically reduce the microbiome's diversity and can cause transitions to dysbiotic states. Conceptually, this is often described as a 'stability landscape': the microbiome sits in a landscape with multiple stable equilibria, and sufficiently strong perturbations can shift the microbiome from its normal equilibrium to another state. However, this picture is only qualitative and has not been incorporated in previous mathematical models of the effects of antibiotics. Here, we outline a simple quantitative model based on the stability landscape concept and demonstrate its success on real data. Our analytical impulse-response model has minimal assumptions with three parameters. We fit this model in a Bayesian framework to data from a previous study of the year-long effects of short courses of four common antibiotics on the gut and oral microbiomes, allowing us to compare parameters between antibiotics and microbiomes, and further validate our model using data from another study looking at the impact of a combination of last-resort antibiotics on the gut microbiome. Using Bayesian model selection we find support for a long-term transition to an alternative microbiome state after courses of certain antibiotics in both the gut and oral microbiomes. Quantitative stability landscape frameworks are an exciting avenue for future microbiome modelling.
See the associated paper (in press, *The ISME Journal*): https://www.biorxiv.org/content/10.1101/222398v3

The application of evolutionary approaches to the study of microbial communities and microbiomes
SMBE-PO-600

Biodiversity assessment of foraminiferal communities at different ecological scales using High-throughput sequencing

R. Thakur [1 2,*], J.-D. Grattepanche [3], L. A. Katz [1 2]
[1]Department of Biological Sciences, Smith College, Northampton, MA, [2]Program in Organismic and Evolutionary Biology, University of Massachusetts Amherst, Amherst, MA, [3]Department of Biology, Temple University, Philadelphia, Pennsylvania, United States

Abstract: Microbial eukaryotes exist everywhere on Earth and play key roles in shaping ecosystems. Historically, biodiversity assessment for eukaryotic microbes has relied on morphological assessment, which requires specialists and can be time-consuming. Recent application of 'universal' eukaryotic primers for amplicon sequencing have been hampered by the rapid rates of evolution and considerable length variation among eukaryotic ssu-rDNAs. Here we focus on one such group Foraminifera, a clade of mostly marine shell-building amoebae that act as biomarker of past and present climatic conditions. Previous biodiversity assessment of Foraminifera have relied mostly on the morphological data and many inserts in their ssu-rDNA require a customized approach to characterize their diversity. Therefore, we designed Foraminifera specific primers from hypervariable regions of their ssu-rRNA genes to assess the foraminiferan communities through high-throughput sequencing. We apply our methods across several ecological scales, including the intertidal habitats of the North Atlantic Ocean, and open ocean habitats. This approach is powerful for studying the present diversity within foraminiferan communities.

The application of evolutionary approaches to the study of microbial communities and microbiomes
SMBE-PO-599

Testate Amoeba Community Diversity Across Time and Space in New England Bogs

O. Dufour [1], K. Munyenyembe [1,*], L. A. Katz [1 2], A. K. Weiner [1]
[1]Department of Biological Sciences, Smith College, Northampton, [2]University of Massachusetts Amherst, Program in Organismic and Evolutionary Biology, University of Massachusetts, Amherst , Amherst, United States

Abstract: Testate amoeba of the order Arcellinida are a group of unicellular eukaryotic microorganisms that build shells (tests) using materials generated intracellularly and/or isolated from the environment. The shells can be calcareous, siliceous or proteinaceous. Arcellinida are sensitive to environmental changes such as temperature, moisture and pH; as a result they are useful models for paleoenvironmental studies and excellent bioindicators of environmental changes. Therefore, tracking Arcellinida communities can show the broader implications of changing environments on biodiversity. So far such studies have relied on the morphological species concept for the identification of Arcellinida diversity. However, genetic studies have shown that morphology does not resolve the entire diversity of Arcellinida, due to the existence of cryptic diversity. Given this, we are developing methods for amplicon sequencing of Arcellinida communities to identify their diversity based on genetics, in addition to carrying out traditional counting methods by microscopy. Our study explores Arcellinida testate amoebae communities in New England bogs and fens across time and space. We analyze the amplicon sequence data using a Python-based pipeline. The resulting data will allow us to gain more insight into the way these communities react to dynamic changes in their environments, and become a valuable resource for understanding the effects of environmental change on the ecology of *Sphagnum* bogs.

The application of evolutionary approaches to the study of microbial communities and microbiomes
SMBE-PO-620

Gut microbiome composition and its association with the evolution of toxin sequestration in poison frogs

K. Siu-Ting [1 2,*], J. Friedersdorff [3], D. Carreño [4], B. Thomas [1], G. Griffiths [3], M. O'Connell [5], J. Newbold [6], C. Creevey [1]
[1]IGFS, School of Biological Sciences, Queen's University Belfast, Belfast, United Kingdom, [2]Museo de Historia Natural, UNMSM, Lima, Peru, [3]IBERS, Aberystwyth University, Aberystwyth, [4]GGB, University of Leicester, Leicester, [5]School of Life Sciences, University of Nottingham, Nottingham, [6]Scotland's Rural College, Edinburgh, United Kingdom

Abstract: Poison frogs secrete alkaloid toxins in their skin as defence mechanisms against predators. Numerous studies have shown that the origin of alkaloid toxins in the skin is through "sequestration from diet", i.e. uptake and storage of toxins or their chemical precursors, mostly from consumed arthropods. There exists the intriguing possibility that the gut microbiome of these frogs may play a role in this process. We address this question by looking at the organism together with its associated microbial communities, an effective symbiotic relationship between host and microbiome that could have allowed phenotypic adaptation of the host to a toxic diet. We sequenced the Bacterial and Archaeal 16S rRNA and fungal ITS1 regions of the gut microbiome of 7 poison frog species and 9 outgroup frog species caught in the rainforest of Eastern Peru. Frog species were selected based on sharing similar microhabitats and comparable individual sizes. A comparative analysis of the microbiome composition across all our samples allowed us to identify a core group of abundant symbiotic microbes unique to poison frogs in spite of intrinsic variation within species. Furthermore, using full high-throughput DNA sequencing of targeted samples, we reconstructed MAGs for a number of poison frogs and their outgroups, which enabled us to identify bacterial genomes that are very similar in composition and function in poison frogs. We speculate that some of the gene families identified in these MAGs may be associated with the evolution of toxin sequestration in these frogs.

The Causes of parallel molecular evolution

SMBE-PO-629

2b-RAD genotyping for linkage mapping in a pair of polymorphic butterfly species

J. J. Hanly [1,*], A. Martin [1], C. Wheat [2], A. Porter [3], V. Ficarrotta [1], P. Rastas [4]

[1]Department of Biological Sciences, The George Washington University, Washington, United States, [2]Department of Zoology, Stockholm University, Stockholm, Sweden, [3]Organismic and Evolutionary Biology, University of Massachusetts, Amherst, United States, [4]University of Helsinki, Helsinki, Finland

Abstract: Butterfly wing patterns frequently vary at the within-species and between-species level. A combination of association and linkage analyses has allowed the recent identification of several gene regulatory loci responsible for such pattern differences in several linages of butterflies, including the ligand WntA and the transcription factors Optix and BarH1. The butterfly wing presents a particularly useful system for examining the evolution of regulatory differences between linages, because while the patterns change, the substrate upon which they vary is an approximately 2D structure that does not undergo complex morphogenetic rearrangements during development.

The species *Colias eurytheme* and *C. philodice* recently came into sympatry in the Eastern United States due to anthropogenic changes to host plant distributions, and can be hybridized in the lab. The two species have several differences in wing patterns, including the level of pterin pigmentation and the shape of melanic bands. Both species also have a low-frequency female-limited white morph, *alba*, which is found in many *Colias* species and is likely an ancient balanced polymorphism. We sequenced the genome of *C. eurytheme,* and linkage maps were constructed for hybrid broods in which wing pattern differences segregated, allowing the identification of genomic loci responsible for these traits. Variation in recombination rate across the genome suggests there are barriers to hybridisation between the two species. The Eastern-US *Colias* system will allow the identification of additional causative loci for butterfly wing pattern evolution and could provide a good genomic case study for the maintenance of species barriers after secondary contact.

The Causes of parallel molecular evolution

SMBE-PO-648

Defining the Role of Cis-Acting DNA Sequence in Evolutionary Divergence of the Bivalent Chromatin State in Mammals

K. Griffin [1,*], B. Lesch [1]

[1]Genetics, Yale University, New Haven, United States

Abstract: Changes in gene regulation play a major role in evolution. Bivalency is an important developmental gene regulatory state consisting of the activating histone mark H3K4me3 and repressive histone mark H3K27me3 occurring on the same nucleosomes. Bivalency is present in sperm and spermatogenic cells throughout mammalian evolution, and is located in the promoters of transcriptionally silent genes important for embryonic somatic development. This is thought to poise developmental genes in sperm to be quickly activated or repressed as cells differentiate during embryogenesis. It is currently not understood how the bivalent state is regulated. We are using a comparative evolutionary approach to determine what role genomic sequence plays in recruiting epigenetic writers to cause a gene to be bivalent. The gene TRAF6 is transcriptionally silent and bivalent in mouse embryonic stem cells (mESCs) and spermatogenic cells, while it is transcriptionally active and marked only by the activating modification H3K4me3 in spermatogenic cells of other mammalian species. In order to determine what genomic sequence designates TRAF6 as bivalent in mouse but not in other species, we replaced the beginning segment of mouse TRAF6, promoter, first intron, and first exon, with the corresponding region from the human genome in mESCs. We inserted GFP immediately after the replaced region, allowing us to track the transcriptional consequences of changes in chromatin state at the inserted sequence. We evaluated histone modification and gene expression states in the engineered mESCs, allowing us to determine the effects of divergence in promoter sequence on bivalency at the TRAF6 locus.

The Causes of parallel molecular evolution

SMBE-PO-647

Differential codon usage preferences are conserved in in-paralogous genes throughout Chordata

F. Borveto [1,*], J. Bourret [1], I. G. Bravo [1]

[1]CNRS, Montpellier, France

Abstract: The Polypyrimidine Tract Binding protein 1 (PTB) acts as a key splicing regulator, expressed constitutively in most cell types.In chordate genomes,PTB1 presents two in-paralogs, derived from two ancient duplication events, and whose expression is reported to be tissue-restricted: PTB2, expressed in neurons, and PTB3,expressed in hematopoietic cells. The PTB2-3 in-paralogs are identical up to 70-80% to PTB1 at the amino acid level. Here we have used the PTB1-2-3 genes of 60 divergent chordate species in order to study their phylogenetic relationships and codon usage preferences, and assess the relationship between both.The phylogenetic reconstruction of these sequences shows three well-supported subtrees, one for each paralog,where the internal structure of each subtree essentially recapitulates the species tree.This suggests that the emergence of the in-paralogs occured before the split of the different chordate clades.The codon usage analysis revealed that, independently of the species and its taxonomic clade, the three paralogs have distinct codon usage preferences: PTB1 displays high GC3 while PTB2 is AT3-rich in all species.PTB3 also shows an AT3-rich profile, albeit less marked than PTB2.

Our results suggest that a sustained combination of directional mutational and/or selection pressure has differentially shaped codon usage preferences of the PTB in-paralogs in chordates.We interpret that this directional evolution and the tissue-restriction of the present day PTB proteins is compatible with Ohno's model of evolution after gene duplication, where duplicates accumulate mutations allowing the gene/s to explore the sequence space and to acquire new functions or maintain the original function with a different spatio-temporal expression pattern.

The Causes of parallel molecular evolution

SMBE-PO-630

The search for bird mitochondrial genome adaptations to high altitude, migration, diving, wintering and flight

V. Burskaia [1,*], G. Bazykin [1], I. Artyushin [2], N. A. Potapova

[1]Skolkovo Institute For Science And Technology, [2]Moscow State University, Moscow, Russian Federation

Abstract: There is a quite widespread opinion that mitochondrial genes could be involved in adaptation to lifestyle that requires elevation of metabolism rate. We aligned mitochondrial gene coding regions of 250 bird species. All the species were divided into 6 phenotypic groups. The groups of interest are: high-altitude birds, divers, long-range-migrators, flightless birds and wintering at high latitude birds. As control group we used tropical birds. We applied phylogenetic GWAS approach to detect mutations, associated with phenotype changes. We found no significant associations. Probably it means that mitochondrial genes do not participate in appearence of aforementioned adaptations. It is also possible that different mutations in one gene may cause similar adaptations. Now we are testing last suggestion.

The Causes of parallel molecular evolution

SMBE-PO-627

Same destination, but different roads? No evidence for genome-wide parallel evolution in Arctic plants

S. Birkeland [1,*], A. L. S. Gustafsson [1], A. K. Brysting [2], C. Brochmann [1], M. D. Nowak [1]

[1]Natural History Museum, University of Oslo, [2]Centre for Ecological and Evolutionary Synthesis, Department of Biosciences, University of Oslo, Oslo, Norway

Abstract: Most published examples of parallel molecular adaptation involve only a few genes, but it is not until relatively recently that we could investigate evolutionary repeatability at the scale of entire genomes. Genome-wide parallel evolution could be expected when species independently evolve similar broad suites of adaptations, e.g. in response to the same extreme environmental conditions. Arctic plants are such an example, as they must have evolved a range of adaptations in response to extremes in light and temperature. In this study, we aimed to evaluate the extent of parallel molecular adaptation in three Arctic plant species representing different clades within the Brassicaceae family. We used the branch-site test to detect positively selected genes in each of the Arctic focal species, and compared the results to a set of non-Arctic controls (~13,000 alignments/species). As a complementary approach, we also calculated the posterior expected numbers of convergent and divergent substitutions between pairs of Arctic branches. We found no evidence that the same genes were under selection in the three Arctic species, and only fourteen genes appear to be under selection in two Arctic species. The posterior expected numbers of convergent substitutions, greatly reflected the results of the selection tests, but only two genes co-occurred in the results of both tests (i.e. genes with convergent substitutions in Arctic plants that were also under positive selection). Comparison of the positively selected gene sets gave some evidence for functional overlap associated with Arctic adaptation. In all Arctic species, the positively selected gene sets were enriched for gene ontology (GO) terms possibly tied to predefined categories of stresses associated with life in the Arctic. This was also the case for genes with putative convergent substitutions in the Arctic focal species. Our results support the idea that a functional syndrome of Arctic adaptation exists, but that different lineages may have taken very different routes resulting in adaptation to the Arctic habitat.

The Causes of parallel molecular evolution

SMBE-PO-644

Parallel losses of ultraviolet vision in high-duty-cycle echolocating bats

L. Li [1], Y. Xia [1], F. Liu [1], H. Liu [1], Y. Zhu [1], S. Zhang [1], S. J. Rossiter [2], J. S. Patel [3], Y. Liu [1,*]

[1]Shenyang Agricultural University, Shenyang, China, [2]Queen Mary University of London, London, United Kingdom, [3]University of Idah, Moscow, United States

Abstract: Recent molecular data indicate that UV vision is more common in mammals than previously thought. In bats (order Chiroptera), widespread UV vision is conferred by the short wavelength-sensitive type 1 (SWS1) pigment, but has been lost by some lineages. Recorded losses of SWS1 pigments in horseshoe and roundleaf bats have been suggested to have arisen because of a sensory trade-off between UV vision and their derived system of high-duty-cycle (HDC) echolocation. Some support for this idea also comes from bats of the *Pteronotus parnellii* species complex - which have independently evolved HDC echolocation - however, the evidence is less clear-cut. Although some studies have reported UV vision in *Pteronotus parnellii* based on an intact *SWS1* gene sequence, others have revealed species in this clade have either undergone loss-of-function mutations in their coding region (*P. mesoamericanus*), or show no protein expression (*P. pusillus*). Here we sequence two additional HDC species in this group (*P. mexicanus* and *P.* sp1) and determine the phenotype of their SWS1 pigments by *in vitro* assay. Interestingly, although both species show intact SWS1 coding sequences, neither of their SWS1 pigments show any spectral sensitivity. On the other hand, the SWS1 pigments of their common ancestor and also a close relative (*P. quadridens*) using low-duty-cycle echolocation still show UV sensitivity. The results suggest the losses of UV vision in *Pteronotus parnellii* clade occurred independently. Our findings demonstrate that relaxed selection on visual ecology might initially lead to a loss of spectral tuning, with more obvious molecular changes occurring later. More generally, we show that the presence of an intact gene and protein do not equate to a functional phenotype.

The Causes of parallel molecular evolution

SMBE-PO-643

Remarkably Repeatable Rewiring of Gene Regulation in Pseudomonas fluorescens

J. Horton [1,*], A. Wilcox [1], R. Jackson [2], M. Silby [3], N. Priest [1], T. Taylor [1]

[1]Biology and Biochemistry, University of Bath, Bath, [2]Biological Sciences, University of Reading, Reading, United Kingdom, [3]Biology, University of Massachusetts, Dartmouth, United States

Abstract: As mutation is inherently a random process, we may expect populations to follow numerous genotypic trajectories to improved fitness during adaptive evolution. Yet there is a bulk of experimental evidence showing that genetic change can be surprisingly repeatable following strong directional selection. The force of selection may well constrain the pool of available genotypes that can reach fixation to those that offer optimal phenotypes, but this effect will be the product of a given adaptive landscape which is dictated by both the genetic and environmental background. The impact such variables have on extreme parallel evolution, however, is not fully understood. Here we document a case of an astoundingly repeatable genetic change leading to a hyper-spreading phenotype in *Pseudomonas fluorescens* SBW25 and reveal the contingencies behind this phenomenon. When placed under strong directional selection for motility we observed an immensely repeatable genetic change – pertaining to a single adenine to cytosine switch at the same nucleotide site – within a nitrogen-regulatory histidine kinase locus in 97% of sequenced mutant lines. Furthermore, we observed a tight correlation between nutrient condition and the evolution of the hyper-spreading phenotype. Phenotype occurrence was found in higher frequencies as populations approached starvation – appearing in 91% of cases in minimal media over 120h. We subsequently found that this repeatability was lost over different genetic backgrounds. Our findings reveal that parallel evolution is strongly contingent on both environmental and genetic context and highlight how particular adaptive landscapes can drive repeatability to an extreme degree.

The Causes of parallel molecular evolution

SMBE-PO-628

Compensatory back mutation in mammalian genome

K. Satomura [1,*], N. Osada [1], T. Endo [1]

[1]Information Science and Technology, Hokkaido University, Hokkaido, Japan

Abstract: According to the neutral theory, most of amino acid substitutions are selectively neutral or deleterious. The deleterious substitutions would be removed before fixation by purifying selection, but fortuitously weakly deleterious substitutions are fixed as selectively neutral substitutions, for example in the case of small effective population size. To repair the fixed weakly deleterious substitution, back substitution would arise by higher probability than the expected value because the original state of amino acid is relatively adaptive compared to the current state. To test this hypothesis is the objective of this study. The amino acids substitution have high bias to some amino acids having similar biochemical features. Therefore, the probability of the amino acid back substitution is also not random. However, if the back substitution was slightly adaptive, the fixation should trend to occur earlier. In order to verify the number of back substitutions and the selective force on it, we examined the time span of multiple substitutions on mammalian homologous genes. The number of multiple substitutions were detected by maximum likelihood method and the time spans were estimated by using branch length. These values were compared to the expected values under selectively neutral. We discussed about the influence of natural selection on amino acid back substitutions in mammalian genomes.

The Causes of parallel molecular evolution

SMBE-PO-635

An experimental test of the genomic consequences of local adaptation in deer mice

S. Laurent[*], R. D. Barrett[1], R. Mallarino[2], S. Pfeifer[3], M. Foll[4], H. E. Hoekstra[5], J. D. Jensen[3]

[1]1Redpath Museum and Department of Biology,, Montreal, Canada, [2]Molecular Biology, Princeton University, Princeton, [3]School of Life Sciences, Arizona State university, Tempe, United States, [4]International Agency for Research on Cancer, Lyon, France, [5]Department of Organismic and Evolutionary Biology, Harvard University, Harvard, United States

Abstract: Cryptic coat color evolution in the Nebraskan deer mice occurred as a response to the formation of a dune field (Sand Hills) around 12000 years ago. On the Sand Hills, dark-coated wild-type mice are subject to higher predation by visually hunting predators. Light-coated mutants increase their fitness by improving their camouflage, which explains the observed correlation between substrate color and coat color. Recently, we showed that the genetic variation at Agouti, a gene known to control coat color, exhibits signatures of past positive selection. During this talk, I will present the results of a manipulative field experiment designed to elucidate the relations between phenotype, genotype, and fitness in semi-natural populations, for the coat-color/Agouti system. In this experiment, survival rates and changes in genome-wide allele frequencies (before/after predation) have been measured in controlled populations of dark and light mice in both environments. Results highlight 1) the evidence of positive selection acting on Agouti variants during the experiment, 2) the reproducibility of positive selection on Agouti across replicates of the experiment, 3) the importance of coding and regulatory variation as the genetic basis of adaptation in this system, and 4) the indirect effect of selection on linked neutral variants. Finally, I will discuss statistical issues related to the estimation of fitness in such field experiments.

The Causes of parallel molecular evolution

SMBE-PO-634

Population genomics map divergent routes to phenotypic convergence in classic Muellerian mimics

K. M. Kozak*, O. McMillan [1]

[1]Smithsonian Tropical Research Institute, Gamboa, Panama

Abstract: Mimicry has served as key examples of evolution in the wild, and especially of seemingly predictable convergence since the times of Darwin and Mueller. A classic case are the butterflies Heliconius erato and H. melpomene, which over the last three million years have spread across the American tropics, where each displays dazzling variation of aposematic wing patterns. Although the 29 races of each species form a perfectly matching mosaic of mimetic pairs across their range, only with the inception of population genomics have we gained the ability to investigate the trajectories of specific populations towards their present day location and appearance.

I present an extensive dataset of whole genomes (271 *H. erato*, 202 *H. melpomene*) from 25 wing pattern races in each species, sampled comprehensively across the the Neotropics. Phylogenomics demonstrates that *H. erato* arose during the final phase of orogeny in the Northern Andes, but only recently invaded the vast expanses of Amazonia. Strikingly, coalescent analyses and statistical cartography of *H. melpomene* reveal contrasting timing and pattern of population expansion, and gene flow across the East of the continent. Comparison of the two mimics suggests that the pattern loci governing the most complex and widespread originated recently and swept across Amazonia, leading to the exclusion of phenotypes presently restricted to the Andean valleys.

Contrary to naive expectations, the classic example of co-evolution in complex adaptive phenotypes is far from strictly coevolved. These findings suggest relatively rapid and recent evolution of the mechanisms governing pattern in *Heliconius*.

The Causes of parallel molecular evolution

SMBE-PO-642

Limits of thermal adaptation in a cold-adapted bacterium

M. Toll-Riera [1 2,*], M. Olombrada [1 2], A. Wagner [1 2 3]

[1]Department of Evolutionary Biology and Environmental Studies, University of Zurich, Zurich, [2]Swiss Institute of Bioinformatics, Lausanne, Switzerland, [3]The Santa Fe Institute, Santa Fe, United States

Abstract: Temperature is a key environmental factor that affects all organisms. In the context of on-going global warning, it is of crucial importance to understand how organisms respond to temperature increases, especially organisms inhabiting cold-environments, which cover most of the Earth's surface. To study the limits of thermal adaptation of cold-adapted bacteria we used the bacterium *Pseudoalteromonas haloplanktis TAC125,* a gamma-proteobacterium isolated from the Antarctic sea. We carried out a laboratory evolution experiment in which we evolved 30 replicate populations to high temperatures by gradually increasing the temperature, until we reached their upper thermal limit of 30°C. To understand the molecular mechanisms underpinning adaptation to temperature, we sequenced 200 clones isolated from different time points of the experiment, and therefore, adapted to different temperatures. We identified over 900 mutations, most of them involving non-synonymous single nucleotide polymorphisms. We observed a high degree of parallel evolution among clones adapted to high temperatures. Higher temperatures coincide with the appearance of mutations in the protease Lon: 91% of the clones adapted to 30°C have mutations affecting it. The Lon protease is involved in degrading mutant and misfolded proteins, suggesting, therefore, that one of the main limits to adaptation to high temperatures is protein misfolding. Other frequent mutations affected genes involved in cell wall biosynthesis and energy conversion. Lastly, clones adapted to 30°C were still able to grow at 20°C, showing that adaptation to 30°C did not entail fitness trade-offs at 20°C.

The Causes of parallel molecular evolution

SMBE-PO-636

Intercontinental genomic parallelism in multiple adaptive radiations

I. Magalhaes [1,*], J. Whiting [2], D. D'Agostino [3], P. Hohenloe [4], M. Bell [5], S. Skulason [6], A. MacColl [3]

[1]Life Sciences, University of Roehampton, London, [2]College of Life and Environmental Sciences, University of Exeter, Exeter, [3]Life Sciences, University of Nottingham, Nottingham, United Kingdom, [4]University of Idaho, Moscow, [5]Stony Brook University, Stony Brook, United States, [6]Holar University, Sauðárkrókur, Iceland

Abstract: Parallelism, the evolution of identical traits in populations diversifying in similar conditions, provides good evidence of adaptation by natural selection. Adaptive radiations are celebrated as outbursts of biological diversity on the tree of life and are useful in the study of parallelism because of often substantial evolutionary replication. Many studies of parallelism have focused on comparisons of strongly different ecotypes or sharply contrasting environments, defined *a priori*, which could upwardly bias the apparent prevalence of parallelism. In our study we used a RADseq approach to test the extent of genomic parallelism associated with continuous variation in both environments and phenotypic traits, across four independent adaptive radiations of three-spined stickleback fish (*Gasterosteus aculeatus*) spanning the northern hemisphere. We used quantitative characterizations of environmental and phenotypic variation, which are often lacking from such studies, and quantified convergence of these axes across continents. We found substantial evidence that evolution within radiations is associated with even relatively modest environmental change. Genomic parallelism is significantly greater than expected by chance for several variables and similar in its extent for phenotypic traits and environmental variables. Genetic similarity appeared to be the best predictor of the extent of parallelism between adaptive radiations with intra-continental pairs of radiations the strongest sources of parallelism. However, we also found evidence of common environments promoting genomic parallelism coming from the common acid-alkali axis experienced by all radiations, and the observations that Calcium and pH were associated with strong patterns of gene-reuse. Overall, our results show that genome-wide evolution continues to be repeatable at intercontinental scales and after hundreds of thousands of years of divergence, and provide useful insight into factors likely to influence genomic divergence at large geographical scales.

The Causes of parallel molecular evolution

SMBE-PO-637

The role of parallel molecular evolution in the emergence of highly-pathogenic avian influenza A viruses

M. Escalera-Zamudio [1,*], M. Golden [1], B. Gutiérrez [1], J. Thézé [1], J. R. Keown [2], L. Carrique [2], T. A. Bowden [2], O. G. Pybus [1]

[1]Department of Zoology, [2]Division of Structural Biology, University of Oxford, Oxford, United Kingdom

Abstract: Avian Influenza A viruses (AIVs) circulate among wild and domestic bird populations worldwide. While some strains only cause mild to asymptomatic infections, known as low pathogenicity avian influenza viruses (LP), high pathogenicity avian influenza viruses (HP) can have an extremely high mortality rate in both domestic and wild bird populations. Although virulence is a polygenic trait, the molecular determinants of virulence have been well characterised for AIVs, such as a polybasic proteolytic cleavage site within the hemagglutinin protein that enables a systemic viral spread within the avian host. We hypothesise that the parallel evolution of HP lineages from LP ancestors may have been facilitated by permissive or compensatory secondary mutations occurring anywhere in the viral genome, preceding or following the appearance of a polybasic proteolytic cleavage site. We used a comparative phylogenetic and structural approach to detect shared mutations evolving under positive selection across the whole genome of HP AIVs of the H7NX and H5NX subtypes, and developed a model that statistically assesses genotype-phenotype associations. We present cumulative evolutionary and structural evidence that supports for the association between parallel mutations and the independent evolution of the HP phenotype. Parallel mutations occur frequently among HP lineages of the same subtype. Many of the mutations detected increase viral fitness in terms of their biological properties, and are ranked as stabilising to protein structure, supporting that these are permissive/compensatory rather than preventive. The mutational panel provided here may function as an early detection system for strains under transitional virulence stages.

The Causes of parallel molecular evolution

SMBE-PO-645

Trehalase gene as a molecular signature of dietary diversification in bats and other mammals

H. Jiao [1], L. Zhang [2], H. Zhao [1,*]

[1]Wuhan University, Wuhan, [2]Guangdong Institute of Applied Biological Resources, Guangzhou, China

Abstract: Diet is a key factor in determining and structuring animal diversity and adaptive radiations. The repeated occurrences of dietary shifts represent a remarkable signature during the evolution of mammals. The mammalian fossil record preserves phenotypic evidence of many dietary shifts, whereas genetic changes followed by dietary diversification in mammals remain obscure. To test whether living mammals preserve molecular evidence of dietary shifts, we examined the trehalase gene (*Treh*), which encodes an enzyme capable of digesting trehalose from insect blood, in bats and other mammals with diverse diets. Bats represent the largest dietary radiation among all mammalian orders, with independent origins of frugivory, nectarivory, carnivory, omnivory, and even sangivory in an otherwise insectivorous clade. We found that bats have repeatedly lost *Treh* in unrelated lineages as they independently radiated into non-insectivorous niches. Consistently, purifying selection has been markedly relaxed in non-insectivorous bats compared with their insectivorous relatives. Enzymatic assays of intestinal trehalase in bats suggest that trehalase activity tends to be lost or markedly reduced in non-insectivorous bats compared with their insectivorous relatives. Furthermore, our survey of *Treh* in 119 mammal species, which represent a deeper evolutionary timeframe, additionally identified a number of other independent losses of *Treh* in non-insectivorous species, recapitulating the evolutionary pattern that we found in bats. These results document a molecular record of dietary diversification in mammals, and suggest that such molecular signatures of dietary shifts would help us understand both historical and modern changes of animal diets.

The Causes of parallel molecular evolution

SMBE-PO-646

Transitions to subterranean life drive parallel and species-specific gene expression changes in isopods

L. F. Grice [1,*], T. Lefébure [2]

[1]Université Claude Bernard Lyon 1,, Université Claude Bernard Lyon 1, CNRS, [2]Laboratoire d'Ecologie des Hydrosystèmes Naturels et Anthropisés, Université Claude Bernard Lyon 1, Villeurbanne, France

Abstract: Subterranean spaces are harsh environments, characterised by complete darkness, low oxygen levels, and limited nutrient availability. Nevertheless, surface-to-subterranean transitions have occurred repeatedly across the animal kingdom. Such transitions are often associated with phenotype changes including pigmentation and eye loss, increased lifespan, and decreased metabolic rate; thus, subterranean animals are valuable models of parallel or convergent evolution. However, it remains unclear whether these predictable phenotypes might result from altered gene expression, and if so, whether changes occur in parallel between subterranean lineages. Isopods are an ideal group for exploring this question. Although ancestrally surface-dwelling, many isopod lineages have independently adapted to life underground. Here, we have generated transcriptomes from 22 isopod populations and species. First, we searched for intraspecific gene expression changes by comparing surface and subterranean conspecifics within three species. On average, 10% of genes are differentially expressed between species' ecotypes, including orthologues of specific pigmentation and eye genes (e.g. *cytochrome P450*, *disconnected*, *rhodopsin*, and *scarlet*). Genes downregulated in subterranean populations are enriched for functions including metabolism, oxygen transportation, cuticle formation, and others. However, we find that most changes in gene expression are species-specific, suggesting that there are multiple adaptive paths to a subterranean lifestyle. We extend our analyses to interspecific comparisons in ongoing work, using eleven pairs of surface and subterranean sister species, effectively serving as natural independent replicates. Our analyses provide novel insights into the ways that gene expression can evolve in parallel, and how both shared and unique changes can underlie parallel adaptation to subterranean habitats.

The Causes of parallel molecular evolution

SMBE-PO-641

Convergent adaptation to alpine environments in three Brassicaceae species

C. Rellstab*, S. Zoller, C. Sailer, A. Tedder, F. Gugerli, K. K. Shimizu, R. Holderegger, A. Widmer, M. C. Fischer

Abstract: Theory predicts that related species develop similar genetic mechanisms to adapt to similar environmental conditions. Most studies documenting such convergent evolution have used within-species comparisons or surveyed only a limited portion of the genome. Here, we investigated whether the same or a different set of genes was involved in genetic adaptation to similar environmental gradients in natural alpine populations of three plant species. We used whole-genome pooled population sequencing to study genome-wide SNP variation in 18 natural populations of three Brassicaceae (*Arabis alpina, Arabidopsis halleri, Cardamine resedifolia*) from the Swiss Alps. First, we de-novo assembled draft reference genomes for all three species. Second, we ran population and landscape genomic analyses to look for shared signatures of selection and adaptation in response to similar habitat contrasts acting on these species. Our results show that the genomic signature of selection in heterogeneous alpine environments is partly convergent among species. The shared signals of adaptation are found at a higher frequency (26-83%) than expected by chance. The most closely related species pair showed the highest level of common adaptation signals. Moreover, genes as identified having undergone convergent adaptation were enriched for non-synonymous mutations. This suggests that these genes have a functional relevance, despite most of the identified genes having currently unknown functions. We conclude that adaptation to heterogenous alpine environments is partly convergent among related species.

The Causes of parallel molecular evolution

SMBE-PO-638

Convergent Evolution of Defensive Venom Components in Spitting Cobras

T. Kazandjian [1,*], D. Petras [2], S. Robinson [3], E. Undheim [4], K. Arbuckle [5], G. Whiteley [1], L.-O. Albulescu [1], S. Ainsworth [1], S. Wagstaff [6], W. Wüster [7], R. Harrison [1], I. Vetter [3], J. Calvete [8], N. Casewell [1]

[1]Tropical Disease Biology, Liverpool School of Tropical Medicine, Liverpool, United Kingdom, [2]Skaggs School of Pharmacy and Pharmaceutical Sciences, University of California, San Diego, San Diego, United States, [3]Institute for Molecular Bioscience , [4]Centre for Advanced Imaging, The University of Queensland, Brisbane, Australia, [5]Department of Biosciences, Swansea University, Swansea, [6]Liverpool School of Tropical Medicine, Liverpool, [7]School of Natural Sciences, Bangor University, Bangor, United Kingdom, [8]Genómica y Proteómica, Instituto de Biomedicina de Valencia, Valencia, Spain

Abstract: Snake venoms are used primarily for prey immobilisation, however three groups of spitting cobras have independently evolved the ability to forcibly eject their venom into the eyes of aggressors as a defence mechanism. This adaptation is underpinned by differences in their fang morphology, while differences in pathology following bites to humans suggests that the molecular composition of venom varies between spitting and non-spitting cobras. To investigate whether the origin of spitting has resulted in parallel molecular evolution of venom composition across spitting cobras, and whether their venoms cause increased pain in comparison with non-spitting cobras, we undertook a multi-disciplinary approach consisting of transcriptomics, proteomics and functional assays for 17 species. We found that spitting cobras have a higher abundance of phospholipase A2 (PLA_2) toxins in their venom and increased enzymatic PLA_2 activity in relation to non-spitting counterparts. Using a cell-based calcium influx assay as a proxy for the activation of sensory neurons, we show that spitting cobra venoms likely cause significantly increased pain than those of non-spitting cobras. Analyses of venom fractions revealed that PLA_2s significantly potentiate this activity. Our findings thus demonstrate that all three spitting cobra lineages have independently increased the abundance of PLA_2 toxins to increase the defensive efficacy of their venom, suggesting that defensive adaptations can drive variation in venom composition in snakes. In a wider context, our findings show that the convergent origin of morphological and behavioural adaptations can stimulate convergent evolution at the molecular level, which in turn results in complex functional phenotypes.

The Causes of parallel molecular evolution

SMBE-PO-639

Correlated Evolution of Gene Families

C. Casola [1,*], A. M. Lawing [1]

[1]Ecosystem Science and Management, Texas A&M University, College Station, United States

Abstract: Gene families are largely considered independent units that increase or decrease in size according to the evolutionary trajectory of gene gains and losses within each family. However, biological pathways usually contain genes from multiple gene families. Thus, one could expect the gain or loss of a gene in a given family might facilitate the fixation of subsequent gains or losses in functionally linked gene families. This expectation is supported, for instance, by the observed preferential retention of genes whose proteins form multi-subunit complexes following polyploidization. However, it remains unclear whether correlated changes in gene family size are widespread and if they tend to occur more often in functionally linked gene families. Using a novel phylogeny-based statistical framework we found a high proportion of correlated expansions and contractions in pairs of gene families in well-annotated genomes of angiosperm, yeast and *Drosophila* species. Randomization tests showed that 13-21% of these pairs were significantly correlated in both polyploid and nonpolyploid groups; unexpectedly, a higher proportion of significant pairs were found in nonpolyploid lineages. Most significant pairs of families were positively correlated, suggesting coordinated expansions or contractions of gene families. Using functional interaction data from the STRING and KEGG databases we showed that significantly correlated pairs of gene families were 18-60% more likely to be functionally linked—that is, shared the same biological pathway—compared to nonsignificant pairs in all taxa. These findings imply that current models of gene duplication and gene family evolution should be revisited to integrate nonindependent gene gains and losses.

The Coalescent in the Era of Population-Scale Genomics
SMBE-PO-654
Bits to Bases: Using Generative Models to Produce Synthetic Genetic Data

B. Yelmen [1], A. Decelle [2], L. Ongaro [1], F. Montinaro [1], C. Tallec [2], L. Pagani [1 3], F. Jay [2,*]
[1]Institute of Genomics, University of Tartu, Tartu, Estonia, [2]Laboratoire de Recherche en Informatique, CNRS UMR 8623, Université Paris-Sud, Université Paris-Saclay, Inria, Orsay, France, [3]University of Padova, Padova, Italy

Abstract: Availability of genetic data has increased tremendously due to advances in sequencing technologies and reduced costs. The vast amount of genetic data is used in a wide range of fields, from medicine to evolution. However, the majority of the data held by private companies and government institutions are not accessible to researchers due to privacy issues.

Using machine learning, we could generate synthetic genomes that successfully mimic the real ones but are not identical to any of them. We relied on two type of neural network architectures: (1) Generative Adversarial Networks, that were a breakthrough in the domain of computer vision, allowing the generation of extremely realistic images using deep architectures; (2) Restricted Boltzmann Machines, another family of generative models capable of learning complex data distributions. We measured the quality of the generated genomes in terms of data hidden structure, population structure, linkage disequilibrium and haplotype diversity, and demonstrated that they provided an accurate representation of the real ones. Without duplicating any of the individuals, most key characteristics of the data were conserved. We also showed a drastic improvement compared to simpler Markovian models.
A major application will be the conversion of private datasets to synthetic genomes that can then be made public without any privacy constraints. A direct implication is the increase in richness and diversity of publicly available datasets, e.g. thanks to the inclusion of minority populations currently under-represented in genetic studies.

To highlight the high potential of our approach we further demonstrated how synthetic genomes could be used in reference panels for imputation analyses leading to performances equally good as with real (hypothetically private) genomes.

The Coalescent in the Era of Population-Scale Genomics
SMBE-PO-653

Demographic inference of human population from haplotype sharing information

Y. Kawai [1,*]
[1]Department of Human Genetics, The University of Tokyo, Tokyo, Japan

Abstract: Genome-wide SNP data provides greater resolution of human demographic events such as changes in population size, migration and admixture. Identical-by-decent (IBD) or haplotype sharing between chromosomes contain information about such demographic events during recent past. Previous studies showed that the utilizing IBD information is feasible for the inference of demographic event within a few thousand years. The genetic makeup of modern Japanese people comprises of two ancestral populations; migrants in Jomon period (16 000–3000 year before present; YBP) and migrants in Yayoi period (3000–1700 YBP). However, timing and mode of demographic events after spread of population into the Japanese Archipelago is not well understood. In this study, we conducted the demographic inferences of Japanese population from genome-wide SNP data to investigate the detailed history of the population. The times of population split among Japanese populations and between Japanese population and other east Asian population were estimated from the statistics of inter-population haplotype sharing under the parametric model. For instance, the split time of Hondo (Honshu islands) and Korean population was estimated to be 101 generations ago while that of Hondo and Ryukyuan population was 70 generations ago.

The Coalescent in the Era of Population-Scale Genomics

SMBE-PO-664

Fine-Scale Structure of the Estonian Population Reveals Signature of 16-17th Centuries Bottlenecks

V. Pankratov [1,*], L. Pagani [1 2], L. Saag [1], G. Hudjashov [1 3], F. Montinaro [1], R. J. Flores [1], A. Kushniarevich [1], D. J. Lawson [4], F. Jay [5], C. Taccioli [2], M. Mitt [1], M. Kals [1], A. Metspalu [1], R. Magi [1], M. Metspalu [1]

[1]Institute of Genomics, University of Tartu, Tartu, Estonia, [2]University of Padova, Padova, Italy, [3]Statistics and Bioinformatics Group, Institute of Fundamental Sciences, Massey University, Palmerston North, New Zealand, [4]University of Bristol, Bristol, United Kingdom, [5]Laboratoire de Recherche en Informatique, Universite Paris-Sud, Paris, France

Abstract: Genetic bottlenecks played an important role in shaping genetic diversity and structure of human populations, with the Out-of-Africa event being one of the classical examples. However, when it comes to much more recent times when human populations were already big in size, it is not well understood if population declines due to wars, famine and disease could be strong enough to affect genetic diversity.

Here we aimed at reconstructing Estonian population dynamics by applying IBDNe to a dataset of 2305 whole genomes to see whether documented historical events of population drop did in fact leave a detectable genetic trace. We focus on two successive episodes of wars and associated famine and disease outbreaks (1558-1625 and 1695-1711), separated by a short period of growth, each known to cause an approximately two-fold reduction in total Estonian population size.

When treating the whole dataset together we see no sign of the above mentioned population drops in the IBDNe curve. However we hypothesize that presence of population structure at the time of the bottleneck could potentially mask its' effect, because lineages from different subpopulations will not coalesce at the time of the bottleneck. To account for that we grouped Estonian samples into genetic clusters using fineSTRUCTURE clustering algorithm and total length of IBD-segments as a measure of genetic similarity between each pair of samples. This resulted in about 90 clusters, most of which were highly geographically localized, while some had broad geographic distribution and were likely comprised of individuals with their ancestry coming from different parts of the country (these were not used in downstream analyses). We next pulled clusters together based on their position on the tree to get 4 groups corresponding to South-East (SE), South-West (SW), North-West (NW) and North-East (NE) Estonia with at least 130 samples in each group. Applying IBDNe to those 4 groups independently revealed an almost 10-fold drop in the SW subset and 2-3 fold drop in NE and NW subsets, with the lowest Ne corresponding to the period of 10-15 generations ago, and subsequent exponential growth in all 3 groups. On the other hand, SE shows long-lasting low Ne with exponential growth starting 10-15 generations ago. We assessed IBD-sharing between those 4 groups and neighboring European population as potential recent admixture with external population could affect our IBDNe results. We found a signal of recent admixture with Finns for NE and to a lesser extend NW Estonia, while SE and SW didn't differ in their levels of admixture with neighboring populations. We next applied MAPS to discover historical migration barriers and showed that SE had little gene flow from the rest of Estonia for at least 50 generations, while barriers between other parts of the country and much more recent and subtle. Such isolation could potentially explain SE having a constantly lower Ne then other Estonian subpopulations.

We conclude that when aiming at reconstructing very recent population history using coalescence framework fine-scale population structure should be taken into account as the assumption of panmixia can be violated even in populations as small and geographically localized as the Estonian one.

The Coalescent in the Era of Population-Scale Genomics
SMBE-PO-652

Fast and accurate identity-by-descent inference despite haplotype and phasing errors

W. A. Freyman [1,*], K. F. McManus [1], S. S. Shringarpure [1], E. M. Jewett [1], S. Das [1], A. Auton [1]
[1]23andMe, Inc., Mountain View CA, United States

Abstract: Estimating the genomic location and length of identical-by-descent (IBD) segments among related individuals is a central step in many genetic analyses. Because IBD segments are broken up by meiotic recombination they are expected to be longer for close relatives. However, long IBD segments are more likely to be impacted by haplotype and phasing errors compared to short segments. This makes accurate inference of phased IBD among close relatives particularly challenging. Here we present a method based off the positional Burrows–Wheeler transform (PBWT) and a probabilistic hidden Markov model (HMM) to make fast and accurate IBD estimates. We use haplotype data simulated over pedigrees to explore the performance of our algorithm against other IBD inference approaches for both distant and close relatives. Additionally we calculate our method's false positive rate and power to detect IBD segments of varying lengths.

The Coalescent in the Era of Population-Scale Genomics
SMBE-PO-663
Space is the Place: Impacts of Continuous Spatial Structure on Demographic Modeling and Association Studies
C. J. Battey [1,*], P. Ralph [1], A. Kern [1]
[1]Biology, University of Oregon, Eugene, United States

Abstract: Individuals exist in continuous space, but standard models in population genetics are based on discrete randomly-mating populations exchanging migrants. As the availability of population-level whole-genome data allows inference of increasingly fine-scale patterns of ancestry in many species, models incorporating realistic demographic and spatial processes are needed to accurately describe spatial structure and control for its impacts on analyses of selection and demography. Here we implement a forward-time simulation of molecular evolution in continuous space and use it to study the impacts of dispersal and density-dependent competition on population genetic summary statistics, demographic inference, and association studies. Low dispersal slows the geographic spread of ancestry and increases branch lengths corresponding to mid-frequency alleles because of slow coalescence among distant groups of individuals. As a result demographic inference from the site frequency spectrum is biased towards inferring a turbulent past demographic history and a recent decline in population size. The interaction of isolation by distance and spatially correlated phenotypes creates spurious signals of genetic association in GWAS and this effect is only partially corrected by including principal component positions as covariates in the analysis. We demonstrate that stratification bias is worst when environmental effects occur in spatially discrete clusters, and assess the loss of power associated with stratification corrections by simulating polygenic phenotypes in populations subject to isolation by distance.

The Coalescent in the Era of Population-Scale Genomics
SMBE-PO-651
Temporal Change of Protein Stability in Human Mitochondria during Out-of-Africa Migration
K. Fujiwara [1,*], K. Satomura [1], T. Endo [1], N. Osada [1]
[1]Graduate School of Information Science and Technology, Hokkaido University, Sapporo, Japan

Abstract: How mutations affect the fitness of the organisms is one of the central issues in evolutionary biology research. As a measurement for the effect of mutations, the thermodynamic stability of protein is one of the fundamental properties related to the protein structure and evolutional history.We approach to the issue by estimating thermodynamic stability effects of protein mutations using empirical computational optimization. In this study, we evaluated the impacts of single nucleotide variants and the combination of mutations observed in human mitochondria on COX protein stability and investigated how stability of human mitochondrial proteins changed over evolutionary time scale.
We particularly focused on mitochondria DNA haplogroups to explore how human populations have been accumulated mutations within their genomes during Out-of-Africa migration. We found rare single nucleotide variants decreased stability than common ones; especially singletons were most destabilized. This tendency supports rare alleles largely destabilize protein structure compared with common alleles, and common alleles maintain protein stability. Moreover, the COX protein of the present human populations is distributed with wide range protein stability. The European population has destabilized protein compared with the African and Asian populations with statistical significance. However, from the ancestral sequence reconstruction, the ancestral state of protein stability among all human populations were identical to each other, and these states had average neutral range stability. These results suggest that negative selection has been constantly worked and deleterious mutations were removed through evolutionary time scale.

The Coalescent in the Era of Population-Scale Genomics
SMBE-PO-650
Human haplotype genealogy and population history using massive individual genome data
M. Shimada*

Abstract: The application of the current genome-wide sequencing techniques toward human population sheds light on considerable gene flows among genus *Homo*, including both modern humans and extinct varieties of archaic humans. Because of this, discrimination between introgression from archaic humans and incomplete lineage sorting (ILS) of ancient polymorphisms is critical to estimate genetic diversity, population structure and population history in humans. Gene flow among human populations have been detected using frequencies of single nucleotide polymorphisms (SNPs) within current human populations. SNP frequency data is focusing on quantity in each population, while haplotype data is focused on association between alleles of neighboring SNPs. Thus, approaches focusing haplotype data is more suitable for detecting genomic location and estimating molecular function of genomic segments, compared to methods depending on SNP frequency data,. This is especially noteworthy when some haplotypes show characteristic distribution among populations.
To develop a strategy that can discriminate between introgression and incomplete lineage sorting of ancient polymorphisms found in modern human genomes, we focused on eight loci of human genome that have been marked or noticed for their unusual gene genealogy and geographic distribution inconsistent with out-of Africa model. For each locus, we constructed gene genealogies for haplotypes in the 1000 genomes project. Then, we performed S* analysis to estimate distinct gene flow events than out-of Africa event. Furthermore, we also estimated unevenness of selective pressure between the most diverged haplotype and others in each locus by EHH analysis.
By utilizing haplotype data of modern human populations, we could estimate haplotypes originated from gene flow event into ancestral modern humans that is discriminated from ILC. This provides a strategy of inferring population structure and history from massive individual genome data.

The Coalescent in the Era of Population-Scale Genomics
SMBE-PO-657

A General Coalescent Hidden Markov Model Framework for Inferring Complex Demographic Histories

G. Upadhya [1,*], M. Steinrücken [2]
[1]Physics, [2]Evolution and Ecology, University of Chicago, Chicago, United States

Abstract: We present a novel Coalescent Hidden Markov Model (CHMM) framework for inferring complex demographic histories from full-genome sequencing data. Our method can efficiently and accurately infer demographic parameters such as population size trajectories, divergence times, admixture proportions, and migration rates. Besides anthropologic interest, historical demography and population structure are important for establishing accurate null models when detecting selection and for correctly stratifying genome-wide association studies. Our proposed method improves upon drawbacks of existing methods by allowing accurate inference in the recent past, use of large sample sizes, and more complex demographic models.

In our CHMM framework, we use summaries of the full genealogical tree as hidden states to allow efficient computation of likelihoods for parameter inference. We present an implementation of our framework that uses either tree height (TMRCA) or total branch length as the hidden state. The requisite transition and emission probabilities are obtained by numerically computing these distributions under the ancestral process with recombination. Demographic parameters underlying data can be inferred using an EM framework.

We demonstrate, using simulated data, that our method can robustly infer the parameters of complex demographic models. The use of the total branch length, in particular, has the potential to improve inference in the recent past. Our flexible framework can be easily augmented to use other hidden states (like those in PSMC, MSMC, SMC++). Possible further extensions include combining hidden states to improve the robustness of inference and using the posterior distribution on hidden states to identify genomic variation under adaptation while controlling for demography.

The Coalescent in the Era of Population-Scale Genomics
SMBE-PO-659

Okinawa Bioinformation Bank Project: understanding human genetic diversity in the Ryukyu archipelago

M. Matsunami [1,*], M. Imamura [1 2], K. Koganebuchi [3], R. Kimura [4], C. Terao [5], Y. Kamatani [5], H. Ishida [4], S. Maeda [1 2]
[1]Department of Advanced Genomics and Laboratory Medicine, Graduate School of Medicine, University of the Ryukyus, [2]Division of Clinical Laboratory and Blood Transfusion, University of the Ryukyus Hospital, [3]Advanced Medical Research Center, Faculty of Medicine, [4]Department of Human Biology and Anatomy, Graduate School of Medicine, University of the Ryukyus, Nishihara-cho, [5]Laboratory for Statistical Analysis, RIKEN Center for Integrative Medical Sciences, Yokohama, Japan

Abstract: Understanding the regional population structure is important for medical and human evolutionary researches. The Ryukyu Archipelago is a chain of Japanese islands that stretch southwest from Kyushu to Taiwan, and is composed of dozens of islands, such as Miyako, Yaeyama, and Okinawa islands. Previous researches have shown that Japanese populations are genetically divided into two main clusters, Hondo and Ryukyu, and the genetic differentiations are observed among the island groups of the Ryukyu Archipelago. However, a detailed population structure of the Ryukyu Archipelago has not been elucidated yet. In this study, we obtained genomic DNA samples from over 10,000 individuals living in Ryukyu islands, including Miyako, Yaeyama, and Kumejima islanders, by a part of the Okinawa Bioinformation Bank Project. Among them, we genotyped over 4,000 individuals for 665,326 single nucleotide polymorphisms using the Asian Screening Array (Illumina, CA, U.S.A.). Principal component and admixture analyses revealed detailed population structure of the Ryukyu Archipelago, which have distinct clusters of each island group. These genetic differences may reflect ancient migrations and admixtures of the Ryukyu people.

The Coalescent in the Era of Population-Scale Genomics
SMBE-PO-660

Efficient simulation of admixture and local ancestry

G. Tsambos [1 2,*], P. Ralph [3], J. Kelleher [4], S. Leslie [1 2], D. Vukcevic [1 2]
[1]School of Mathematics and Statistics, [2]Melbourne Integrative Genomics, University of Melbourne, Melbourne, Australia, [3]Department of Mathematics, University of Oregon, Eugene, United States, [4]Big Data Institute, University of Oxford, Oxford, United Kingdom

Abstract: To assess the performance of methods in population genetics, we often wish to simulate realistic genetic datasets while retaining detailed information about the history of the simulated genomes. This is especially important when the consequence of admixture on patterns of genetic diversity is of primary interest.

Many existing methods can infer the ancestral origin of chromosomal segments, including MOSAIC (Salter-Townshend & Myers 2018), RFMix (Maples et al. 2013) and HapMix (Price et al. 2009). However, it is difficult to simulate chromosomes for which the true origin of those segments is known; existing approaches are approximate and ad-hoc. Recent advances implemented in the software msprime (Kelleher et al. 2016) and SLiM (Haller et al. 2018) allow us to efficiently record genetic information using a succinct tree sequence data structure, which provides unprecedented detail about the genealogy of the sample. However, for the purposes of studying admixture and ancestry, this detail can be overwhelming and difficult to analyse. We are often most interested in the ancestral population that particular genomic segments have been inherited from (i.e. the *local ancestry* of the sample). Recovering this information from the overall genealogies is challenging.

In this presentation, we will outline a method that combines these existing state-of-the-art tools with a processing step to efficiently extract local ancestry information. The simulation procedure combines a forward-in-time step to simulate admixture with a backwards-in-time step to simulate genetic diversity in the ancestral populations. These techniques allow the user to track local ancestry in large simulations under realistically complex demographic scenarios, with minimal computational overhead.

To illustrate the usefulness of this procedure, we will also show how it might be used to benchmark the performance of ancestry inference methods on various admixed populations, and to assess the degree of incomplete lineage sorting. More broadly, we anticipate that this procedure will make it easier to explore the impact of complex demographic hypotheses on detailed patterns of genetic diversity.

The Coalescent in the Era of Population-Scale Genomics
SMBE-PO-656
Identifying human population structure in South Asia
G. D. Poznik [1,*], K. Bryc [1], T. 23andMe Research Team [1], A. Auton [1]
[1]23andMe, Mountain View, United States

Abstract: To identify human population structure in South Asia, we conducted a series of unsupervised analyses on genome-wide SNP array data from more than 3,500 unrelated 23andMe research participants, all of whom indicated that their four grandparents spoke a single regional language or were born in one state or country of the subcontinent. Analyzing these data with uniform manifold approximation and projection (UMAP) and *t*-distributed stochastic neighbor embedding (*t*-SNE) revealed seven robust and identifiable clusters. One cluster consisted of Central Asian individuals, a second was comprised of individuals with Bengali ancestry, and a third was defined by a distinct subset of Gujaratis. Upon re-contacting a sample of individuals from the third cluster, we learned that, despite no pair of individuals sharing more than 100 cM identical by descent, the majority shared a single last name, indicating membership in a well-known clan. Individuals with North Indian and South Pakistani ancestry formed a fourth cluster. Keralan individuals formed a fifth cluster and were clearly differentiable from other individuals with South Indian ancestry who, together with Sri Lankans, composed a sixth cluster. Within this cluster, Sri Lankan individuals whose four grandparents spoke Sinhalese were distinct from those descending from Tamil speakers. Finally, a seventh cluster, consisting of a mixture of North and South Indian individuals, was highly enriched for individuals mentioning the word "Brahmin" in the free-text response field of our ancestry survey. Participants with ancestry from intermediate-latitude states (e.g., Maharashtra) were distributed across the fourth and sixth clusters. Results were concordant with Admixture analysis, and including ~650 individuals from well-curated external reference panels (1000 Genomes and HGDP) affirmed the quality of our survey data. We leveraged this population structure to train 23andMe's local ancestry inference engine to classify genomic segments with greater granularity than was previously possible. These findings demonstrate the potential for direct-to-consumer genetic testing, combined with online survey data collection, to reveal detailed population substructure in diverse populations from around the world.

The Coalescent in the Era of Population-Scale Genomics

SMBE-PO-661

Understanding the contribution of steppe-related migrations into western Europeans: insights from a French population

I. Alves [1,*], M. G. Blum [2], E. Genin [3], J.-F. Deleuze [4], R. Redon [1], C. Dina [1], F. Consortium

[1]l'institut du thorax, Inserm UMR 1087, CNRS UMR 6291, University of Nantes, Nantes, [2]TIMC-IMAG UMR 5525 CNRS, Univ. Grenoble Alpes, Grenoble, [3]UMR 1078 Génétique, Génomique fonctionnelle et Biotechnologies, Inserm, Univ Brest, EFS, CHU Brest,, Brest, [4]Commissariat à l'Energie Atomique, Institut Génomique, Centre National de Génotypage, Evry, France

Abstract: Multiple large-scale population movements have shaped the genomic makeup of western Europeans since the first arrival of modern humans in the continent approximately 43 kya. The geographic region of modern France, ranging from the northwest fringes of the European platform to the Mediterranean sea, represents the convergence point of most of these migration events. Therefore, understanding the peopling of France is key to shed light on the edge dynamics of ancient migrations across Europe.

In this study, we use ~850 whole-genome sequences of contemporary French individuals and merge it with publically available modern and ancient genome-wide datasets.

Rare variant sharing among the French samples shows fine-scale genetic structure, revealing a major north-southwest axis of differentiation. We also found evidence for considerable batch effect when analysing rare variant distribution between samples from different datasets, such as those from the 1000 Genome Project. By using a combination of statistical tools to test for admixture, we found that the French samples, specially those from Brittany and Normandy, show significant excess of allele-sharing with northwestern European populations from Iceland, British Isles, and Norway, and the Basques. The Basque-related ancestry was found to be larger among southwestern samples, nevertheless it is not driven by recent admixture with modern Spanish. By analysing modern samples together with publicly available aDNA, we found that the north-southwest axis of differentiation is mainly explained by larger levels of steppe-related ancestry in the northern populations in comparison with southwestern samples, which contrarily show increased Neolithic-related ancestry.

In sum, strong genetic similarities between Brittany and Normandy and other Celtic/Viking populations and their increased levels of steppe-related ancestry suggest that the ancestors of Celtic populations might have been predominantly of steppe origin likely associated with the spread of the Beaker complex across northwestern Europe.

The Ecological Genomics of Extremophile Eukaryotes

SMBE-PO-669

Population genomics of the sleeping chironomid (Polypedilum vanderplanki)

N. M. Shaikhutdinov [1 2,*], G. V. Klink [3], S. K. Garushyants [1 3], O. A. Gusev [2], G. A. Bazykin [1 3]

[1]Life Sciences, Skolkovo Institute of Science and Technology, Moscow, [2]Institute of Fundamental Medicine and Biology, Kazan (Volga region) Federal University, Kazan, [3]Institute for Information Transmission Problems (Kharkevich Institute), Russian Academy of Sciences, Moscow, Russian Federation

Abstract: The larvae of the sleeping chironomid (*Polypedilum vanderplanki)* from Nigeria is a unique example of survival in desiccated state (anhydrobiosis) and returning to an active life after rehydration. In a comparative genomic analysis of a desiccation-tolerant midge and a related desiccation-sensitive (*P. nubifer)* it was discovered that *P. vanderplanki* have multiple paralogs encoding desiccation-specific proteins that organized in clusters. The study of extremophiles with methods of population genomics and evolutionary analysis allows us to better understand the evolutionary adaptations to extreme environmental conditions.

We analyzed the WGS data from 6 populations *of P. vanderplanki* which can be divided to northern and southern populations. Analysis of the genetic structure of Nigerian populations showed a strong division (F_{ST} = 0.4 - 0.5) between populations probably due to the low gene flow, since midges live 2 days in imago stage and fly poorly. Also we observed 10% of genetic divergence between Nigerian and anhydrobiotic midge from Malawi that probably can be explained by high mutation rate due to extensive DNA damage caused by reactive oxygen species that increases in desiccation process.

Analyzing the frequency ratios of polymorphisms (Pn/Ps) in paralogous desiccation-specific genes in Nigerian populations data shown that most anhydrobiotic genes under negative selection. In comparison of Dn/Ds ratios of nigerian and malawian ortologs involved in anhydrobiosis desiccation-specific paralogs showed relaxation of negative selection that can be explained by ongoing selection or balancing selection.

All these results show that anhydrobiotic midges can be interesting evolution model of genome adaptation to extreme environmental conditions.

The Ecological Genomics of Extremophile Eukaryotes

SMBE-PO-668

Insights into the molecular evolution of lichen-forming fungi from a genome obtained by metagenomic sequencing

B. Greshake Tzovaras [1], F. Segers [2 3,*], I. Ebersberger [2 3]

[1]Lawrence Berkeley National Laboratory, University of California, Berkeley, United States, [2]Applied Bioinformatics, Goethe University Frankfurt, [3]LOEWE Center for Translational Biodiversity Genomics, Frankfurt, Germany

Abstract: Lichens grow on a variety of substrates in habitats ranging from tropical rainforests to cold deserts. This great niche diversification is thought to be due to the symbiosis of a photosynthetic organism, which generates energy, and a mycobiont, which can produce various substances for nutrient harvesting and protection against abiotic stresses. There are an estimated 20,000 lichen species, representing multiple independent evolutionary origins of the lichen lifestyle among fungi, yet so far the evolutionary path towards lichenization on a molecular level has seldom been investigated. To fill this void, we assembled highly contiguous genomes of both the mycobiont and the photobiont from metagenomic sequencing of the rock-dwelling lichen *Lasallia pustulata*. By a comparative analysis of fungal genomes we reconstructed the genomic changes towards the evolution of the mycobiont *L. pustulata* and its lichen clade members, the Lecanoromycetes. Overall, the lecanoromycete ancestor did not undergo a large contraction of gene families, but did lose genes involved in the breakdown of complex sugars, while secreted proteins and secondary metabolite clusters were gained on the path to lichenization. The size of the secretome, the proteins secreted by fungi to interact with their surroundings, was smaller in rock-dwelling lichens compared to clade members with different niches, which could be due to a less need for these substances in a rocky habitat. Our comparative genome analysis gives new insights into lichen evolution and diversification and our metagenomics assembly demonstrates how highly contiguous genomes can be obtained from unculturable symbionts.

The evolution of senescence: from theory to molecular data

SMBE-PO-672

Evolution of mammalian brain ageing patterns

M. Somel [1,*], Z. G. Turan [2], H. Izgi [2], D. D. Han [3], H. M. Donertas, E. Kocabiyik, P. Parvizi, J. Tung [4], P. Khaitovich [5]

[1]Biological Sciences, [2]Middle East Technical University, Ankara, Turkey, [3]Partner Institute for Computational Biology, Shanghai, China, [4]Duke University, Durham, United States, [5]Skolkovo Institute of Science and Technology, Moscow, Russian Federation

Abstract: Medawar's mutation accumulation hypothesis explains ageing by the act of genetic drift shaping old-age-specific phenotypes. Under this model, one may predict that ageing-related expression patterns should be less conserved than development-related patterns among species, and further, that genes highly expressed during ageing should evolve under weaker purifying selection than other genes. Studying mammalian brain transcriptome and comparative genome data, we find support for both predictions. In particular, genes that show elevated mRNA expression levels with age are evolutionarily less conserved than other genes, as measured by inter-species genomic comparisons of their protein coding and regulatory sequences. These late age-elevated genes are mainly enriched in immune system and apoptosis-related processes; we also observe that they are up-regulated in Alzheimer's Disease patients compared to age-matched controls, implying that the observed expression patterns may have functional consequences. Studying the recent evolution of these late age-elevated genes in human populations, we find no consistent indication of positive or balancing selection. Our results are compatible with the view that late age-expressed genes are subject to significant mutational load, which might contribute to ageing phenotypes.

The evolution of senescence: from theory to molecular data

SMBE-PO-674

Combining GWAS to detect genetic trade-offs between complex diseases and longevity. Insights from Schizophrenia.

G. Muntane [1 2,*], X. Farré [1], A. Navarro [1], E. Vilella [2]

[1]Department of Experimental and Health Sciences, Universitat Pompeu Fabra, Barcelona, [2]Hospital Universitari Institut Pere Mata, Institut Investigació Sanitaria Pere Virgili, Reus, Spain

Abstract: The genetic trade-off between fitness and late life mortality has been proposed as an evolutionary mechanism to explain the process of ageing. These trade-offs, known as antagonistic pleiotropy, are common, if not ubiquitous. This implies that a number of molecular mechanisms of longevity may be widely common among traits with fitness effects. The current availability of hundreds of genome-wide association studies (GWAS) results provides the unique opportunity to gain insight into a global view of pleiotropy and trade-offs of ageing. Combining GWAS data on early- and late-life diseases with genomics of longevity offer the opportunity of elucidating shared pathophysiology between longevity and complex diseases, testing novel hypothesis emerging from classical evolutionary theories of senescence.

We studied pleiotropy between genetic variants associated to specific complex diseases, Schizophrenia (SCZ) and Bipolar Disorder (BD), and variants liked to human longevity. Using a genome-wide enrichment analysis we demonstrated extensive polygenic overlap between SCZ and longevity; identifiying specific loci showing pleiotropic links between human longevity and SCZ, but not BD. Our results are consistent with decades of research that have demonstrated premature mortality amongst persons with SCZ, with a 15–20 year shorter life expectancy compared to the general population. We also provide novel information about pleiotropic links between variants associated to SCZ, longevity and other diseases, including antagonistic pleiotropic effects such as variants that increase both, the risk of SCZ and the expected lifespan. We finally suggests the application of a cross-phenotype approach as a powerful tool for comprehensively explore the genetic trade-offs of ageing.

The evolution of senescence: from theory to molecular data

SMBE-PO-671

Reconstruction of a cell lineage tree by using single-cell level somatic mutations with a distance matrix-based approach

S. Oota [1,*], K. Ikeo [2]

[1]Center for Advanced Photonics , RIKEN, Wako, [2]Genomics and evolutionary biology, National Institute of Genetics, Mishima, Japan

Abstract: Somatic mutations are powerful signals to trace a 'history' of cells of an individual. With the somatic mutations, it is theoretically possible to infer intraindividual 'evolution' in terms of postzygotic genomes. The intraindividual evolution provides a two-fold insight: (1) developmental cell lineage trees; (2) malignant deviations from the original zygotic genome, which may lead to cancers, psychiatric diseases and ageing. Therefore, detection of genuine somatic mutations has a large biological significance. In actual data acquisitions, however, typical somatic samples are a bulk of somatic cells which are subject to mosaicism. As a result, it would be a difficult task to retrieve an informative mutation pattern from them. Recently, emerging single-cell technologies can potentially solve this issue: i.e., we are able to obtain a genuine somatic mutation pattern in a single cell without a complex and often intractable post process. By using single-cell RNA data, furthermore, we can avoid a potential problem that each single cell contains a limited amount of DNA and we have to amplify the DNA for subsequent analysis, taking a risk for unwanted errors. Here we detected candidates of somatic mutations in a sample by using Nx1-Seq, one of the emerging single-cell technologies. We constructed a cell lineage tree by using a distance matrix method, the neighbor-joining (NJ) method which has a more topology search power rather than maximum parsimony methods. Our objective is to reconstruct early-onset mutations that potentially derived diseases and ageing.

The evolution of senescence: from theory to molecular data

SMBE-PO-673

Evolution of parrots' mitochondrial genomes in terms of duplication and its relationships with longevity and body mass

P. Mackiewicz [1,*], A. D. Urantowka [2], A. Kroczak [1]

[1]University of Wroclaw, [2]Wroclaw University of Environmental and Life Sciences, Wroclaw, Poland

Abstract: Mitochondrial genomes of vertebrates are generally considered to evolve under strong selection for size reduction and gene order conservation. However, a growing number of mitogenomes with duplicated regions changes our view on the mitogenome evolution. Among Aves, order Psittaciformes (parrots) is especially noteworthy because it shows large morphological, ecological and taxonomical diversity. It offers an opportunity to study genome evolution in various aspects. Former analyses showed that tandem duplications comprising the control region with adjacent genes are restricted to several lineages in which the duplication occurred independently. However, using an appropriate PCR strategy, we demonstrate that early diverged parrot groups contain mitogenomes with the duplicated region. These findings together with mapping duplication data from other mitogenomes onto parrot phylogeny indicate that the duplication was an ancestral state for Psittaciformes. The state was inherited by main parrot groups and was lost several times in some lineages. The duplicated regions were subjected to concerted evolution with a frequency higher than the rate of speciation. The duplicated control regions may provide a selective advantage due to a more efficient initiation of replication or transcription and a larger number of replicating genomes per organelle, which may lead to a more effective energy production by mitochondria. We found that the mitogenomic duplications are associated with several phenotypic features. Parrots with the duplicated region can live longer, show larger body mass as well as predispositions to a more active flight. The results have wider implications on the presence of duplications and their evolution in mitogenomes of other avian groups. This work was supported by the National Science Centre Poland (Narodowe Centrum Nauki, Polska) under Grant no. 2015/17/B/NZ8/02402. Some computations were carried out at the Wrocław Center for Networking and Supercomputing under the grant no. 307.

The evolution of senescence: from theory to molecular data

SMBE-PO-677

Somatic mutations and genome stability maintenance in clonal coral colonies

E. H. López [1,*], S. R. Palumbi [1]

[1]Biology, Stanford University, Pacific Grove, United States

Abstract: One challenge that multicellular organisms face is maintaining genome stability in the face of mutagens and repeated cell divisions during long life spans. Imperfect genome maintenance leads to somatic mutation accumulation, which is associated with tumors and senescence in vertebrates. Colonial reef-building corals can often grow large in size, can live for thousands of years while maintaining reproductive viability, almost never develop recognizable tumors, and are thought not to have clear germ-soma segregation, so they are a pivotal group in which to understand genome maintenance. To measure rates and patterns of somatic mutations, we analyzed transcriptomes from 17-22 branches from each colony for four *Acropora hyacinthus* corals. We called putative single nucleotide variants that differed within a single colony, then re-sequenced them to identify verified somatic mutations. There is no signature of mutations caused by UV damage, indicating either higher efficiency of repair than in vertebrates, or strong sunscreen protection in these shallow water tropical animals. The frequency of mutations per branch increases with colony size, a rough proxy for age. The somatic mutation load per nucleotide in *A. hyacinthus* is on the same order of magnitude (10^{-7}) as noncancerous human somatic cells. Unlike mammals, loss of heterozygosity variants outnumber gain of heterozygosity mutations about 2:1 and there is moderate purifying selection on gain of heterozygosity mutations. Although the mutation load is similar in mammals and corals, the loss of heterozygosity and selection against amino acid changes may limit the deleterious effects of somatic mutations on the coral organism.

The evolution of senescence: from theory to molecular data

SMBE-PO-679

Comprehensive multiomics approach for the uncovering aging mechanism using animal model

J.-R. Lee [1,*], S.-H. Choe [2 3], H.-M. Cho [2 3], S.-J. Park [2], J.-S. Kim [1 3], Y.-H. Kim [2 3], J.-W. Huh [2 3]

[1]Primate Resources Center, Korea Research Institute of Bioscience and Biotechnology, Joengeup, [2]National Primate Research Center, Korea Research Institute of Bioscience and Biotechnology, [3]Department of Functional Genomics, KRIBB School of Bioscience, Korea University of Science and Technology (UST), Cheongju, Korea, Republic Of

Abstract: Because chronological age is a central risk factor for many aging-related pathogenesis including cancer, cardiovascular disease, and degenerative disease, understanding the aging mechanisms were used great potential for discoveries of therapeutic measures about aging-related pathogenesis. Recent technological advances of next generation sequencing is useful for unravel the genetic mechanisms of senescence, aging, and aging-related diseases. Furthermore, omics profiling using whole genome, epigenome, and transcriptome can offer detailed information about aging related phenomenon in an integrative manner. To further explore and characterize the relationship between somatic mutation, DNA methylation, gene expression in aging, we performed the whole genome sequencing, the targeted bisulfite sequencing, and the RNA sequencing analysis during last three years in blood samples of the old African green monkey. We identified novel differentially expressed gene (DEG) that was affected by DNA methylation according to aging. Also we identified novel transcripts via somatic mutation with age. However, these novel transcripts were not affected by DNA methylation. The extensive assays used in this study identified molecular changes associated with aging processes.

The evolution of senescence: from theory to molecular data

SMBE-PO-678

Telomere dynamics and senescence in the Seychelles warbler

D. S. Richardson [1,*], M. Hammers [2], H. S. Dugdale [3], K. Bebbington [2], L. G. Spurgin [1], S. Kingma [4], E. Fairfield, J. Komdeur [2], T. Burke [5]

[1]Biological sciences, University of East Anglia, Norwich, United Kingdom, [2]university of Groningen, Groningen, Netherlands, [3]school of Biology, University of Leeds, Leeds, United Kingdom, [4]Animal sciences, Wageningen university, Wageningen, Netherlands, [5]APS, Sheffield University, Sheffield, United Kingdom

Abstract: All organisms face challenges as they struggle to survive and reproduce. The costs and trade-offs associated with these challenges underpin the differential patterns of senescence that occur within and across species and, therefore, are fundamental to our understanding of the evolution of senescence. Telomere attrition is thought to be largely the result of oxidative stress – an inability to fully buffer the free radicals produced by physiological processes during the challenges of life. Thus, telomere dynamics may act as an indicator of biological costs and provide a generic currency with which to measure the relative impact of the various challenges individuals face. We investigate this idea by determining the causes and consequences of telomere dynamics in a natural population of the cooperatively breeding Seychelles warbler *Acrocephalus sechellensis,* studied for over 30 years. We show that adult telomere length predicts an individual's future lifespan and can be used as a marker of biological age. We also outline results that evidence the cost (in terms of telomere attrition) of key environmental, genetic and life-history factors, including early-life conditions, inbreeding, cooperative breeding and antagonistic interactions. For example, results indicate that being helped during breeding delays female senescence and may contribute to the evolution of disproportionately long lifespans within cooperative breeders. By elucidating the sources of individual variation in senescence in a natural population and linking this to genetic variation among individuals, we hope to provide a deeper understanding of the evolution of senescence.

The evolution of senescence: from theory to molecular data

SMBE-PO-681

Revisiting the influences of mutation accumulation and antagonistic pleiotropy on human senescence and disease: the case of inflammaging

A. Navarro*, E. Bosch, G. Muntané, J. A. Rodriguez, X. Farre

Abstract: The rapid progress of medical genomics is affording new data that allow testing hypothesis related to senescence and aging, both within and across species. Recently, we studied the effects of genetic variants associated with non-infectious, complex human diseases appearing at different periods in life, and made observations that fitted the Mutation Accumulation and the Antagonistic Pleiotropy theories of ageing. In particular, we observed higher risk allele frequencies and large effect sizes for late-onset diseases, and detected a significant excess of early–late antagonistically pleiotropic variants that, strikingly, tend to be harboured by genes related to ageing.

We revisit all these results in the light of new data and refine our analysis in two crucial aspects. First, we focus on the risk alleles of early-onset diseases with strong fitness effects, which should be subject to stronger selective pressures and observe that their fit to theories of senescence is even better. Secondly, we add infectious diseases into the picture, which unveils abundant pleiotropy between immune-mediated and inflammatory diseases that support ideas on inflammaging. Altogether, our results provide further systematic, genome-wide evidence for evolutionary theories of senescence in our species and contribute to the long-standing question of whether senescence is the result of adaptation.

The genetic architecture of polygenic adpatation: sweeps, small shifts and everything in between

SMBE-PO-696

Identifying the genetic basis of variation in adaptive phenotypes between two Arabis species

Y. Özoglan [1,*], J. de Meaux [1], H. Dittberner [1]

[1]University of cologne, Cologne, Germany

Abstract: Due to rapid adaptation, even sibling species can occur in contrasting ecological conditions. The endangered Arabis nemorensis is almost exclusively found in floodplain meadows and is able to withstand the harsh conditions imposed by the frequent flooding in this habitat. In contrast, its sibling species Arabis sagittata prefers dry soil and calcareous grasslands. Despite these strong differences in ecology, we recently discovered that these species sometimes co-occur and naturally hybridize leading to offsprings sufficiently fertile to allow gene flow. In semi-natural common garden experiments, we have found substantial phenotypic differences in morphology, phenology, and stress response potentially resulting from adaptation to their specific environment. We used over 1000 individuals of an F2 generation of an interspecific cross to characterize the genetic basis of these potentially adaptive phenotypes. Further, we measured hybrid fertility to identify regions of genetic incompatibilities between the parental species. We discuss our results in light of the specific ecology of the two species and propose avenues for future research aiming at identifying trait specific QTLs and linking them with regions of increased gene flow between the species. Our work supports restoration efforts aiming at the maintenance of species-rich floodplain meadows in the Rhine area.

The genetic architecture of polygenic adpatation: sweeps, small shifts and everything in between

SMBE-PO-687

Comprehensive measurements of genetic architecture in a diverse yeast cross through Barcoded Bulk QTL mapping

K. R. Lawrence [1 2,*], A. Rego-Costa [2], A. N. Nguyen Ba [2], M. M. Desai [2]

[1]Physics, MIT, [2]Organismic and Evolutionary Biology, Harvard University, Cambridge, United States

Abstract: Across organisms, studies of the genetic basis underlying complex traits have consistently observed two trends: high polygenicity, where quantitative trait loci (QTLs) are numerous, dispersed, and contributing at small effect sizes; and missing heritability, where the detected QTLs do not explain all of the genetic variance in the phenotypes. Potential sources of missing heritability include numerous small-effect QTLs below the studies' sensitivity as well as epistatic interactions between QTLs. Recent work has significantly advanced the spatial resolution of QTL mapping for identifying causal nucleotides, but at the expense of resolution for effect size and epistatic interactions. Here we demonstrate a QTL mapping analysis of a pool of 100,000 yeast segregants. Our approach combines the advantages of individual phenotyping/genotyping and bulk segregant analysis, allowing detection of QTLs with sub-1% effect sizes and, in many cases, identification of their locations down to single nucleotides. Collecting high-resolution genotype and phenotype data on this scale is achievable and cost-effective due to a suite of novel techniques: robotic liquid handling, lineage barcoding, combinatorial indexing, custom enzyme purification, and bulk fitness assays. We identify hundreds of small-effect QTLs across dozens of complex traits and quantify their contributions to missing heritability, as well as their epistatic interactions and pleiotropic effects. This flexible and powerful advance in QTL mapping enables comprehensive measurements of genetic architecture in diverse yeast crosses, which has profound implications for the evolutionary trajectories of recombining populations.

The genetic architecture of polygenic adpatation: sweeps, small shifts and everything in between
SMBE-PO-685

Detection of pervasive horizontal pleiotropy in highly polygenic human traits

D. Jordan [1,*], M. Verbanck [1], R. Do [1]
[1]Charles Bronfman Institute for Personalized Medicine, Icahn School of Medicine at Mount Sinai, New York, NY, United States

Abstract: Horizontal pleiotropy, where one variant has independent effects on multiple traits, is important for our understanding of the genetic architecture of human phenotypes. We developed a method to quantify horizontal pleiotropy using genome-wide association summary statistics. This method uses a statistical whitening procedure to remove causal relationships between phenotypes, so that any variant that affects multiple phenotypes represents a case of true horizontal pleiotropy. In simulation, our method has power to detect pleiotropy across a range of realistic scenarios, including in the presence of linkage and under varied assumptions about the heritablity of traits and the number of pleiotropic and non-pleiotropic causal loci. Applying this method to 372 heritable phenotypes measured in 361,194 UK Biobank individuals, we observed that horizontal pleiotropy is: 1) pervasive throughout the human genome and across a wide range of phenotypes; 2) primarily driven by highly polygenic phenotypes; 3) detected in 24,968 variants in 7,831 loci, a majority of which have no previously known association with any particular trait; and 4) enriched in active regulatory regions. Our detected pleiotropic loci replicate in independent datasets at a much higher rate than expected, indicating that they likely represent biologically important loci, rather than artifacts of the dataset. These results highlight the central role horizontal pleiotropy plays in the genetic architecture of human phenotypes, especially among highly polygenic phenotypes. Our pleiotropy score method, as well as a graphical Shiny app, which allows visualization and manipulation of the pleiotropy score datasets, is implemented in an R package, available on GitHub. Precomputed scores are also available for download.

The genetic architecture of polygenic adpatation: sweeps, small shifts and everything in between

SMBE-PO-683

A Whole-Genome Association Approach using genome overall feature as marker for polygenic Inter-species Trait

Q. Wu [1,*], H. Fan [2], L. Chen [3], Y. Hu [2], F. Wei [2]

[1]The State Key Laboratory of Mycology (SKLM), the Institute of Microbiology, CAS, [2]Key Laboratory of Animal Ecology and Conservation Biology, The institute of Zoology, CAS, Beijing, [3]Northwestern Polytechnical University, Xi'an, China

Abstract: GWAS is one of the most important and popular methodology developments on genetic analysis in the genome age. In the past decade, the results of GWAS have been leaning more towards the so-called ominigenic model, arguing that the polygenic traits can be determined by the majority of genes all over the genome. It suggests that the statistical feature of the whole genome has a potential role in trait determination. If a new, genome-based, global statistic can be introduced as marker, it can be expected to better describe the genetic determination of polygenic trait than the present one based on the local variation (SNP). Here, we propose a practical statistical approach that is using kmer frequencies as the genetic markers to associate genome-scale variants with polygenic traits. Unexpectedly, we found that such an approach could deal with a large scale of inter-species traits. We applied this new approach to the trait of chromosome number in 96 mammalian proteomes, and prioritized 130 genes, of which 6 were candidate genes, including TP53 and BAD. These genes were proved in the association with the cellular reactions of DNA double-strand breaks caused by chromosome fission/fusion, such as apoptosis. By explaining the kmer frequency with the oligonucleotide (oligopeptide) copy number in genome (proteome), we suggested a hypothesis that the copy number of oligopeptide allover the proteome may vary among species, which contributes to the determination of the given traits. We also found that the kmers (oligopeptides in this work) may overlap each other to cover certain protein domain, particularly the intrinsic disorder (ID) region It is noticeable that the ID region involves several omic function with global features including alternative splicing or protein-protein interaction in interactomes. It confirms to the hypothesis proposed in the ominogenic model, although in a different, inter-species level. In that sense, our study has provided a new effective genomic strategy to perform association studies by identifying genes with global effect. By using the chromosome number as a case, we hope this approach could provide an idea for exploring more various traits.

The genetic architecture of polygenic adpatation: sweeps, small shifts and everything in between
SMBE-PO-689

The genetic architecture of colonization in the presence of hybridization

T. Vision [1,*], B. Reatini [1]
[1]Biology, University of North Carolina at Chapel Hill, Chapel Hill, United States

Abstract: The role genetic architecture plays in adaptation to novel environments has received much interest when the source of adaptation variation is *de novo* mutation. Comparatively little is known about the genetic architecture of adaptation to novel environments when the source of adaptive variation is inter- or intraspecific hybridization. Here we model hybridization between divergent source populations and colonization of an unoccupied novel environment using individual-based simulations in order to interrogate the influence of genetic architecture on the timing of colonization and the mode of adaptation by hybrid populations. We find that two distinct categories of genetic architectures facilitate rapid colonization, and that they do so in qualitatively different ways. The first category is comprised of architectures with few loci and or tightly linked loci, and the mode of adaptation to the sink for such architectures is the recovery of adaptive parental genotypes. The second category is comprised of architectures with many unlinked loci and high genetic variance, and the mode of adaptation to the sink for this category is the generation of novel hybrid genotypes. We further find a tradeoff between these categories as a result of their differing modes of adaptation; the first category results in the shortest colonization lag phases across the widest range of parameter space, but further evolution may be mutation limited, whereas the second category takes longer and is more sensitive to genetic variance and dispersal rate, but may facilitate adaptation to environmental conditions that exceed the tolerance of parental populations via transgressive segregation. This influence of genetic architecture on when and how hybridization contributes to colonization may have important downstream effects on the impact of invasions and further evolution of colonizing hybrid lineages.

The genetic architecture of polygenic adpatation: sweeps, small shifts and everything in between
SMBE-PO-690
Genome-wide sexually antagonistic variants reveal longstanding constraints on sexual dimorphism in fruit flies
M. Reuter*

Abstract: The evolution of sexual dimorphism is constrained by a shared genome, leading to 'sexual antagonism' where different alleles at given loci are favoured by selection in males and females. Despite its wide taxonomic incidence, we know little about the identity, genomic location and evolutionary dynamics of antagonistic genetic variants. To address these deficits, we use sex-specific fitness data from 202 fully sequenced hemiclonal D. melanogaster fly lines to perform a genome-wide association study of sexual antagonism. We identify ~230 chromosomal clusters of candidate antagonistic SNPs. In contradiction to classic theory, we find no clear evidence that the X chromosome is a hotspot for sexually antagonistic variation. Characterising antagonistic SNPs functionally, we find a large excess of missense variants but little enrichment in terms of gene function. We also assess the evolutionary persistence of antagonistic variants by examining extant polymorphism in wild D. melanogaster populations and closely related species. Remarkably, antagonistic variants are associated with multiple signatures of balancing selection across the D. melanogaster distribution range and in their sister species D. simulans, indicating widespread and evolutionarily persistent (~1 million years) genomic constraints on the evolution of sexual dimorphism. Based on our results, we propose that antagonistic variation accumulates due to constraints on the resolution of sexual conflict over protein coding sequences, thus contributing to the long-term maintenance of heritable fitness variation.

The genetic architecture of polygenic adpatation: sweeps, small shifts and everything in between

SMBE-PO-692

Polygenic adaptation signals for height are confounded by population structure

M. Sohail[1 2 3,*], R. M. Maier[3 4 5], A. Ganna[3 4 5 6 7], A. Bloemendal[3 4 5], A. R. Martin[3 4 5], M. C. Turchin[8 9], C. W. Chiang[10], J. Hirschhorn[3 11 12], M. J. Daly[3 4 5 7], N. Patterson[3 13], B. Neale[3 4 5], I. Mathieson[14], D. Reich[3 13 15], S. R. Sunyaev[1 2 3]

[1]Division of Genetics, Department of Medicine, Brigham and Women's Hospital/Harvard Medical School, [2]Department of Biomedical Informatics, Harvard Medical School, Boston, [3]Program in Medical and Population Genetics, [4]Stanley Center for Psychiatric Research, Broad Institute of MIT and Harvard, Cambridge, [5]Analytical and Translational Genetics Unit, Massachusetts General Hospital, Boston, United States, [6]Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, Stockholm, Sweden, [7]Institute for Molecular Medicine Finland, University of Helsinki, Helsinki, Finland, [8]Center for Computational Molecular Biology, [9]Department of Ecology and Evolutionary Biology, Brown University, Providence, [10]Department of Preventative Medicine, Keck School of Medicine, University of Southern California, Los Angeles, [11]Departments of Pediatrics and Genetics, Harvard Medical School, [12]Division of Endocrinology and Center for Basic and Translational Obesity Research, Boston Children's Hospital, [13]Department of Genetics, Harvard Medical School, Boston , [14]Department of Genetics, Perelman School of Medicine, University of Pennsylvania, Philadelphia, [15]Howard Hughes Medical Institute, Harvard Medical School, Boston, United States

Abstract: Height has served as the primary example of polygenic adaptation in humans. Several studies in the last decade have claimed polygenic adaptation signals using shifts in allele frequency at height-associated loci from GWAS in present-day and ancient European and global populations. These studies have relied on the GIANT consortium meta-analysis. Here, we show that these results are significantly confounded by population stratification. Computing height polygenic scores using the recent larger and more homogenous GWAS for height in the UK Biobank (UKB), we show that polygenic adaptation signals are no longer significant (GIANT p = 1E-220, UKB p = 4E-1 for present-day Europeans). Height adaptation was also independently confirmed by a test based on the singleton density score (SDS). We show that this signal too is only present when using GIANT but not UK Biobank summary statistics. Specifically, the Spearman correlation between GWAS p-value and SDS statistic is 2E-65 using GIANT statistics but only 0.077 using UKB. We show that correlations between effect size estimates and allele frequency differences between Northern and Southern European populations underlie most of these discrepancies. The confounding with population stratification appears to be most severe for less significant SNPs; restricting the analyses to genome-wide significant SNPs results in a higher concordance between GIANT and UKB polygenic scores, and a smaller adaptation signal using both (GIANT p = 2E-4, UKB p = 2E-1). Our results suggest the need to take a step back and revisit GWAS methodology before application of existing polygenic adaptation tests to other traits.

The genetic architecture of polygenic adpatation: sweeps, small shifts and everything in between

SMBE-PO-695

Hitch-hiking laterally-acquired genes contribute to delayed adaptation

J. Olofsson [1,*], L. Dunning [1], P.-A. Christin [1]

[1]APS, University of Sheffield, Sheffield, United Kingdom

Abstract: Evidence of eukaryote-to-eukaryote lateral gene transfer (LGT) has accumulated in recent years, but the selective pressures governing the evolutionary fate of these genes as well as their adaptive value within recipient species remain largely unexplored. Here we utilize the grass system of *Alloteropsis semialata*, whose genome contains DNA fragments laterally acquired from distant grass species, to test the hypothesis that LGT spread through the recipient species and integrate into its genome by positive selection, rather than drift alone. Combining whole genome and reduced population-level sequencing, we show that long LGT fragments containing genes that add novel functions were rapidly integrated in the recipient genome through strong selective forces. Some of these LGT fragments showed genomic erosion of segments encompassing protein-coding genes, creating neutral absence/presence polymorphisms of LGT genes that persist in multiple geographic locations. Our analyses further revealed that these neutral foreign genes can become part of the standing genetic variation and are therefore spread within the recipient species by genetic hitch-hiking. One of these hitch-hiking genes underwent secondary strong selection, showing that LGT can provide a delayed advantage which can contribute to rapid local adaptation and intraspecific ecological diversification. Therefore, while short-term LGT retention is mediated by positive selection on a few genes, physically linked hitch-hikers can remain functional and contribute to the standing genetic variation with delayed adaptive consequences.

The genetic architecture of polygenic adpatation: sweeps, small shifts and everything in between

SMBE-PO-691

Transposable elements contribute to adaptive response in experimentally evolved maize populations

M. Stitzer [1,*]

[1]Department of Plant Sciences, University of California, Davis, Davis, United States

Abstract: The vast majority of most plant genomes consists of transposable elements (TEs), yet our understanding of their contributions to functional genetic variation has been limited by our inability to identify and characterize individual insertions and the difficulty identifying the targets of selection in many natural systems. Here, I capitalize on a recent structural annotation of individual TEs in the maize genome to address the contribution of TEs to adaptation in three maize populations experimentally evolved under five independent selection regimes. I assayed allele frequencies of more than 200,000 TE insertions both before and after 30 generations of phenotypic selection using pooled whole-genome sequencing. Although genome-wide copy number decreases post-selection in all populations, TE frequecy differences among populations are dominated by variation in effective population size. Nonetheless, a number of insertions disrupt likely candidate genes and show clear evidence of selection from standing variation. For example, six of the eight insertions showing the strongest change in allele frequency under selection for kernel size disrupt transcription factors important in starch biosynthesis. Overall, these populations provide a clear example of the potential of transposable elements to contribute to adaptive phenotypic change.

The genetic architecture of polygenic adpatation: sweeps, small shifts and everything in between
SMBE-PO-693

Using singletons to infer recent polygenic selection in domestic cattle (Bos taurus)

M. Hartfield [1 2,*], N. A. Poulsen [3], B. Guldbrandtsen [4], T. Bataillon [1]
[1]Bioinformatics Research Centre, Aarhus University, Aarhus, Denmark, [2]Institute for Evolutionary Biology, The University of Edinburgh, Edinburgh, United Kingdom, [3]Department of Food Science, [4]Center for Quantitative Genetics and Genomics, Aarhus University, Aarhus, Denmark

Abstract: Detecting regions of the genome that have been subject to natural selection is a major focus of evolution research. While there has been considerable success with discovering individual genes exhibiting footprints of historical selection, quantifying ongoing evolution has proved more challenging. This is especially true when multiple genes contribute to a phenotype ('polygenic' selection). Domesticated species are characterised by strong artificial selection on traits that are useful for humans (e.g., milk production in dairy cattle). Yet these phenotypes have a polygenic basis. While many individual genetic regions have been associated with them, it remains to be determined how those elements act in concert to influence phenotypic evolution, and over what timescales selection acts. The 'Singleton Density Score' (or SDS) was recently introduced to detect ongoing selection in humans, by identifying SNPs that are surrounded by very few singletons (variants in a dataset that are unique to an individual). Here we investigate the extent of contemporary selection in *Bos taurus*, using a modified SDS method. We first perform a genome-wide scan to determine individual genetic regions that appear to be subject to ongoing selection. We then study the distribution of SDS scores close to genetic regions associated with milk protein content to determine how artificial selection has shaped the recent evolution of this domesticated trait. Our study clarifies the genetic basis underlying the evolution of complex phenotypes.

The genetic architecture of polygenic adpatation: sweeps, small shifts and everything in between

SMBE-PO-694

Maintenance of adaptive dynamics in a bottlenecked population that retained strict outcrossing.

M. Takou [1,*], T. Hämälä [2], K. Steige [1], H. Dittberner [1], V. Castric [3], L. Yant [4], X. Vekemans [3], O. Savolainen [5], J. de Meaux [1]

[1]Botanical Institute, University of Cologne, Cologne, Germany, [2]University of Minnesota, Minneapolis, United States, [3]University of Lille, Lille, France, [4]University of Nottingham, Nottingham, United Kingdom, [5]University of Oulu, Oulu, Finland

Abstract: During range expansion, populations at the front of the expansion wave deal with decreased genetic variation and accumulation of deleterious mutations, which may alter their ability to adapt locally. We used whole genome sequences to investigate the genetic differentiation and demographic history of two *Arabidopsis lyrata* spp *petraea*, populations growing at the range edge and in the likely refugium of the species. We further tracked signatures of positive and negative selection in the genomes of the two local populations. In the Northern population, we find signs of a stronger bottleneck than in Southern population. Yet, S-allele diversity remained unaltered, indicating that, in contrast to what has been observed in other plant species, negative frequency-dependent selection on self-incompatibility alleles has remained efficient. Moreover, signatures of positive selection are not less frequent in the Northern population, despite its lower genetic diversity. These results suggest that, in *Arabidopsis lyrata* spp *petraea*, the range-edge bottleneck has not compromised the adaptive dynamics.

The molecular basis of major transitions in evolution

SMBE-PO-725

Improving orthology and paralogy detection after whole genome duplication sheds light on Teleosts evolution

E. Parey [1,*], H. Roest Crollius [1], C. Berthelot [1]

[1]Institut de Biologie de l'ENS (IBENS), Département de biologie, École normale supérieure, CNRS, INSERM, Université PSL, Paris, France

Abstract: Whole Genome Duplications (WGD) have a significant impact on the long term evolution of species, and correlate with several major evolutionary transitions across the eukaryotic kingdom. Yet, the evolutionary mechanisms affecting genomes after WGD are still poorly understood. Here we study the WGD that occurred at the root of the teleost fish species tree, dated 320 Mya. This WGD was followed by a major evolutionary radiation, leading to the impressive diversity of teleosts, which make up half of the extant vertebrate species, and provide a dataset of particular interest to study the evolutionary consequences of WGD.

Shortly after a WGD, redundant copies of genes are massively lost, but an important subset remain in two copies after hundreds of million years of evolution, in a process that is not well understood (35% in the human genome after the vertebrate WGDs, 26% in zebrafish after the teleost WGD). The evolutionary history of these paralogous genes can be reconstructed in phylogenetic gene trees. However, although sophisticated methods have been developed to compute gene trees, errors remain when sequences lack sufficient signal to correctly sort out orthologs and paralogs across species.

We have developed a gene tree correction method that leverages the conservation of relative gene positions along chromosomes (conserved synteny) to identify orthologous from paralogous genomic regions across species. Indeed, orthologous genomic segments share a longer common ancestry and are more similar in their patterns of gene retention, loss, and overall molecular evolution. This observation remains true despite disruption of the strict order of genes due to small-scale rearrangements. On this basis, we identify and correct gene tree topologies that are inconsistent with the gene synteny context. By integrating information from the sequence alignment, the species phylogeny as well as local synteny, we propose optimized gene evolutionary histories consistent with the known WGD event and gene sequence evolution.

We applied our correction to a dataset of almost 14,000 gene trees from the Ensembl Compara database (version 94) containing 47 teleost species. We correct 23% of teleost subtrees, thus providing a WGD-consistent solution for a substantial fraction of Ensembl Compara trees. Further, application to gene trees containing only ten teleost genomes reveals that the corrected fraction increases with the number of teleost species in Ensembl Compara.

Finally, we explored functional annotations of genes retained as duplicates in all species or differentially retained in a subset of species. The corrected tree set allows us to link duplicate gene retention to developmental and signaling pathways that coincide with teleostean evolution. Moreover, our results suggest that the reciprocal loss of alternative copies of essential genes may have contributed to reproductive isolation and emergence of new species in Teleosts. We will next extend our comparative analysis to integrate over sixty teleost species and provide insight into the evolutionary significance of whole genome duplications.

The molecular basis of major transitions in evolution

SMBE-PO-726

The origin of novel genes through sequence divergence

N. Vakirlis [1,*], A.-R. Carvunis [2], A. McLysaght [3]

[1]Benaki Phytopathological Institute, Athens, Greece, [2]Department of Computational and Systems Biology, Pittsburgh Center for Evolutionary Biology and Medicine, School of Medicine , University of Pittsburgh, Pittsburgh, United States, [3]Department of Genetics, Trinity College Dublin, Dublin, Ireland

Abstract: Novel genes are important drivers of adaptation and phenotypic innovation, including major evolutionary transitions. For example, the evolution of eusociality in ants and bees went together with the appearance of many novel genes that show caste-biased expression. It has been widely presumed that novel coding sequences evolve mainly through duplication and divergence of existing genes, with homologous sequences eventually diverging beyond any detectable similarity and thus presenting as completely novel. To date, no study has produced quantifiable evidence from real, non-simulated data to characterize this major evolutionary mechanism and measure its overall impact on genomes.

Here we harnessed conserved synteny to understand, for the first time, how the process of divergence produces entirely novel genes. We measure the rate of this process and how it varies across lineages. Our results show that divergence accounts for a minority of genes without similarity, the majority of which can therefore be attributed to processes such as de novo gene emergence. Finally, we show that lineage-specific divergence beyond detectable similarity is frequently associated with a drastic reduction in length of the coding sequence, pointing to a process akin to pseudogenization that could partly account for the short gene lengths often associated with novel genes. Our work offers important insights into a poorly understood, universal evolutionary phenomenon.

The molecular basis of major transitions in evolution

SMBE-PO-702

Heat production may have been the initial driver of proto-mitochondrial endosymbiont maintenance in the early eukaryotic cell

C. D. Dunn [1 2,*]

[1]Institute of Biotechnology-HiLIFE, University of Helsinki, Helsinki, Finland, [2]College of Sciences, Koç University, Sariyer-Istanbul, Turkey

Abstract: Eukaryotic cells are characterized by a considerable increase in subcellular compartmentalization when compared to prokaryotes. Most evidence suggests that the earliest eukaryotes consisted of mitochondria derived from a bacterial ancestor enclosed within an archaeal host. However, what benefits the archaeal host and the proto-mitochondrial endosymbiont might have obtained from one another at the inauguration of this endosymbiotic relationship remains unclear. Ancestral archaea were hyperthermophiles, and I argue that heat generated by internalized proto-mitochondria initially permitted an archaeon living at high temperatures to sample cooler environments. Furthermore, endosymbiont heat generation would have provided phenotypic flexibility not available through piecemeal acquisition of alleles selected for fitness at specific temperatures. A role for heat production by the proto-mitochondrion bridges a conceptual gap between initial endosymbiont entry to the archaeal host and a later role for mitochondrial ATP production within the eukaryotic cell.

The molecular basis of major transitions in evolution

SMBE-PO-727

A comparative genomics approach to identify genomic elements implicated in the evolution of placental mammals

A. S. Taylor [1 2,*], T. A. Walsh [3], B. Constantinides [1], L. Hume [2], D. Orr [1], H. Tinning [2], N. Forde [2], M. J. O'Connell [4]

[1]School of Biology, [2]School of Medicine, University of Leeds, Leeds, United Kingdom, [3]School of Biology, Dublin City University, Dublin, Ireland, [4]School of Life Sciences, University of Nottingham, Nottingham, United Kingdom

Abstract: The mammal placenta emerged once on the therian lineage ~180 million years ago. Eutherian placenta is unique in the level of morphological variation observed. It is most likely that the evolution and diversification of therian placenta required changes in both protein coding genes and regulatory elements. Considering the dynamic nature of gene expression during the development of placenta, miRNAs are likely to be key regulatory molecules in this process. Here we present the results of our analyses into the role of innovation at the protein coding and regulatory level in the emergence of this important tissue. We have identified 115 genes that have undergone positive selection on the stem eutherian lineage and those sites have not subsequently altered in any extant eutherian lineage. These genes function in for example angiogenesis, cell proliferation and immune response. We extracted 232 microRNAs from the literature with a putative role in placental function, and placed their emergence on the phylogeny. We show that 6 novel miRNAs emerged on the stem therian lineage, 2 of which were never subsequently lost, and 106 novel miRNAs emerged on the stem eutherian lineage - 12 of which were never subsequently lost. In total, 81 of the 115 gene families with evidence of positive selection were found to be regulated by the 14 conserved miRNAs. We show some of these miRNAs are implicated in physiological cues important for early pregnancy events.

The molecular basis of major transitions in evolution

SMBE-PO-724

Selection on a voltage-gated sodium channel gene (scn4aa) is associated with ecologically mediated changes in electric organ discharge in South American electric fishes

F. Hauser [1,*], A. Van Nynatten [1 2], W. Crampton [3], N. Lovejoy [1]

[1]Biological Sciences, University of Toronto Scarborough, [2]Cell and Systems Biology, University of Toronto, Toronto, Canada, [3]Biological Sciences, University of Central Florida, Orlando, United States

Abstract: Electrocommunication – the ability to interact with the environment through the generation of electric signals – is an evolutionary innovation that has convergently evolved in several fish lineages. South American weakly electric fishes (Gymnotiformes) are among the most species-rich groups of fishes in the Neotropics, likely in part due to the evolution of a specialized electric organ that can produce diverse, species-specific electric organ discharges (EODs) of varying complexity. Neofunctionalization of a voltage-gated sodium channel gene (scn4aa) enables this novel organ to produce an electric discharge. Here, we leverage the link between variation in scn4aa and differences in EOD waveform to investigate the relationship between genotype, phenotype, and ecology in this complex trait. We combine a comprehensive sampling of scn4aa sequences from Gymnotiformes, EOD waveform recordings, and biogeographic data to test whether ecological transitions from regions of high to low predation (selecting for EODs of high and low complexity, respectively) have shaped selection on scn4aa. Using molecular evolutionary and comparative phylogenetic analyses, we find that convergent shifts in the strength of selection in scn4aa from different electric fish lineages coincide with transitions between EOD types. We also identify increases in scn4aa molecular evolutionary rates in species that have evolved in the absence of predators, and therefore likely require less complex EODs. Finally, we model amino acid substitutions in scn4aa that may underlie parallel shifts in protein function. Together, this work sheds light on the selective forces underpinning major evolutionary transitions in electric signal production at the molecular and phenotypic level.

The molecular basis of major transitions in evolution

SMBE-PO-729

Evidence for post-transcriptional regulation of dosage compensation in platypus

W. Horspool [1,*], S. Waters [2], C. Conlan [3], P. Waters [1]

[1]School of Biotechnology and Biomolecular Sciences, [2]Faculty of Medicine, University of New South Wales, Sydney, [3]Research School of Biology, Australian National University, Canberra, Australia

Abstract: Therian (placental and marsupial) mammals possess an XX female: XY male sex chromosome system. It was proposed that the single X in males was upregulated to be functionally equivalent to two X chromosomes. This over expression from the X was proposed to carry through to females, which would result in a functional tetrasomy. However, one X in females is subject to X inactivation restoring parity between the sexes and with the autosomes. In more recent years, several RNA-seq studies have cast doubt about whether or not X upregulation occurs. In male platypus (a monotreme) there are five X and five Y chromosomes. Unlike in therian mammals, there is no evidence for chromosome wide transcriptional inactivation of any X. Here, we consider the possibility of post-transcriptional regulation of X gene dosage. We analysed platypus X gene dosage at three different points from the genome to the proteome: 1) total mRNA; 2) mRNAs bound to ribosomes; 3) total protein abundance. As expected we observed that male X total mRNA levels were roughly half that of both the female Xs and autosomes. However by the protein level, the X dosage was roughly equivalent between males and females. Ribosome bound mRNA levels were observed to be in between protein and total mRNA levels. These results highlight that mRNA and protein levels do not always correlate and is the first large scale evidence that the dosage of X genes can be corrected post transcriptionally. This elucidates how X gene dosage is compensated for in platypus and has wider implications for the understanding of dosage compensation of other species.

The molecular basis of major transitions in evolution

SMBE-PO-703

Blattodean genomes offer valuable insights into the molecular basis of the transition to termite eusociality.

M. C. Harrison [1,*], E. Jongepier [1], H. M. Robertson [2], L. P. Kremer [1], C. Schal [3], S. Richards [4], X. Belles [5], J. Korb [6], E. Bornberg-Bauer [1]

[1]Institute for Evolution and Biodiversity, University of Munster, Münster, Germany, [2]Department of Entomology, University of Illinois at Urbana-Champaign, Urbana, IL, [3]Department of Entomology & Plant Pathology, North Carolina State University, Raleigh, [4]UC Davis, Davis, California, United States, [5]Institut de Biologia Evolutiva (CSIC-UPF), Barcelona, Spain, [6]Evolutionary Biology and Animal Ecology, University of Freiburg, Freiburg, Germany

Abstract: The major transition from multicellularity to eusociality has occurred several times independently within the insects. So far most genomic investigations into this transition have concentrated on Hymenoptera, while the evolution of termites from within the cockroaches has remained rather understudied. The recent availability of termite and cockroach genomes are, however, allowing insights to be gained into the molecular processes underpinning the evolution of eusociality in this diploid, hemimetabolous insect group. By comparing genomes and transcriptomes of four Blattodean species across three levels of sociality, we have been able to identify many adaptive changes of protein sequences as well as extensive rewiring of expression patterns in several gene families with importance for the creation and maintenance of insect societies. These findings confirm the importance of a sophisticated chemical communication system for eusociality to arise. For instance, we found, several enzymes involved in the production of pheromones, such as desaturases and elongases, to have evolved caste-specific expression in the termites. Although, in contrast to ants, for example, not Odorant but Ionotropic Receptors have been co-opted for the caste-specific perception of chemical cues in termites. Furthermore, we find evidence for the rewiring of developmental pathways, such as insulin/insulin-like growth factor signaling, juvenile hormone and ecdysone synthesis, which was essential for the emergence of sterile castes from the nymphal stages of solitary ancestors. Many of our findings parallel molecular mechanisms of eusocial evolution in Hymenoptera. However, the specific solutions are remarkably different, thus revealing a striking case of convergence in one of the major evolutionary transitions in biological complexity.

The molecular basis of major transitions in evolution

SMBE-PO-701

Genome organisation and the evolution of eusociality in the honeybee (Apis mellifera)

E. J. Duncan [1,*], M. P. Leask [2], P. K. Dearden [2]

[1]School of Biology, University of Leeds, Leeds, United Kingdom, [2]Department of Biochemistry, University of Otago, Dunedin, New Zealand

Abstract:

The evolution of eusociality represents one of the major and most successful life history transitions in animal evolution. The defining feature of eusociality is the reproductive division of labor, where only one female caste is reproductively active. Understanding how this division of labour functions and how reproduction is constrained in the non-reproductive female castes is critical to our understanding of how eusociality functions and how it evolved.

In the worker caste of the honeybee (*Apis mellifera*) reproductive constraint is conditional: in the absence of the queen and brood, worker bees can activate their ovaries and lay unfertilised eggs. In previous work we have shown that a pheromone, queen mandibular pheromone (QMP) causes worker ovary activity to be constrained via a process involving Notch cell signalling. Here, we investigate the genome-wide response to QMP in honeybee ovaries using RNA-seq. We find that as the ovaries become active, a range of genes organised in clusters in the genome, become co-regulated. These clusters of genes occur more frequently than predicted by chance, and have complex evolutionary history, some evolving in the lineage leading to honeybees, and some ancient complexes found in all Hymenoptera. Further, using ChiP-seq against histone modifications, we have shown that these complexes are marked differently in repressed worker ovaries by common chromatin modifications, *before* the loss of QMP suggesting that these regions of the genome are prefigured to respond to the loss of QMP.

The presence of gene complexes and chromatin modifications imply that the genome is poised respond to QMP; and is hard-wired to produce coordinated gene regulation in response to QMP. Such mechanisms imply that QMP responsiveness has shaped the evolution of the honeybee genome. These findings will help us understand the genomic basis of the evolution of eusociality.

The molecular basis of major transitions in evolution

SMBE-PO-731

Co-evolution of opportunistic pathogen Trichuris Trichuria, has been facilitated by perpetuation of ancient genetic fragility, as a result of whole genome duplication.

A. Martin-Geary [1,*], A. Ogunkanbi [2], D. W. Newman [1], J. I. Pennock [3], D. L. Robertson [1 4]

[1]Evolution and Genomic Science, [2]Infection, Immunity and respiratory medicine, [3]Infection, Immunity, and respiratory medicine, University of Manchester, Manchester, [4]MRC-University of Glasgow Centre for Virus Research, University of Glasgow, Glasgow, United Kingdom

Abstract: Trichuriasis, caused by gut dwelling parasites in the Helminth family, is on the WHO watch-list of neglected tropical diseases, and is currently estimated to infect in the region of 1.5 billion people worldwide. Infection by the helminth *Trichuris Trichuria* has been shown to stimulate the activation of latent TGFβ1 protein in the host. However, the mechanism by which this occurs is unclear. Using the newly sequenced Trichuris genome, and large-scale human genomic data, such as duplication history, evolutionary age, haplosufficiency, and sequence divergence, we have identified an important ancient orthologue of the human TGFβ superfamily within this group of helminths. We propose that the relationship between Trichuris and vertebrates is directly linked to the co-evolution of TGFβ superfamily genes within the two species. This co-evolution has been facilitated by the expansion of the superfamily genes within vertebrates, by whole genome duplication. Whole genome duplicates (ohnologs), are known to be important ancient genes, tending to be refractory to both copy-number and gene loss due to dosage-associated deleterious phenotypes. We show how this ancient and on-going importance, and relative evolutionary inflexibility, has led to an inability to evolve away from compatibility with opportunistic pathogens such as Trichuris. This has provided ideal evolutionary conditions for co-evolution of human TGFβ superfamily genes, and the Trichuris orthologue.

The molecular basis of major transitions in evolution
SMBE-PO-706
The ARHGAP family evolutionary history

V. R. Paixão-Côrtes [1,*], H. Souza Santos [1 2], G. Reales [1 2]
[1]Instituto de Biologia, UFBA, Salvador, [2]Genetics, UFRGS, Porto Alegre, Brazil

Abstract: The ARHGAP family (SLIT-ROBO RHO GTPase activating protein), active during development to adulthood, has around 48 paralogue genes in mammals, and is involved in several cellular signaling processes such as the SLIT-ROBO pathway and the activation of Rho GTPases. The aim of this study is to describe the ARHGAP family evolutionary history, by analyzing the patterns of gene gain and loss, and estimating the evolutionary rates of ARHGAP family members. We compared the family members, in 69 vertebrate genomes, in terms of synteny, similarity and protein domains. We did not found a pattern of genes losses and gains in any specific vertebrate class, besides recent known duplications in humans. All species surveyed showed events of gene loss -and less frequently, duplications- during their evolution. We found evidences for the arising of some neurodevelopmental genes (*SRGAP1, SRGAP2, SRGAP3* and *ARHGAP4*) during whole genome duplication events. Purifying selection was the predominant force acting on the majority of genes. *ARHGAP11* showed the highest number of sites with relaxed selection constraints (11.33%), while *SRGAP3* showed the lowest (0.09%), being the most conserved gene. We investigated the evolutionary history of the ARHGAP family in vertebrates and showed an overall constant pattern in the number of genes, and also in terms of protein similarity.

The molecular basis of major transitions in evolution

SMBE-PO-705

Adaptive evolution of chemical communication in ant societies

E. Privman [1,*], R. Saad [1], A. B. Cohanim [1], S. Inbar [1], P. Cohen [1], B. Saied [1]

[1]University of Haifa, Haifa, Israel

Abstract: In the transition from solitary to social life, ants evolved complex chemical communication systems. Their diverse chemical vocabulary requires a correspondingly diverse set of chemosensory receptors. Ant odorant receptors (ORs) are the largest known family in insects, characterized by frequent expansions of subfamilies, in which duplicated ORs may adapt to detect new signals through positive selection. We inferred positive selection along every branch of the OR gene tree, identified amino acid sites under positive selection, and mapped them onto the insect OR structure. Two clusters of sites mapped to the extracellular side of the receptor, on either side of a cleft that was previously implicated in ligand activation. These results provide insights into the specific ORs and individual residues that underwent adaptive evolution, potentially explaining the elaboration of chemical signaling in ant societies.

On the complementary side, we used genomic mapping approaches to identify the genes responsible for the synthesis of pheromones. We focus on the cuticular hydrocarbons (CHCs), which are the basis for social recognition and communication. We applied multiple mapping approaches using genomic sequencing from 400 population samples of the desert ant *Cataglyphis niger* to map loci responsible for variation in CHC profiles (QTLs). We identified multiple QTLs for most of the CHCs on most of the 26 chromosomes in this species, including candidate genes coding for enzymes of long-chain fatty acid biosynthetic pathways. Together, these results form the basis for unraveling the evolution of social communication via the elaboration of complex chemical signaling and olfactory perception.

The molecular basis of major transitions in evolution
SMBE-PO-707

A chronology of multicellularity evolution in cyanobacteria

K. Hammerschmidt [1,*], G. Landan [1], F. Tria [2], T. Dagan [1]
[1]CAU Kiel, KIel, [2]HHU Düsseldorf, Düsseldorf, Germany

Abstract: The transition from unicellular to multicellular organisms is one of the most significant events in the history of life. Key to this process is the emergence of Darwinian individuality at a higher level: groups must become single entities capable of reproduction for selection to shape their evolution. Evolutionary transitions in individuality are characterized by cooperation between the lower level entities and by division of labour. Theory suggests that division of labour may drive the transition to multicellularity by eliminating the trade-off between two incompatible processes that cannot be performed simultaneously in one cell. Here we examine the evolution of the most ancient multicellular transition known today, that of cyanobacteria. We developed a novel approach for the precedence polarization of phenotypic traits that employs gene phylogenies and does not require a species tree. Applying our procedure to cyanobacterial genomes we reconstruct the chronology of ecological and phenotypic trait evolution in cyanobacteria. Our results show that the prime driver of multicellularity in cyanobacteria was the expansion in metabolic capacity offered by nitrogen fixation, which was accompanied by the emergence of the filamentous morphology and a reproductive life cycle. This was followed by a range of niche expansions and interactions with other species, and the progression of multicellularity into higher complexity in the form of differentiated cells and patterned multicellularity.

The molecular basis of major transitions in evolution

SMBE-PO-713

Phylogenomic insights into the origin of primary plastids

I. Irisarri [1,*], F. Burki [2]

[1]Biodiversity and Evolutonary Biology, Museo Nacional de Ciencias Naturales, Madrid, Spain, [2]Organismal Biology, Uppsala University, Uppsala, Sweden

Abstract: The origin of plastids (=chloroplasts) represents a key evolutionary transition in the history of eukaryotes, launching an astonishing diversification of micro– and macroscopic algae and land plants. Although the endosymbiotic theory explains well the origin of plastids from cyanobacteria, the evolution of plastids within eukaryotes remains highly contentious. In particular, the origin of primary plastids in glaucophytes, rhodophytes, and chlorophytes (collectively known as Archaeplastida) is still controversial due to continuous inconsistencies between plastid- and nuclear-based evidence. While plastids strongly support a single endosymbiosis in the ancestor of Archaeplastida, nuclear phylogenies have remained ambiguous with respect to the monophyly of Archaeplastida. Thus, discriminating among competing hypotheses is crucial to clarify the origin of primary plastids and spread across eukaryotes. Importantly, a common origin of the plastid in Archaeplastida can only be demonstrated if the host lineages are convincingly monophyletic. Here, we review the phylogenomic evidence for Archaeplastida by reanalyzing four nuclear phylogenomic datasets assembled by independent labs. We study the effect of gene and taxon sampling, which partially overlap in these four datasets, and assess data contaminations and systematic errors (model misspecifications). Using a novel alignment-splitting algorithm in combination with complex mixture models, we obtain for the first time robust evidence for the monophyly of Archaeplastida with both maximum likelihood and Bayesian inference. Our study clarifies one of the last major uncertain nodes in the broad-level tree of eukaryotes and represents convincing support from the host lineages for a single origin of primary plastids in the Archaeplastida ancestor.

The molecular basis of major transitions in evolution

SMBE-PO-718

Baby genomics: tracing the evolutionary changes that gave rise to placentation

Y. Hao [1,*], H. J. Lee [2], M. Baraboo [3], K. Burch [4], T. Maurer [5], J. A. Somarelli [6 7], G. C. Conant [1 8 9 10]

[1]Bioinformatics Research Center, North Carolina State University, Raleigh NC, [2]Division of Biological Sciences, University of Missouri-Columbia, Columbia MO, [3]Department of Computer Science, [4]Department of Psychology, Truman State University, Kirksville MO, [5]Department of Biology, Kenyon College, Gambier OH, [6]Duke Cancer Institute, [7]Department of Medicine, Duke University Medical Center, Durham NC, [8]Division of Animal Sciences, University of Missouri-Columbia, Columbia MO, [9]Program in Genetics, [10]Department of Biological Sciences, North Carolina State University, Raleigh NC, United States

Abstract: It has long been a challenge to determine the molecular mechanisms behind morphological innovations and striking evolutionary transitions such as the origin of the mammalian pregnancy. We hypothesize that the genetic toolkit of mammalian pregnancy involved repurposing of many genes that had pre-existing roles, and we want to trace how deeply in time this toolkit originated using the comparative method, which is a powerful tool for identifying key entities involved in biological functions. We have developed an orthology assignment pipeline that uses both robust, distance-based measures of gene relationships and synteny to clarify homology relationships that are ambiguous when using sequence data alone. We inferred orthology relations between human genes and genes from each of 43 other vertebrate genomes, resulting in ~20,000 orthologous pairs for each genome comparison. We then reconstructed the ancestral states of these orthologs and pinpointed orthologs that appeared before and after the divergence of eutherian mammals from marsupials. By identifying genes that are associated with the origin of placental mammals, we hope to obtain a subset of the genome that is enriched for genes that played a role in placental evolution in therian mammals. We found orthologs shared by eutherian ancestor are enriched in functions such as transcription regulations by KRAB-ZNFs, innate immune response and the MAGE protein class. Moreover, since the cellular mechanism of invasive placentas is very similar to that of the metastatic cancer cells, we also identified a list of genes that could be involved in both placenta invasion and cancer metastasis.

The molecular basis of major transitions in evolution

SMBE-PO-719

Passing it on: Serial horizontal transfer of vitamin-biosynthetic genes enables the establishment of new nutritional symbionts in aphids' di-symbiotic systems

A. Manzano Marín [1,*], A. Coeur d'acier [1], A.-L. Clamens [1], C. Orvain [2], C. Cruaud [2], V. Barbe [2], E. Jousselin [1]

[1]INRA UMR 1062 Centre de Biologie pour la Gestion des Populations, Montpellier, [2]CEA, Genoscope, Institut de Biologie François-Jacob, Every Cedex, France

Abstract: Many insects with a nutrient-restricted diet depend on obligate mutualistic bacteria for the provisioning of essential nutrients lacking from their food source, namely essential amino acids and B vitamins. Most aphids (Hemiptera: Aphididae), whose diet consists of phloem, rely on the bacterial endosymbiont *Buchnera* for the supply of the aforementioned compounds. However, in some aphid lineages *Buchnera* have lost the capability of producing these nutrients and thus the symbiotic consortium has accommodated an extra bacterial partner to supplement *Buchnera*'s deficiencies. In this work, we explore the di-symbiotic nutritional endosymbiosis of a group of *Cinara* aphids which has been found to harbour both *Buchnera* and an *Erwinia*-related symbiont. Using fluorescence *in situ* hybridisation, we have located this symbiont to the bacteriome where it inhabits its own bacteriocytes. Through whole-genome sequencing of the endosymbionts of 9 species of *Erwinia*-associated *Cinara* aphids, we have found that *Ewrinia* genomes are highly syntenic and all show significant genome reduction. Additionally, *Erwinia* symbionts display phylogenetic congruency with *Buchnera*, suggesting long-term co-divergence. Most significantly, we found that not only is *Erwinia* capable of complementing *Buchnera*'s auxotrophies, but that the genes involved in the biosynthesis of two B vitamins have actually been horizontally acquired from a *Sodalis*-related bacterium. Finally, this B-vitamin biosynthetic genes have been further transferred to a new *Hamiltonella*co-obligate symbiont in a specific *Cinara* lineage, thus displaying a tri-symbiotic system. These results highlight the important role horizontal gene transfer plays in the establishment of new obligate nutritional symbionts.

The molecular basis of major transitions in evolution
SMBE-PO-710
Genetic control of male production in Daphnia pulex
Z. Ye [1,*], C. Molinier [2], C. ZHAO [1], C. R. Haag [2], M. Lynch [1]
[1]Center for Mechanisms of Evolution, Arizona State University, Tempe, United States, [2]Centre d'Ecologie Fonctionnelle et Evolutive (CEFE), Centre National de la Recherche Scientifique (CNRS), Montpellier, France

Abstract: *Daphnia* normally reproduce by cyclical parthenogenesis, with offspring sex being determined by environmental cues. However, some females have lost the ability to produce males. Our results demonstrate that this loss of male-producing ability is controlled by a dominant allele at a single locus. We identified the locus by comparing whole-genome sequences of 67 non-male-producing (NMP) and 100 male-producing (MP) clones from five *Daphnia pulex* populations, revealing 132 NMP-linked SNPs within a single 1.1 Mb nonrecombining region on chromosome I. These markers include seven nonsynonymous mutations, all of which are located within one unannotated protein-coding gene (gene 8960). Within this single gene, all of the marker-linked NMP haplotypes from different populations form a monophyletic clade, suggesting a single origin of the NMP phenotype, and we further found evidence that the NMP haplotype did not originate within *D. pulex* but has been introgressed from a sister species, *D. pulicaria*. Methyl farnesoate (MF) is the innate juvenile hormone in daphnids, which induces the production of males and whose inhibition results in female-only production. We found that gene 8960 is sensitive to treatment by MF in MP clones, but that such responsiveness is greatly reduced in NMP clones. Thus, we hypothesize that gene 8960 is located downstream of the MF-signaling pathway in *D. pulex*, with the NMP phenotype being caused by expression change of gene 8960.
Key words: *Daphnia pulex*; methyl farnesoate; non-male producing; sex determination

The molecular basis of major transitions in evolution
SMBE-PO-709
Support for a viral origin of the eukaryotic nucleus
P. J. L. Bell [1,*]
[1]Genetics, Microbiogen, Sydney, Australia

Abstract: The prokaryote to eukaryote transition was one of the foremost evolutionary transitions in the history of life on Earth and raises two of the most challenging questions in evolutionary biology; the origin of sex, and the origin of the eukaryotic cell itself. The Viral Eukaryogenesis hypothesis makes the radical proposal that the eukaryotic cell and sex arose because the 'eukaryotic cell' is a consortium of three separate organisms: an archaeal ancestor of the eukaryotic cytoplasm, a bacterial ancestor of the mitochondrion, and a viral ancestor of the nucleus. That the eukaryotic cytoplasm is descended from an archaeon and the mitochondrion is descended from a bacterium is now widely accepted, but a viral origin for the nucleus is much more controversial. Since a membrane bound nucleus is the defining feature of the eukaryotic cell and introduced major changes at a molecular level to the way that genetic information is processed from DNA into protein, its origin is surely core to understanding the origin of the eukaryotes. Support for a viral origin of the nucleus has recently come from the discovery that some Jumbophage construct nucleus-like viral factories that share deep similarities with eukaryotic nuclei including the ability to uncouple transcription from translation, a feature previously thought to be restricted to cells possessing a eukaryotic nucleus. If the nucleus has a viral origin, the complex molecular machinery uniquely required by eukaryotes to uncouple transcription from translation should also have a viral origin. In this talk evidence is provided that the complex molecular changes underlying the ability of the eukaryotes to uncouple transcription from translation do indeed have a viral origin.

The molecular basis of major transitions in evolution

SMBE-PO-715

Ontogeny and phylogeny of gene expression in chordate embryonic development

S. Guo [1,*], H. Hu [2], C. Xu [2]

[1]Skolkovo Institute for Science and Technology,, Moscow, Russian Federation, [2]CAS-MPG Partner Institute for Computational Biology, 上海市, China

Abstract: The relationship between development and evolution has long been discussed for many generations of biologists in evolutionary developmental embryology. Early conservation model, hourglass model, and adaptive penetrance model were three main observations, but none of it is favored using the heterochrony data due to different evolutionary scale on species chosen in studies. Therefore, the relationship between ontology and phylogeny is far more intricate than we thought. And on the other side, we lack of a direct overview on the relationship between morphology and temporal gene expression change with the scale of evolutional history. In this study, we want to know whether embryonically developmental stage among species is comparable or functionally equivalent. In this study, we analyzed the dynamic gene expression of eight chordate species (amphioxus, ciona, zebra fish, two species of frogs, turtle, chicken and mouse) and one out-group species (oyster) to extend the comparison. To perform meaningful transcriptom comparison, we pair-wisely aligned developmental stages based on developmental stage-specific genes which considered as indicators of functional conserved among species. Contrary to early conservation model and hourglass model, we found parallel relationship on two species embryonic development stage is dominated for all comparisons. And based on parallel stage alignment, we found the most conserved gene expression module is the gene with highly expressed in 2cell/8 cell with the conserved Splicesome related cellular process function.

The molecular basis of major transitions in evolution

SMBE-PO-712

Amplification of signalling, regulation and protein interactions in complex prokaryotes

M. Mahajan [1,*], B. Yee [2], E. Hagglund [1], L. Guy [1], S. G. Andersson [1]

[1]Molecular Evolution, Department of Cell and Molecular Biology, Uppsala, Sweden, [2]N/A, Singapore, Singapore

Abstract: Eukaryotes show complex cell architectures, sexual reproduction, multicellularity, and many different levels of cellular organization. Prokaryotes on the other hand lack such complex traits. On average, eukaryotes have longer and more disordered proteins than prokaryotes. In fact, eukaryotes can be clearly distinguished from the prokaryotes based on these two factors. However, some bacteria were found to defy this trend, and could hardly be to distinguished from the eukaryotes. *Planctomycetes*, *Myxococcales*, *Candidatus* Magnetomorum, *Ktedonobacter* and a few other bacteria were among these bacteria. These bacteria are known to show complex traits in form of multicellularity, cell specialization and intracellular structures.

Bacteria of the *Planctomycetes* phylum, specifically the *Gemmataceae* family, are known for their complex cell architectures. A comparative genomic analysis showed massive expansions and evolution of novel domains combinations in these bacteria. These expansions and novel genes consisted of signalling, regulation and protein interaction domains. Multidomain proteins consisting of such domains are said to be associated to higher complexity in Eukaryotes. We observed similar expansions consisting of these domains in other complex bacteria as well. Additionally, we observed a higher number percentage of multidomain proteins in the complex prokaryotes. In fact, simpler protists like *Fungi* seemed to be simpler than some of the prokaryotes. Thus, the hard boundary between prokaryotes and eukaryotes is starting to blur as the novel lineages are being discovered. Furthermore, our results show that it might to be possible to systematically define complexity in terms of the genomic content.

The molecular basis of major transitions in evolution

SMBE-PO-730

Why Sex? Testing the benefits of recombination with an antibiotic resistance enzyme

D. Pesce [1,*]

[1]University of Wageningen, Wageningen, Netherlands

Abstract: The origin and evolution of sex present a major problem in biology even after a century of investigation. Sex typically involves random recombination of related DNA-sequences, which both creates novel and breaks-up existing beneficial allele combinations. Experimental studies of recombination effects are not straightforward, because benefits may occur only over many generations and key variables are difficult to manipulate in experiments with living organisms. We use a radically different approach: the laboratory directed evolution of an antibiotic resistance enzyme. Directed evolution experiments combine in vitro mutation and recombination protocols with in vitro or in vivo selection and allow extreme control over evolutionary conditions and parameters. The enzyme we use - TEM β-lactamase - catalyzes the degradation of certain antibiotics, causing resistance of its bacterial host. By using random mutagenesis and selection of bacteria expressing different β-lactamase alleles in the presence of a novel antibiotic, we follow the step-wise evolution of the enzyme towards increased specificity for the new antibiotic. By introducing *in vitro* recombination at various rates and time intervals, we systematically explore conditions and time scales for recombination benefits and test the causes of the observed recombination benefits. Our ability to perform highly controlled evolution experiments allows rigorous tests of the evolutionary effect of recombination on a complex fitness landscape.

The molecular basis of major transitions in evolution

SMBE-PO-697

The origin and evolution of vertebrate head mesoderm: a cross-species transgenic approach

T. Takahashi [1,*], H. Wang [1]

[1]Faculty of Biology, Medicine and Health, University of Manchester, Manchester, United Kingdom

Abstract: The head is the most elaborate and characteristic structure in the vertebrate body, and its evolutionary origin has long attracted the interests of scientists. This question has often been addressed by comparing vertebrates with amphioxus: a marine invertebrate closely related to vertebrates. Previous studies indicate that the vertebrate head has evolved through the elaboration of a pre-existing structure, with evolutionary inventions of cranial neural crest cells and epidermal placodes, to establish the "New" head. What still remains unresolved is the evolutionary origin of the head mesoderm of vertebrates. This is mainly because the vertebrate head mesoderm is unsegmented whereas the comparable region in amphioxus is explicitly segmented.

Using whole-mount *in situ* hybridization, we compared the spatio-temporal distribution of several marker genes for the zebrafish head mesoderm with those for the amphioxus anterior somites. We then cloned the enhancer/promoter regions of these amphioxus genes and established zebrafish transgenic lines to investigate the homologous relationship between the head mesoderm in zebrafish and the anterior somites in amphioxus.

The molecular basis of major transitions in evolution

SMBE-PO-717

Experimental evolution of collective action despite genetic conflict

N. Jahren [1,*], D. Snyder [2], J. Dewey [1], S. M. Hong [1], V. Cooper [2], M. Travisano [1]

[1]Department of Ecology, Evolution & Behavior, University of Minnesota, Saint Paul, MN, [2]Department of Microbiology & Molecular Genetics, University of Pittsburgh School of Medicine, Pittsburgh, PA, United States

Abstract: The earliest known transition to multicellular life forms occurred among cyanobacteria. This understudied transition contributed to the great oxygenation event and reshaped the planetary ecology for all evolving organisms. Current interest in the holobiont concept and investigations of the microbiome have motivated hypotheses that multicellular organisms evolved from multispecies communities. *Pseudomonas fluorescens* cells evolve a collective action strategy as an adaptation to settling selection. Specifically, cells evolve to express an extra-cellular matrix (ECM) that mediates clustering, which in turn facilitates settling. Thus, these clusters provide a system to investigate how collective action can evolve in communities. Genetic uniformity is sometimes theorized as a requirement for reducing conflict between units as they engage in collective action. However, while previous work with settling selection in microbes has produced populations that are swept by one highly-adapted strain, we observe a long-term coexistence of strains with distinct adaptations. Close inspection of the clusters reveals the presence of a "smooth" strain that does not produce ECM but is able to survive settling selection by physically associating with clusters of ECM-producers. Because ECM production comes at a cost to growth, we call this insinuation into ECM-mediated clusters a "free-riding" strategy. Comparison with the ancestral strain found that this free-riding behavior is an evolved adaptation. Whole genome sequencing on isolates from our evolved populations revealed an extraordinary degree of parallelism (and convergence) across replicates. In six of eight evolving populations, an ECM-producing isolate did not share any mutations with a smooth isolate, indicating that these phenotypes can be the product of lineages that separated early in their evolutionary history. Notable mutations included those affecting cyclic-di-GMP levels which is known to regulate cellulosic ECM production in *P. fluorescens* "wrinkly spreader" strains recovered from biofilms. Additionally, 4 of 16 isolates carry mutations in *mutS* and *mutL*, causing mutator strains. This system is notable because the free-riding strategy that evolves in the smooth strains does not disrupt the collective action of the ECM-producers. Our experiment provides a venue for multi-level selection where the ECM-producing strains adapt to the settling selection and thereby create a niche that can be exploited by the smooth strains. The emergence of evolutionary individuals at higher levels of organisation is usually conceptualized as the establishment of a mechanism that allows the exclusion of units that do not participate in the benefits of collective action. We hypothesize that the creation of new niches is a general by-product of higher level organization. Thus individuality at higher levels actually facilitates the coevolution of taxa still occupying lower levels of organization.

The molecular basis of major transitions in evolution

SMBE-PO-722

The molecular diversity and developmental expression of doublesex, a master regulator of sexual dimorphism in the Papilio polytes butterfly

R. Deshmukh [1,*], K. Kunte [1]

[1]National Centre for Biological Sciences, Bangalore, India

Abstract: The evolution of sexual dimorphism was a major transition for life on earth, resulting in considerable intra-specific morphological diversity. The molecular and developmental genetic bases of these sexual dimorphisms are poorly understood in non-model organisms. Here we study the molecular diversity and developmental expression of *doublesex* (*dsx*), which regulates sexual dimorphism in insects. *dsx* triggers sex-differentiation cascade in egg stages. It is co-opted in pupal stages to produce sexually dimorphic adaptive traits in adults. We investigated whether this co-option has a molecular basis: are co-opted functions facilitated by molecular diversification? We addressed this problem in the non-model *Papilio polytes* butterfly with whole-transcriptomes from multiple relevant tissues through the entire metamorphosis from eggs to adults. *Papilio polytes*, an iconic example of female-limited Batesian mimicry and polymorphism, has three female forms – two mimetic and one non-mimetic – produced by alleles of *dsx*. With an RNAseq dataset, we first characterised the isoform diversity and developmental regulation of *dsx* across the metamorphosis. We found that the co-opted function (wing colour patterns) was associated with diversification of isoforms in females, which were expressed in developmental stage- and tissue-specific manner. Two key stages – 5th instar larval and 3-day pupal stages – showed wing-specific *dsx* expression in mimetic females. To determine potential targets of *dsx*, we identified genes that positively co-expressed with *dsx* throughout development, which included a few known wing patterning and pigmentation genes. Our study indicates that *dsx* might achieve its dual regulation of sex determination and mimetic wing patterns by means of novel downstream targets and differential isoforms expression.

The molecular basis of major transitions in evolution

SMBE-PO-723

Macroevolutionary patterns within the eukaryotic epigenetic toolkit inferred from phylogenomic analyses

A. K. Weiner [1], Y. Yan [1], M. A. Cerón Romero [1 2], L. A. Katz [1 2,*]

[1]Biological Sciences, Smith College, Northampton MA, [2]Program in Organismic and Evolutionary Biology, University of Massachusetts - Amherst, Amherst MA, United States

Abstract: We present data on the nature of the epigenetic toolkit in the last eukaryotic common ancestor (LECA) through phylogenomic analyses of diverse eukaryotic lineages, and speculate on the impact of this toolkit on eukaryotic genome evolution. Epigenetic processes in eukaryotes are well known to play important roles in the generation of phenotypes through regulation of gene expression, control of genome rearrangements and other dynamic genome processes. As epigenetic modifications have been found to be stable across generations, they may not only change the biology of organisms but also influence the evolution of lineages. We and others have previously hypothesized about the epigenetic toolkit to be present in LECA. However, detailed studies on the composition of this toolkit so far have been largely restricted to studies of animals and plants. We use PhyloToL, which is our taxon and gene rich phylogenomic pipeline that contains transcriptomes and genomes of 714 species of all major eukaryotic clades, to make inferences on macroevolutionary patterns of gene families involved in epigenetic processes. In addition to collecting data from GenBank, we added single-cell transcriptomes from understudied clades of SAR (Stramenopila, Alveolata and Rhizaria) and Amoebozoa in order to increase taxonomic sampling. We identified a total of 120 genes from the literature that we considered as part of the eukaryotic epigenetic toolkit as they are related to either processes of chromatin modification (e.g. DNA methylation, histone acetylation) or small non-protein-coding RNAs (e.g. miRNAs, siRNAs, piRNAs). Analyzing the gene trees generated by PhyloToL, we find a broad distribution of these genes across most, but intriguingly not all, major eukaryotic clades. We also observe contrasting patterns based on function: genes involved in chromatin processing are generally conserved, while many of those involved in small RNAs have more punctate distributions. These data indicate differential macroevolutionary patterns operating on the epigenetic toolkit, and add to speculations that small RNAs may contribute to macroevolutionary phenomena such as speciation and genome conflicts.

The molecular basis of major transitions in evolution

SMBE-PO-720

An application of selective pressure analyses to guide gene discovery in amelogenesis imperfecta

G. Nikolopoulos [1,*], C. E. L. Smith [1], P. O. Mulhair [1 2], C. Inglehearn [1], A. J. Mighell [1], M. J. O'Connell [2]

[1]University of Leeds, Leeds, [2]University of Nottingham, Nottingham, United Kingdom

Abstract: Amelogenesis imperfecta (AI) is a heterogeneous group of conditions characterised by inherited developmental enamel defects. These defects are caused by pathological variants in genes involved in all stages of amelogenesis, which are inherited in a Mendelian pattern. They result in discoloured, weak enamel, great pain and social embarrassment. So far 19 genes have been associated with a non-syndromic AI phenotype, explaining 60% of the AI cases on a molecular level. High throughput technologies have increased our understanding of the genetic basis of AI, however the rate of new candidate genes has reached a plateau. We propose an alternative approach to gene discovery by studying the evolutionary history of 19 known AI implicated genes across 30 toothed, 5 toothless and 3 enamel-less mammal species. We theorise that there is a distinct pattern of selective pressure shared by all genes responsible for amelogenesis, resulting from their coevolution, that has the potential to lead us to associating variants in candidate genes with an AI phenotype. In addition, we attempt to examine whether the change in selective pressure on these genes has lead to loss of function and pseudogenisation in the group of toothless and enamel-less mammals. Some of the sequences in the toothless and enamel-less set of species have internal stop codons, while maintaining an extraordinarily high level of conservation that is more typical of functional coding sequences. Here we present our findings thus far on the identification of novel candidate AI genes and selective pressure variation across toothed, toothless and enamel-less mammals.

The molecular basis of major transitions in evolution

SMBE-PO-721

Loss of homologous pairing in Drosophila interspecies hybrids provides insight into meiosis, speciation, and the origin of somatic pairing

J. Baldwin-Brown [1,*]

[1]Biology, University of Utah, Salt Lake City, United States

Abstract: The pairing of homologous chromosomes is an essential process in all eukaryotes. Although we often associate chromosome pairing with meiosis, pairing can also occur in somatic cells. In humans, somatic pairing is not typically detected in healthy cells, and often occurs in cancer.

Drosophila interspecies hybrids are a powerful system for investigating long-standing and fundamental questions about chromosome pairing. Somatic pairing is the wild type state in flies, but highly reproducible patterns of unpaired regions are observed across the genome in interspecific hybrids between D. melanogaster and its closest sister species, D. simulans. This pattern of pairing loss opens the door to precise, powerful experiments to address unsolved problems in chromosome pairing biology.

I generated a set of within-species and between-species crosses using D. melanogaster and D. simulans, and used Hi-C and Illumina sequencing to measure the rate of chromosome pairing with high resolution across the genome. I correlated the rate of reduced pairing in hybrids with genomic features, and will discuss the patterns identified. This first study reveals much about the nature of homologous chromosome pairing and sets the stage for further experiments that will characterize the mechanism of homologous pairing through functional genetic manipulations of hybrid flies. These experiments provide new insight into the question of how Dipterans such as Drosophila transitioned from obligate somatic non-pairing to obligate somatic pairing in the distant past.

The molecular evolution of cancer

SMBE-PO-733

Modeling the Allele Frequency Spectrum and Cell Population Dynamics of Cancer Clones

H. Chen*

Abstract: The allele frequency spectrum (AFS) is commonly used to summarize genomic polymorphism of cancer cells, and has been extensively applied in studies of tumorigenesis and cancer dynamics in recent years. However, these existing studies assume a simplified exponential growth model, which is unrealistic for most tumor growths. In this study, we present a framework for modeling AFS for arbitrary cancer growth models. We also develop a method for inferring population dynamics using AFS by taking into account the effect of the mixed proportion and sequencing depth etc. We apply the theoretical results to two data sets, one from leukemia and one from hepatoma to demonstrate the performance. Our work provides a foundation for further investigation of cancer cell evolution using genomic polymorphism pattern.

The molecular evolution of cancer

SMBE-PO-744

Controlling transcription through long non-coding RNA interactions between PTEN and its pseudogene

N. C. Lister [1,*], G. Shevchenko [2], J. Walshe [3], J. Groen [1], P. Johnsson [4], L. Vidarsdóttir [5], D. Grander [5], S. Ataide [3], P. Waters [1], K. Morris [2]

[1]BABS, UNSW, Sydney, Australia, [2]Center for Gene Therapy, City of Hope, Duarte, United States, [3]School of Life and Environmental Sciences, University of Sydney, Sydney, Australia, [4]Ludwig Institute for Cancer Research, [5]Department of Oncology-Pathology, Cancer Center Karolinska, Stockholm, Sweden

Abstract: Regulation of gene transcription has been observed through interactions between chromatin and RNA. However, the mechanism by which these interactions occur in humans between long non-coding RNAs (lncRNAs) and chromatin at a site of interest is not well known. We use PTEN and its pseudogene as a model to examine both epigenetic and transcriptional regulation of gene expression. Our findings demonstrate an interaction between the 5' UTR of a transcript, which spans the PTEN promoter, and an antisense lncRNA. This interaction results in the recruiting of DNA methyltransferase 3a (DNMT3a) to the promoter of interest, and is dependent on both sequence and structure of the antisense lncRNA. This example presents the idea that RNA structure may be subjected to evolutionary pressures more so than conservation of sequence. The observations we present provide some insight into a much more exciting and intricate role for the regulation of gene expression by RNA than previously thought.

The molecular evolution of cancer

SMBE-PO-734

PSiTE: a Phylogeny guided Simulator for Tumor Evolution

H. Yang [1], B. Lu [2], W. Zhai [1,*]

[1]Institute of Zoology, Chinese Academy of Sciences, Beijing, China, [2]University of College London, London, United Kingdom

Abstract:

Simulating realistic clonal dynamics of tumors is an important topic in cancer genomics. Here, we present PSiTE (Phylogeny guided Simulator for Tumor Evolution), a tool that can simulate different types of tumor samples including single sector, multi-sector bulk tumor as well as single-cell tumor data under a wide range of evolutionary trajectories. PSiTE provides an efficient tool for under-standing clonal evolution of cancer. PSiTE is implemented in Python and is available at https://github.com/hchyang/PSiTE.

The molecular evolution of cancer

SMBE-PO-737

Quantifying the evolutionary dynamics of human cancers.

T. Graham [1,*]

[1]Barts Cancer Institute, London, United Kingdom

Abstract: The fundamental evolutionary parameters that define cancer evolution, such as the mutation rate per cell division and selective advantage conferred by each mutation, remain poorly characterised. Here I will discuss how these parameters can be derived from routinely-available cancer genome sequencing data, via statistical inference of mathematical population genetics models of clonal evolution. We measure that positively selected mutations can cause fitness increases as large as 50%, and also explore the dynamics of negatively-selected mutations (neoantigens) in a growing tumour. These quantitative measurements of cancer evolution enable mechanistic forecasting of the future evolution of a tumour.

The molecular evolution of cancer
SMBE-PO-736
Subclonal structure in tumors
G. Tibély*, G. Szöllősi, I. Derényi

Abstract:
Subclonal structure in tumors
Gergely Tibély, Gergely Szöllősi, Imre Derényi
Intratumor heterogeneity appears as a consequence of imperfect DNA copying during division of tumor cells, leading to a somatic evolutionary process. Recent studies imply that the number of subclonal mutations is significantly larger than expected, compared to current estimates of the human somatic mutation rate, assuming neutrality of mutations (1, 2). We hypothesize that the elevated number of mutations are caused either by a significantly higher mutation rate, or to prevalent cell death which lengthens lineage lines, resulting in more cell divisions along each lineage. We are developing tools to estimate the likelihood of generative models having different mutation rate-cell death rate parameter values. We also include sequencing errors in the model, which dominate the mutant reads in empirical data. Test results show that the real parameters can indeed be found, the effects of cell death and mutation rate can be differentiated.
(1) Williams M J, Werner B, Barnes C P, Graham T A, Sottoriva A. Identification of neutral tumor evolution across cancer types. Nat Genet 2016; 48:238-244.
(2) Martincorena I, Campbell P J. Somatic mutation in cancer and normal cells. Science 2015; 349:1483-1489.

Intratumor heterogeneity appears as a consequence of imperfect DNA copying during division of tumor cells, leading to a somatic evolutionary process. Recent studies imply that the number of subclonal mutations is significantly larger than expected, compared to current estimates of the human somatic mutation rate, assuming neutrality of mutations (1, 2). We hypothesize that the elevated number of mutations are caused either by a significantly higher mutation rate, or to prevalent cell death which lengthens lineage lines, resulting in more cell divisions along each lineage. We are developing tools to estimate the likelihood of generative models having different mutation rate-cell death rate parameter values. We also include sequencing errors in the model, which dominate the mutant reads in empirical data. Test results show that the real parameters can indeed be found, the effects of cell death and mutation rate can be differentiated.

(1) Williams M J, Werner B, Barnes C P, Graham T A, Sottoriva A. Identification of neutral tumorevolution across cancer types. Nat Genet 2016; 48:238-244.(2) Martincorena I, Campbell P J. Somatic mutation in cancer and normal cells. Science 2015; 349:1483-1489.

The molecular evolution of cancer

SMBE-PO-735

Investigation of EMT-mediated anti-cancer drug resistance of cancer cells by single-cell RNA-seq analysis

Y. Seto [1,*], R. Katayama [1]

[1]Experimental Chemotherapy, Cancer Chemotherapy Center, Japanese Foundation for Cancer Research, Tokyo, Japan

Abstract: Cancer is described as abnormal cell proliferation due to disruption of normal cell signaling pathways. Various protein kinase genes involving in the signaling pathways have been identified as proto-oncogenes. For example, oncogenic mutation in epidermal growth factor receptor (EGFR) which has intracellular tyrosine kinase domain induces constitutive activation of downstream signaling pathway and induces abnormal cell proliferation. In cancer therapy, various kinds of protein kinase inhibitors, called molecular targeted drugs, are used and result in positive therapeutic effect. However, emergence of drug resistant cancer cells during the cancer therapy is a serious clinical problem. Epithelial to mesenchymal transition (EMT) which is morphological transformation from epithelial cells to mesenchymal cells by loss of cellular polarity and by cytoskeletal modification is known as a major mechanism of anti-cancer drug resistance. Additionally, a tumor consists of variety cancer cells having different genetic and phenotypic features, such as mutation pattern, gene expression pattern, growth rate, motility, and this tumor heterogeneity also plays an important role in acquiring anti-cancer drug resistance. Therefore, we investigated response against anti-cancer drug of cancer cells by using single-cell RNA-seq method to reveal molecular mechanisms of EMT and of anti-cancer drug resistance. In this study, we used cultured cell line established from an EGFR mutation-positive lung cancer patient. The cell line shows reversible EMT in the presence or absence of EGFR-tyrosine kinase inhibitor (EGFR-TKI) and has resistance to the EGFR-TKI. As the results, expressions of EMT-related genes, *vim*, *fn1,* and *wnt2b*, were up-regulated by treating EGFR-TKI. Clustering analysis based on gene expression indicated the cells were clustered in 6 groups. Although expressions of genes involving in cell cycle were decreased in all clusters, expression patterns of EMT-related genes were different among the clusters. These results indicated that EMT-state induced by EGFR-TKI was different among the cancer cells and suggested that this EMT-state heterogeneity is important for cancer cells to acquire anti-cancer drug resistant.

The molecular evolution of cancer

SMBE-PO-740

Specific molecular functions linked to favoured amino acids found in cancer mutations

A. Safadi [1,*], S. Lovell [1], A. Doig [1]

[1]university of Manchester, Manchester, United Kingdom

Abstract: Missense mutations can affect protein structure and function, and so can play a pivotal role in cancer initiation and development. Mutations that are reported in literature and associated with cancer progression are diverse, and there are not clear underlying patterns that has been identified. Such pattern (if found) would provide a link between the impact of these amino acids replacements and cancer biological phenotypes compared to other replacements found in other disease for example. Fortunately, large-scale analyses of sequence variation in exomes allows the comparison of normally-occurring variation in human populations with the pattern of mutations and amino-acid replacements observed in cancer, with the aim of identifying large-scale differences in distributions. Here we analyse the variation amino-acid replacements observed in cancer-associated genes. We compare cancer driver mutations from the COSMIC database, and we compare this distribution with several control distributions. In particular we use a null model that combines the probability of amino-acid replacements arising from the genetic code with transition and transversion frequencies. We also compare the results with wider general spectrum of amino acid replacements found in a population. We show how our finding help in understanding the impact of cancer driver mutations and possibly re-evaluating mutations that were previously unidentified as cancer driver mutations using patterns found. The latter could be important in the identification of potential drug targets

The molecular evolution of cancer

SMBE-PO-732

Linking somatic selection of oncogenes and tumor suppressor genes with clinical phenotypes

L. Liu*

Abstract: Oncogenes (OGs) and tumor suppressor genes (TSGs) play opposite roles in carcinogenesis. Pooling these two groups of genes in genome-wide analysis risk obscuring signals that otherwise may have significant biological or clinical implications. However, the lack of knowledge on tumor activating or suppressing status of most cancer-related genes limits our ability to conduct stratified analyses. To address this problem, we developed an evolutionary approach that jointly estimates selection coefficients on missense mutations and on truncating (nonsense and frameshifting) mutations and integrates directional selection with positional mutational distribution to categorize OGs and TSGs, which is the first *in silico* method in the field. We applied this new method to classifying OGs and TSGs in 33 cancer types using 10,172 tumor exomes from The Cancer Genome Atlas. Our classifications are in high concordance (87%) with the expert reviewed cancer gene consensus panel, and included 57 novel OGs and 101 novel TSGs showing signature selection patterns. Given these classifications, we contrasted OGs and TSGs on their associations with clinical phenotypes. Across cancer types, the number of mutated TSGs increased with the age at diagnosis (Pearson correlation=0.68, p-value=10^{-5}). But the number of mutated OGs did not change significantly with age (p-value=0.15). After adjusting for age at diagnosis and tumor stages, mutational burden of TSGs is not significantly associated patient overall survival in any of the 33 cancer types. However, in low-grade glioma, mutational burden of OGs is an independent predictor of favorable prognosis (Cox regression p-value=$3x10^{-4}$, hazard ratio=0.50). We further identified functional domains that were enriched with mutations in OGs but depleted in TSGs. Many of these domains contain known drug targets and shared between common OGs (found in >10% of tumors) and rare OGs. Using melanoma as an example, we showed that these rare OGs are potential candidates of therapeutic targets. The classification results are available in an online database, named genes under selection in tumors (GUST) at https://compumedlab.net/gust/

The molecular evolution of cancer

SMBE-PO-739

Biosynthetic energy cost for amino acids decreases in cancer evolution

H. Zhang [1], Y. Wang [1,*], J. Li [2], H. Chen [2], X. He [3], H. Zhang [4], H. Liang [2], J. Lu [1]

[1]Peking University, Beijing, China, [2]The University of Texas MD Anderson Cancer Center, Houston, United States, [3]Sun Yat-Sen University, Guangzhou, China, [4]The University of Texas Health Science Center at Houston, Houston, United States

Abstract: Cancer development is a multiple-step evolutionary process in which cancer cells acquire a selective advantage in their competition with neighboring cells. Cancer cells must adapt to their microenvironment for rapid proliferation, and metabolic adaptation is the key to this process. Rapidly proliferating cancer cells have a much higher demand for proteinogenic amino acids than normal cells.

The use of amino acids in human proteomes is largely affected by their bioavailability, which is constrained by the biosynthetic energy cost in living organisms. Conceptually distinct from gene-based analyses, we introduce the energy cost per amino acid (ECPA) to quantitatively characterize the use of 20 amino acids during protein synthesis in human cells. By analyzing gene expression data from The Cancer Genome Atlas, we find that cancer cells evolve to utilize amino acids more economically by optimizing gene expression profile and ECPA shows robust prognostic power across many cancer types. We further validate this pattern in experimental evolution of xenograft tumors. Our ECPA analysis reveals a common principle during cancer evolution. (DOI: 10.1038/s41467-018-06461-1)

The molecular evolution of cancer

SMBE-PO-743

Dramatic change in frequencies by DNA barcoding between in vivo cell culture and in vitro mice tumorigenesis: a case study

Z. Luo[*], H. Wen[1], C.-I. Wu[2]

[1]State Key Laboratory of Biocontrol, School of Life Sciences, Sun Yat-sen University, Guangzhou, China, [2]Department of Ecology and Evolution, University of Chicago, Chicago, United States

Abstract: Genomic and phenotypic studies have demonstrated that tumorigenesis is an ultra-micro evolutionary process characterized by its extremely small divergence in DNA sequences and complicated by demography issues (i.e., rapid clonal expansion). Recently, cellular DNA barcoding technique is widely used in increasing site frequency resolution spatiotemporally as a supplement of genome sequencing. Despite its popularity, herein we report an unusual case where lineage frequencies can be increased from relatively low in source library to unexpected high (60-70% sequencing reads due to only three barcode lineages) in mice xenograft tumors dramatically. This result is reconfirmed in three tumors (39 punched hole samples in total) and sequence preference by PCR amplification is less persuadable indicated by qPCR standard curve method. Notably, new experimental setups using same set of barcodes do not find such accordant pattern.

The molecular evolution of cancer

SMBE-PO-742

Evolution of Tasmanian devil transmissible tumour chromosomes

J. Deakin [1,*], E. Ingles [2]

[1]University of Canberra, Canberra, Australia, [2]Institute for Applied Ecology, University of Canberra, Canberra, Australia

Abstract: The Tasmanian devil (*Sarcophilus harrisii*) is one of only two wild mammalian species with a transmissible tumour spreading through the population, where the tumour cells themselves are the infectious agent. The devil has two, independently derived transmssible facial tumours, referred to as DFT1 and DFT2. Transmissible tumours are seemingly rare yet the discovery of two in the one species suggests devils possess attributes that make them prone to the development of these tumours. Identifying similarities between the two transmissible facial tumours may help to understand how these tumours develop and the features these tumours possess that contribute to their transmissibility. We have examined telomere biology and chromosome evolution in both tumours. We have been particularly interested in following the evolution of DFT2, a more recently emerged transmssible tumour which could help to shed light on the evolutionary proccess involved in the development of the highly rearranged DFT1 tumour karyotype.

Understanding the genomics of climate change response
SMBE-PO-747
Towards the genetic prediction of bleaching response in corals

Z. Fuller*, J. Peng [1], P. Andolfatto [2], M. Matz [3], L. Bay [4], M. Przeworski [2 5]
[1]Ecology and Evolutionary Biology, Princeton University, Princeton, [2]Biological Sciences, Columbia University, New York, [3]Integrative Biology, University of Texas, Austin, United States, [4]Marine Science, Australian Institute of Marice Science, Townsville, Australia, [5]Systems Biology, Columbia University, New York, United States

Abstract: Coral reefs are facing a global crisis, as increasing seawater temperatures and other environmental stressors are triggering mass bleaching episodes. Within coral populations, however, there is considerable phenotypic variation in the bleaching response. In the coral *Acropora millepora* on the Great Barrier Reef (GBR), this variability has been shown to be partly heritable, *i.e.,* due to segregating genetic variation. Unfortunately, it is not known which individuals are more likely to bleach and therefore where to prioritize conservation efforts. To address this urgent problem, we propose to perform phenotypic prediction in corals by conducting a genome-wide association study (GWAS) of bleaching response, borrowing on methods widely used in agriculture and human genetics. To this end, we first constructed a highly contiguous *de novo* assembly of the *A. millepora* genome. We further resequenced 48 whole genomes of individuals collected from 12 locations on the GBR at high coverage in order to assess population structure and characterize the demographic history of sampled *A. millepora* populations. We infer little or no population structure across 1000s of kilometers, yet find strong localized signatures of elevated genetic diversity consistent with models of spatially varying selection across a heterogeneous environment. Using a reference haplotype panel constructed from the individuals sequenced at high coverage, we demonstrate that we can reliably perform genotype imputation, indicating that large-scale, low-coverage sequencing should be feasible in this coral species. Finally, we perform low-coverage sequencing for >250 additional individuals, setting the stage for genome-wide mapping studies of bleaching and other phenotypes.

Understanding the genomics of climate change response

SMBE-PO-746

Multi-species comparative transcriptomics reveals shared and divergent response of marine mollusc larvae to climate change stress

E. Harney [1,*], P. Miner [2], P. Boudry [2], S. Auzoux-Bordenave [3], E. Corre [4], F. L. Nunes [5]

[1]Ecology, Evolution and Behaviour, University of Liverpool, Liverpool, United Kingdom, [2]Functional Physiology of Marine Organisms, Ifremer, Plouzané, [3]Biology of Aquatic Organisms and Ecosystems, National Museum of Natural History, Concarneau, [4]Analysis and Bioinformatics for Marine Science, Roscoff Marine Station, Roscoff, [5]Coastal Environment Dynamics, Ifremer, Plouzané, France

Abstract: In the marine environment, changing environmental conditions such as ocean warming and acidification are likely to have variable effects on species with different ecologies and adaptations. Comparison of gene expression between species may help reveal why some are more responsive or resilient to these changing conditions than others. We carried out common garden experiments to investigate the potential effect of temperature stress, pH stress and combined stress at three stages of larval development in four different marine mollusc species. The species that were considered (great scallop, *Pecten maximus*; Manilla clam, *Ruditapes philippinarum*; European Abalone, *Haliotis tuberculata*; and Pacific oyster, *Crassostrea gigas*) represent a broad phylogenetic spread, and include two intertidal and two subtidal species.

Gene expression across the four species, four environments and three developmental stages was assessed by RNA-Seq. Following *de novo* transcriptome assembly, orthologues among the four species were identified, and those orthologues that showed similar patterns of expression across the four species were clustered into modules by weighted gene co-expression network analysis. Orthologue expression within each module was summarised by a module 'eigengene' value for each species, which could be correlated with treatments, phenotypic measures and the tidal ecology of the larvae in a phylogenetically controlled manner.

This functional transcriptomic approach was able to identify conserved responses to temperature and pH among the four species, and also highlighted a number of developmental and metabolic genes that were more responsive to climate stress in the intertidal species. These genes may play a role in the ability of intertidal species to withstand variable conditions and may provide clues about how intertidal and subtidal species will respond to future climate change.

Understanding the genomics of climate change response

SMBE-PO-748

Evolutionary Dynamics of the Antifreeze Protein III Gene Cluster in Polar Fish

J. Kelley [1], L. S. F. Lins [1], S. Hotaling [1], L. Helou [2 3], A.-S. Fiston-Lavier [2,*]

[1]School of Biological Sciences, University of Washington, Pullman, WA, United States, [2]Genome, Institut des Sciences de l'Evolution de Montpellier (UMR 5554, CNRS-UM-IRD-EPHE), Université de Montpellier, Montpellier, [3]School of Biological Sciences, University of Washington, Washington, France

Abstract: To sustain life at subfreezing temperatures in the poles, organisms must survive in constant cold temperatures and long periods of darkness. Many species in polar regions experience temperatures below their serum freezing point and have independently evolved antifreeze proteins (AFPs). AFPs bind to forming ice crystals and inhibit ice recrystallization, conferring an adaptive advantage to the species that have AFPs. The type III AFP gene in the Antarctic fish, *Lycodichthys dearborni,* arose from the translocation and duplication of another gene (*SAS-B*). Tandem duplications of the AFP gene led to a cluster of *AFP III* genes. A transposable element (TE) is located upstream of the *SAS-B* and the *AFP* cluster, suggesting the putative role of this TE in the formation of the AFP gene cluster. We aim to understand how this important protein (AFP) has evolved in some polar species and the role of TEs in this evolutionary process. To answer these questions, we characterized the *AFP* gene cluster region and analyzed the TEs density in *L. dearborni*. We identified a highly abundant TE in the AFP cluster that has not been identified for this species before: the L2-1 element. These elements are located downstream of each annotated repeat unit of the *AFP III* gene with a conserved distance suggesting recurrent duplication events. We then further characterized the *AFP III* genes by examining the relationships among the gene copies and their nearby TE insertions. This study sheds light on the role of TEs in the evolution and duplication of novel proteins.

Understanding the genomics of climate change response

SMBE-PO-749

Consequences of the last glacial period on genetic gradients of current Asians

C. Branco [1,*], M. Arenas [1]

[1]Biochemistry, Genetics and Immunology, University of Vigo, Vigo, Spain

Abstract: An environmental factor that might have deeply influenced genetic patterns of diverse species, including modern humans, is the last ice age. It is known that this period promoted population range contractions, refugia isolation and posterior range expansions. In 1993, Cavalli-Sforza and coauthors observed a genetic gradient of modern humans in Asia presenting an east-west orientation. However, little was explored concerning the environmental and population genetic processes causing this gradient. Here we investigated the influence of the last glacial period, as well as different levels of admixture between Paleolithic and Neolithic populations, on genetic gradients of current Asians. To do so we simulated genetic data under spatially explicit models to mimic the cited evolutionary scenarios and we estimated the corresponding genetic gradients with principal component analyses. Next, we compared simulated and real gradients to identify the best fitting evolutionary scenario. Our preliminary results showed that (*1*) Paleolithic populations can only present the real east-west gradient if the last glacial period is considered and, under this situation, the genetic gradient was probable caused by *allele surfing*; (*2*) Scenarios of admixture between Paleolithic and Neolithic populations presented the real east-west gradient if two origins of Neolithic expansion (Middle East and Southeast Asia) are considered, this constraint was enough to fix the gradient since it did not change with the level of admixture or the presence of the last glacial period. This gradient could be explained by admixture of genetic sectors caused by the Neolithic expansions. Altogether we conclude that the last glacial period promotes the real genetic gradient observed in Asia although other factors (especially the two Neolithic expansions) could also favor this gradient.

Understanding the genomics of climate change response

SMBE-PO-751

Genetic adaptation and vulnerability to climate change of an Alpine tree species with a long generation time

B. Dauphin*, C. Rellstab, S. Zoller, D. Karger, S. Brodbeck, F. Gugerli

Abstract: Current temperature increase in the European Alps is more pronounced at higher than at lower elevations. Recent evidence indicates that the Alpine flora responds to climate warming through both accelerated elevational range shifts and increased plant community height. However, it remains unknown to what degree Alpine plants, especially those with a long generation time, may adapt at the genetic level to follow the rapid pace of climate change. We combined present-day genomic variation in different age cohorts of Swiss stone pine (*Pinus cembra*), a keystone species at the timberline, with interpolated and modelled climate data across the last two centuries to identify historical shifts in allele frequencies at genomic regions putatively associated with climatic variables. Contemporary juvenile cohorts in the core of today's elevational range, which experience different temperature conditions than those under which their progenitors established, showed higher shifts in allele frequency at temperature-associated compared to neutral loci. Juvenile cohorts at the upper colonization front mostly exhibited adaptation patterns similar to those of adult cohorts of the core elevational range for temperature-associated loci, indicating that beneficial alleles for colder temperature moved uphill during colonization. However, in some putatively adaptive loci, high-elevation juvenile cohorts already showed signs of selection for beneficial alleles for warmer conditions, suggesting adaptation to recent climate change. Nevertheless, when projecting allele frequency changes into the future, the genomic vulnerability of low-elevation juvenile cohorts to future climatic changes appears high under the current climate trend (RCP4.5), and is even doubled under a more extreme scenario (RCP8.5).

Understanding the genomics of climate change response

SMBE-PO-752

A kea's tale: Evolution and survival of the world's only alpine parrot in a warming world

D. Martini [1], H. Cross [1], N. Dussex [2], B. C. Robertson [1], N. J. Gemmell [1], M. Knapp [1,*]

[1]University of Otago, Dunedin, New Zealand, [2]Swedish Museum of Natural History, Stockholm, Sweden

Abstract: The kea is an endangered parrot species endemic to New Zealand and the only parrot in the world which habitually utilises the alpine environment. The use of alpine habitats sets it apart from the kaka, a forest adapted sister species with which the kea shares a common ancestor within the last 2 million years. Differences in habitat preference between the species raise the question whether the kea has developed any functional genomic adaptations to the alpine environment, and, if so, how these may affect its ability to survive in a warming climate with a shrinking alpine zone in New Zealand.

Using whole genome analyses of kea and kaka, we show that both species share similar functional genomic traits characteristic for alpine adaptation, suggesting that ecological differences are more likely underpinned by different behavioural adaptations. However, population dynamic analyses based on whole genome data show that kaka population size increased significantly with climate warming during interglacials, while kea populations stagnated or even declined. This is likely a result of a strong increase in forest habitat and a reduction in alpine habitats in New Zealand during warm times. A further reduction of the alpine zone associated with climate warming will therefore likely further increase pressure on kea populations and may eventually lead to a complete replacement by its sister species.

Using Ancient DNA to Study Natural Selction: New Models and New Data
SMBE-PO-760

The impact of modern human specific sites on human phenotypes

C. R. Robles [1,*], S. Sankararaman [2]
[1]Human Genetics, [2]Computer Science, UCLA, Los Angeles, United States

Abstract: Uncovering genetic changes that make anatomically modern humans "unique" is critical for a comprehensive understanding of human evolution. By analyzing genome sequences from our closest evolutionary relatives, Neanderthals and Denisovans, we can functionally characterize these genetic changes. Towards this end, we identified 321,821 mutations that are nearly fixed for the derived allele in African individuals from the 1000 Genomes project (>90% derived allele frequency) but are absent in at least some of the deeply sequenced Altai Neanderthal or Denisovan genomes. To understand the phenotypic impact of these fixed derived mutations (FDMs), we leverage the observation that interbreeding with Neanderthals likely re-introduced the ancestral allele at a number of these sites. We find that a number of FDMs are polymorphic in European populations so that their phenotypic impact can be analyzed by genotyping these mutations in large cohorts with phenotypic information.
We analyzed ~114,000 FDMs across 107 phenotypes that include anthropometric, blood, bone density, and disease related traits measured in ~475,000 individuals of European ancestry in the UK Biobank dataset. We discovered 3266 independent associations of FDMs in all 107 of our phenotypes with p-values that pass a threshold accounting for both FDM number and phenotypes tested. Controlling for frequencies and linkage disequilibrium patterns at these variants, we find the contribution of FDMs to phenotypic variation is significantly depleted for several traits including those related to body mass and bone density. These results direct us towards genetic and phenotypic changes important for contemporary human biology since diverging from archaic humans.

Using Ancient DNA to Study Natural Selction: New Models and New Data

SMBE-PO-755

Ancient DNA sheds light on the immunogenetic history of Europe

R. Barquera [1,*], M. Rivollat [2], L. Papac [1], M. Spyrou [1], S. Schiffels [1], C. Jeong [1], W. Haak [1], J. Krause [1]

[1]Department of Archaeogenetics, Max Planck Institute for the Science of Human History, Jena, Germany, [2]De la Préhistoire à l'Actuel, Culture, Environnement, Anthropologie, Université de Bordeaux, Bordeaux, France

Abstract: The geographic distribution of allelic variants of the Human Leukocyte Antigen (HLA) immunogenetic system has been extensively studied for the past 50 years. Several alleles exhibit geographic restriction which has led to several hypothesis in an attempt to explain their distribution. Models such as isolation by distance, neutral evolution, pathogen-driven selection and natural selection scenarios have failed to thoroughly explain the patterns found so far. Here we use a novel approach by combining a genome-wide population genetics approach with immunogenetics from ancient populations to explore past diversity and compare it with the HLA diversity exhibited in central Europe through time, going from the Neolithic period to modern populations in an attempt to reconstruct the chronological distribution of HLA alleles. Neolithic populations throughout Central Europe exhibit the same characteristic distribution of HLA alleles in similar frequencies, with alleles such as HLA-B*44 being present throughout the entire time transect. However, other alleles and haplotypes vary in time, which may be related to events such as pandemic episodes, long exposure to pathogens or changes associated with Neolithization or other processes. For the first time, we are able to show that not only geographic distribution, but also chronologic variation, are evidence of adaptive events in human history and provide a dynamic window to the interplay that pathogens and human hosts have played for the past 10 000 years.

Using Ancient DNA to Study Natural Selction: New Models and New Data
SMBE-PO-759
Investigating sheep domestication using ancient DNA
A. J. Hare [1,*]
[1]Trinity College Dublin, Dublin , Ireland

Abstract: Sheep are an important domestic species, being farmed world wide, and being one the first to be involved in agriculture during the Neolithic transition. This makes them an ideal candidate for studies that seek to understand the mechanics of animal domestication, particularly using ancient DNA.

Using data generated from ancient sheep, we have examined the initial changes associated with the transition from mouflon, such as reduction in genetic variation pre and post domestication event. These data are from a broad geographic region, spanning from Iran to Ireland and from pre-Neolithic to medieval times. These are then compared against whole genome data from a wide and varied range of modern domestic sheep, as well as modern mouflon, to contextualize the relative variation in each population and compare them.

Further genetics changes are investigated, particularly those associated with the spread of farming and animals human populations transported with them, across Eurasia. This movement out of Western Asia caused related bottlenecks across the contient. Finally potential admixture between differentiated populations and sub-species of sheep is also investigated, using D-stats from a large population of low coverage individuals.

Using Ancient DNA to Study Natural Selction: New Models and New Data
SMBE-PO-764

The population genetics of prehistoric Portugal.

E. Breslin [1,*], L. M. Cassidy [1], V. Mattiangeli [1], R. Martiniano [2], A. M. Silva [3 4], D. G. Bradley [1]
[1]Department of Genetics, Trinity College Dublin, Dublin, Ireland, [2]Department of Genetics, University of Cambridge, Cambridge, United Kingdom, [3]Laboratory of Prehistory, CIAS, Department of Life Sciences, University of Coimbra, [4]CEF, University of Coimbra and UNIARQ (University of Lisbon), Coimbra, Portugal

Abstract: The field of ancient population genetics has benefited greatly from the development of high-throughput next-generation sequencing (NGS) and the discovery that the petrous part of the temporal bone is a rich reservoir for aDNA, allowing the generation of whole genome sequences for ancient individuals. Portugal occupies a unique position in Europe; located on the south-western extreme of mainland Europe and facing both the Atlantic and the Mediterranean, it was connected to two major maritime trade and migration routes as well as experiencing influx from central Europe throughout its prehistory. However, many open questions remain about demographic and selection processes acting on populations at key transition points in European prehistory, such as the early Bronze Age migrations from the Pontic steppe, the potential source for the R1b Y-chromosome haplogroup which now dominates in European populations. In this study we present whole genome sequences from ancient Portuguese individuals, covering a period of over 3000 years as well as a wide geographic region. We observe changes in both mitochondrial and Y-chromosome haplogroup frequencies over time, reflecting changing demographic processes acting on Iberian populations.

Using Ancient DNA to Study Natural Selction: New Models and New Data
SMBE-PO-763

Accuracy of haplotype estimation and demographic inference of ancient DNA

J. Medina-Tretmanis [1,*], M. C. Ávila-Arcos [2], E. Huerta-Sánchez [3]
[1]Faculty of Sciences, National Autonomous University of Mexico, CDMX, [2]International Laboratory for Human Genome Research, National Autonomous University of Mexico, Querétaro, Mexico, [3]Ecology and Evolutionary Biology and Center for Computational Molecular Biology, Brown University, Providence, United States

Abstract: The increasing availability of ancient human genomes has produced a wealth of information regarding past demographic, admixture, and adaptive events. For inferring evolutionary events, haplotype-based methods exploit more information than frequency-based methods, but this requires accurate haplotype estimation (i.e. “phasing”) which is particularly challenging with ancient DNA (aDNA). Accuracy is affected by modern contamination, low coverage, damage, sample age and genetic distance from the modern haplotype reference panels. To test the effect of these variables on phasing accuracy we simulated ancient genomic data and applied phasing methods that have been previously used for aDNA.

We simulated several combinations of depth of coverage, modern contamination, sample age, and different demographic events, such as population splits and bottlenecks. We further contrasted the accuracy of phasing with and without a haplotype reference panel. As expected, we found that reference panel phasing is preferred over population phasing. We observed a general trend of increasing switch error with increasing contamination and sample age, and with decreasing depth of coverage.

Lastly, we used the phased simulated data to extract IBD tracts, to try to reconstruct simulated demographic events. We identified several scenarios for which phasing of aDNA yields unreliable IBD information and consequently unreliable demographic inferences. We thus provide valuable guidelines for carrying out aDNA phasing under specific values of age, contamination and depth of coverage.

Using Ancient DNA to Study Natural Selction: New Models and New Data
SMBE-PO-758
The Demographic History of the Woolly Rhinoceros
E. Lord [1 2,*], N. Dussex [1], L. Dalén [1]
[1]Department of Bioinformatics and Genetics, Swedish Museum of Natural History, [2]Department of Zoology, Stockholm University, Stockholm, Sweden

Abstract: The Pleistocene was marked by extreme fluctuations in climate and thus provides the ideal conditions to test questions regarding the relationship between climate change, species evolution, and population dynamics. These climatic changes are suggested to have resulted in increased rates of faunal turnover and adaptation, particularly in the Arctic. For instance, recent studies using whole genome data in the woolly mammoth have shown a loss of genetic diversity associated with warming climate and habitat change at the end of the Pleistocene, followed by an increase inbreeding resulting from long-term small population size. Yet, the impact of climate change on genomic diversity in other extinct megafauna has been little explored so far.
The woolly rhinoceros (*Coelodonta antiquitatis*) is another cold-adapted species that was found across Eurasia, and went extinct approximately 14ky BP, coinciding with the abrupt warming of the Bølling-Allerød interstadial. Here we investigate the demographic trajectory of this species during the late Pleistocene using whole genome data from a woolly rhinoceros dating to 18500 Cal BP and mitochondrial DNA from specimens ranging from approximately 50,000-14000 Cal BP. We show that unlike the mammoth, the woolly rhinoceros maintained a relatively constant population size and exhibit higher genetic diversity near to their extinction.

Using Ancient DNA to Study Natural Selction: New Models and New Data

SMBE-PO-757

Phenotype prediction from ancient low-coverage genomes

E. D'Atanasio [1,*], C. L. Scheib [2], R. Hui [3], J. E. Robb [3], C. Cessford [3 4], S. A. Inskip [3], L. Saag [2 5], K. Tambets [2], M. Metspalu [2], T. Kivisild [1]

[1]Department of Human Genetics, KU Leuven, Leuven, Belgium, [2]Estonian Biocentre, University of Tartu, Tartu, Estonia, [3]Department of Archaeology, University of Cambridge, [4]Cambridge Archaeological Unit, Cambridge, United Kingdom, [5]Department of Evolutionary Biology, University of Tartu, Tartu, Estonia

Abstract: Over the past decade, the technological improvements in the shotgun sequencing allowed the analyses of an increasing number of ancient DNA (aDNA) sequences, which can be highly informative about past human evolution and adaptation. However, because of *post-mortem* degradation, low endogenous DNA content and contamination, the average coverage of ancient sample sequences is often not higher than 1×, making the interpretation of the aDNA data challenging. The most used method to overcome these problems is the imputation of missing genotypes comparing the aDNA sequences with a modern reference panel. However, this approach is usually considered reliable for samples with a coverage of at least 1×. We performed several imputation accuracy tests, in order to obtain reliable results for hundreds of ancient samples with an average coverage of about 0.15x. Our sample was composed of two groups of individuals, from Estonia and England respectively, dating back to different cultural and time periods, ranging from Neolithic to Middle Age. We used the local imputation approach to analyze a set of phenotype informative markers, involved in visible features (such as eye, hair and skin colour) and in adaptation to external agents (such as diet and pathogens). We used publicly available modern samples as reference panel, after filtering it to discard the rare variants with a minor allele frequency lower than 0.01. The observed differences in frequencies between time periods and geographic areas allowed us to shed light on past evolutionary changes linked to cultural shifts or pandemics.

Using Ancient DNA to Study Natural Selction: New Models and New Data

SMBE-PO-785

Characterizing ancient genomes embedded into modern human populations

L. Pagani [1 2,*], F. Montinaro [1], D. Marnetto [1], B. Yelmen [1 3], M. Mondal [1], L. Molinaro [1 3], K. Pärna [1 3 4], T. Kivisild [1 5], T. Haller [6], R. Mägi [6], K. Fischer [6 7], M. Metspalu [1]

[1]Estonian Biocentre, Institute of Genomics, University of Tartu, Tartu, Estonia, [2]Department of Biology, University of Padova, Padova, Italy, [3]Institute of Molecular and Cell Biology, University of Tartu, Tartu, Estonia, [4]Department of Epidemiology, University of Groningen, Groningen, Netherlands, [5]Department of Human Genetics, KU Leuven, Leuven, Belgium, [6]Estonian Genome Centre, Institute of Genomics, University of Tartu, [7]Institute of Mathematics and Statistics, University of Tartu, Tartu, Estonia

Abstract: A growing body of ancient DNA evidence is being used to build increasingly more realistic models of human demographic changes in the last few thousand years. However, availability of aDNA of good quality and diversity is often the limiting factor, especially in areas where preservation of archaeological remains is poor. Here we propose to consider modern genomes as being arranged together from pieces of a jigsaw of ancient haplotypes that recombined and admixed in the last few thousand years.

Following what has already been attempted for recently admixed populations, one can use local ancestry methods to extract these “ancient” genomic regions and study them separately. The benefit of this approach stems from our ability to make use of existing high-quality whole genomes, which can be deconvoluted to identify the genetic makeup of the ancient populations that admixed to form contemporary human groups.

We applied the proposed strategy to contemporary South Asian human genomes and managed to retrieve, for the first time, a plausible representation of the South Asian genetic landscape prior to the arrival of West Eurasian populations in the area. We also analyzed North-West African genomes to further characterize the West Asian origin of the non-African genetic component introduced to East Africa in the Iron Age (1st Millenium BCE).

Additionally, we show that the West Eurasian genomic fragments obtained through ancestry deconvolution are sufficient to predict the corresponding West Eurasian full genome Polygenic Risk Score for complex traits such as Type 2 Diabetes in contemporary Egyptians. Here we show that such a prediction is useful to decompose an individual’s genetic risk score into its various ancestry contributions, which has implications for our understanding of the modern and ancient genomic architecture of complex traits, as well as for unlocking the promising field of predictive medicine to individuals of mixed ancestry.

Using Ancient DNA to Study Natural Selction: New Models and New Data

SMBE-PO-784

First ancient DNA evidences of prehistoric human migration and gene flow from Africa into the Iberian Peninsula

G. Gonzalez-Fortes [1,*], F. Tassi [1], E. Trucchi [1], A. Grandal D´Anglade [2], A. Bettencourt [3], C. Barroso [4], A. Manica [5], M. Hofreiter [6], G. Barbujani [1]

[1]Department of Life Science and Biotechnology, University of Ferrara, Ferrara, Italy, [2]Instituto Universitario de Geología, University of A Coruña, A Coruña, Spain, [3]Department of History and Archaeology, University of Miño, Braga, Portugal, [4]Fundación Instituto de Prehistoria, Málaga, Spain, [5]Department of Zoology, University of Cambridge, Cambridge, United Kingdom, [6]Department of Mathematics and Natural Sciences, University of Potsdam, Potsdam, Germany

Abstract: Being at the Western fringe of Europe, Iberia had a peculiar prehistory and a complex pattern of Neolithization. A few studies, all based on modern populations, reported the presence of DNA of likely African origin in this region, generally concluding it was the result of recent gene flow, probably during the Islamic period. Here we provide evidence of much older gene flow from Africa to Iberia by sequencing whole genomes from four human remains from Northern Portugal and Southern Spain dated around 4,000 years BP (from the Middle Neolithic to the Bronze Age). We found one of them to carry an unequivocal Sub-Saharan mitogenome of most likely West or West-Central African origin, never reported before in prehistoric remains outside Africa. Our analyses of ancient nuclear genomes show small but significant levels of Sub-Saharan African affinity in several ancient Iberian samples, which indicates that what we detected was not an occasional individual phenomenon, but an admixture event recognizable at the population level. We interpret this result as evidence of an early migration process from Africa into the Iberian Peninsula through a Western route, possibly across the Strait of Gibraltar.

Using Ancient DNA to Study Natural Selction: New Models and New Data
SMBE-PO-761
Ancient genomic analysis reveals the Yamnaya-related admixture in Iron Age Tianshan, Xinjiang
C. Cui*

Abstract: The recent genetic studies have implied a complex population history in Xinjiang, northwest China with population migrations and admixtures from both East and West Eurasia. However, it remains unclear about the source populations, and when and how frequently the Westerners meet the Easterners in Xinjiang due to the lack of ancient genomic data. Here we reported the genome-wide data of 11 ancient samples from Shirenzi site about 2000 years ago. We observed the Shirenzi people were all admixed deriving ancestry from both the East and West Eurasians. The majority part of the East Eurasian ancestry in Shirenzi was northern Asian related. While the West Eurasian ancestry was most-likely Yamnaya-related ranging from 20% to 80%, which support the Steppe hypothesis of early peopling Xinjiang that this ancestry once expanded further south from Altai region into the northern slope of Tianshan Mountains in Xinjiang.

Using Ancient DNA to Study Natural Selction: New Models and New Data
SMBE-PO-781

The ancient biogeographical history of the Roma

M. M. Johansen [1,*], E. Elhaik [1]
[1]Department of Animal and Plant Sciences, University of Sheffield, Sheffield, United Kingdom

Abstract: The Roma are one of the largest minority group in Europe, with 6-16 million people comprising a mosaic of socially and culturally divergent groups with different languages and religions. Without a written history of their own, the Roma have been subject to investigations that pinpointed them to India based on genetic and linguistic similarities. Four studies have corroborated a Northwest Indian origin by comparing the Roma to modern individuals from South Asia, the Middle East and the Caucasus region using genome-wide autosomal data. To identify the origin and reconstruct the demographic history of the Roma, we employed the Geographic Population Structure (GPS) tool. GPS is an accurate biogeographical tool that converts genetic information into geographic coordinates. We also employed ancient DNA data for the first time to study their ancient origins. GPS localised all but Welsh Roma to Iran. The Roma also exhibited the highest genetic similarity to Iranian, Caucasian and Lebanese Populations highlighting their Middle Eastern origins, at odds with previous reports. Analysis of their modern and ancient admixture components revealed that the Roma share high genetic similarity with Central European and Turkish Neolithic population, and only marginal similarity with Indian populations. Our results challenge the theory of Indian origins for the Roma and suggest at a more complex population history for the Roma.

Using Ancient DNA to Study Natural Selction: New Models and New Data
SMBE-PO-756

Detecting selection from genomic time series: the Beta with spikes approximation.

C. Paris [1], B. Servin [1], S. Boitard [1,*]
[1]GenPhySE, National Institute of Agricultural Research, Castant Tolosan, France

Abstract: Detecting genomic regions under selection is one of the main issues in population genetics. While most methods exploit different kinds of patterns observed in present time data, recent DNA sequencing techniques allow to get more and more time series genomic data. A common modeling approach of these data is to describe the temporal evolution of an allele frequency as a Markov chain. Based on this principle, several methods have been proposed to infer selection intensity for a given polymorphism. One of the main differences between these methods lies in how they model the Markov chain's transition probabilities. Indeed, although the Wright-Fisher model is a natural choice, its computational cost is prohibitive for large population sizes. To overcome this limitation, other models consider approximations to the Wright-Fisher model, based on continuous transition densities. Using simulations, we compared the performance of several of these approximations with respect to their power to detect selection and estimation of the selection intensity. To this aim, we developped a new generic Hidden Markov Model likelihood calculator and applied it on simulations of various scenarii in terms of population sizes, selection intensities and data collection times. We show that the Beta with spikes approximation, which includes fixation probabilities in continuous Beta transitions, provides a very good approximation to the Wright-Fisher process for a computational cost that does not increase with population size. For long sampling time intervals such as those expected from ancient DNA studies, detection power and estimation accuracy are significantly higher with this distribution than with the classical Gaussian or Beta distributions.

Using Ancient DNA to Study Natural Selction: New Models and New Data

SMBE-PO-767

New models to uncover spatiotemporal patterns of adaptation using ancient and present-day DNA

F. Racimo [1,*]

[1]University of Copenhagen, Copenhagen, ,

Abstract: Evolutionary genomics has – in the last two decades – unearthed a rich history of population dynamics while studying species across the planet, including complex patterns of divergence, migration and admixture among differentiated groups. Yet genome-wide studies of selection often assume simple dynamics (e.g. a 3-population tree) or aim to control for complex dynamics without explicitly modeling them (e.g. using the genome-wide covariance matrix). This prevents these rich historical insights from bearing on our understanding of past adaptive events in the organisms we study. Here, I will introduce several new methods we have developed to explicitly account for complex population dynamics while looking for loci with footprints of positive selection. These include programs that can use admixture graphs and latent mixed-membership models to pinpoint where and when in the history of a species a particular selective event took place, while explicitly modeling migration and admixture processes. I will also discuss applications of these methods to new present-day and ancient genomic datasets, including hundreds of ancient human genomes sampled throughout Eurasia, as well as bovine and fish population datasets. Finally, I will talk about on-going research on modeling selection as a spatiotemporal dynamic process – borrowing insights from environmental and paleoclimatic research to uncover the drivers of adaptation (e.g. temperature, vegetation or pathogens), while accounting for the fact that these drivers may have also changed over time and space.

Using Ancient DNA to Study Natural Selction: New Models and New Data

SMBE-PO-765

The rise and fall of the Beringian megafauna: lessons from the Canadian Klondike

J. Oppenheimer [1,*], M. Stiller [2], A. Vershinina [2], M. Mahony [3], T. Sadoway [4], T. Porter [3 5], P. Heintzman [6], M. Wooller [7 8], G. Zazula [9], H. Poinar [10], D. Froese [3], B. Shapiro [2]

[1]Department of Biomolecular Engineering, [2]Department of Ecology and Evolutionary Biology, University of California, Santa Cruz, Santa Cruz, California, United States, [3]Department of Earth and Atmospheric Sciences, University of Alberta, Edmonton, [4]Department of Anthropology, McMaster University, Hamilton, [5]Department of Geography, University of Toronto Mississauga, Mississauga, Ontario, Canada, [6]Department of Ecology and Evolutionary Biology, University of California, Santa Cruz, Santa Cruz, [7]School of Fisheries and Ocean Sciences, [8]Water and Environmental Research Center , University of Alaska, Fairbanks, Fairbanks, Alaska, United States, [9]Palaeontology Program, Government of Yukon, Whitehorse, Yukon, [10]Department of Anthropology, McMaster University, Hamilton, Ontario, Canada

Abstract: Ancient DNA previously retrieved from megafaunal remains throughout Beringia (Siberia through the unglaciated Yukon) and North America has given evidence for drastic population declines of megafauna and has suggested that these declines were driven primarily by environmental change over continental scales. However, such reconstructions have been limited by the study of relatively small numbers of individuals over vast areas, as well as assumptions of panmixia, despite the existence of strong barriers that would have limited dispersal, such as rising sea levels, ice sheets, and ecological factors. In this study, we focus on horse and bison population dynamics in the central Yukon over the last 50,000 years. By coupling ancient DNA, AMS dating, and carbon and nitrogen isotopic analysis in the context of local paleoclimate and paleoenvironmental reconstructions, it is possible to disentangle the effect of local environmental changes on megafaunal populations. We find that during the late MIS 3 interstadial, when open-shrub boreal forest was present, bison were prominent while horses were limited. However, with the development of steppe-tundra toward the last glacial maximum, bison declined and horses increased in abundance. As shrubs expanded in the late Pleistocene, bison increased in abundance, while horses disappeared rapidly by ~13,000 years ago. Taken together, these data demonstrate the large influence of local environment on late Quaternary large mammal population dynamics.

Using Ancient DNA to Study Natural Selction: New Models and New Data

SMBE-PO-778

The ancient genetics of early Neolithic populations in southern China

H. Shafiey [1,*], M. A. Yang [1], Q. Fu [1]

[1]Key Laboratory of Vertebrate Evolution and Human Origins of Chinese Academy of Sciences, Institute of Vertebrate Paleontology and Paleoanthropology, Beijing, China

Abstract:

There is rich literature on population dynamics in Southeast Asia and Oceania and in particular the origin of linguistic div ersity in the region [see REF below]. The prevalent model emphasizes an influx of migrants from southern China into Sou theast Asia around 4000 years ago. This migration is believed to be associated with spreading of Austroasiatic languages in Southeast Asia. Despite several genetic studies focused on Southeast Asia, little is known about ancient populations fr om East Asia, especially southern China. In order to fill this gap, we sequence and analyze genome-wide data from ancient individuals from southern China that date back to the Early Neolithic (around 10000-8000 years ago). Using these data as well as published ancient and present-day human genetic data, we assess the relationship among present-day and ancient populations from Southeast Asia and ancient individuals from southern China.

REF:

H. McColl et.al., Science 361, 88-92 (2018)

M. Lipson et.al., Science 361, 92-95 (2018)

M. Lipson et.al.,Current Biology 28, 1-9 (2018)

P. Skoglund et.al., Nature 538, 510-513 (2016)

CFW. Higham, Farming, social change, and state formation in Southeast Asia, in Oxford Handbook of Zooarchaeology, 20 17.

Using Ancient DNA to Study Natural Selction: New Models and New Data

SMBE-PO-774

Ancient Mitogenomics and the Evolutionary History and Biogeography of Extinct Sloths

F. Delsuc [1,*], M. Kuch [2], G. C. Gibb [3], E. Karpinski [2], D. Hackenberger [2], P. Szpak [4], J. G. Martínez [5], J. I. Mead [6], H. G. McDonald [7], R. D. E. Macphee [8], G. Billet [9], L. Hautier [10], H. N. Poinar [2]

[1]Institut des Sciences de l'Evolution, CNRS - Université de Montpellier, Montpellier, France, [2]McMaster Ancient DNA Centre, McMaster University, Hamilton, Canada, [3]School of Agriculture and Environment, Massey University, Palmerston North, New Zealand, [4]Department of Anthropology, Trent University, Peterborough, Canada, [5]Instituto Superior de Estudios Sociales, CONICET/Instituto de Arqueología y Museo, Universidad Nacional de Tucumán, San Miguel de Tucumán, Argentina, [6]The Mammoth Site, East Tennessee State University Natural History Museum, Johnson City, [7]Utah State Office, Bureau of Land Management, Salt Lake City, [8]Division of Vertebrate Zoology/Mammalogy, American Museum of Natural History, New York, United States, [9]Centre de Recherche en Paléontologie, Muséum National d'Histoire Naturelle, Paris, [10]Institut des Sciences de l'Evolution, CNRS - Université de Montpellier, Montpellier, France

Abstract: Living sloths represent two distinct lineages of small-sized mammals that independently evolved arboreality from terrestrial ancestors. The six extant species are the survivors of an evolutionary radiation marked by the extinction of large terrestrial forms at the end of the Quaternary. Until now the tale of sloth evolution has mainly been told from the morphological point of view. Here we used ancient DNA methods to successfully assemble and sequence 10 extinct sloth mitogenomes encompassing all major lineages. This includes the iconic continental ground sloths *Megatherium*, *Megalonyx*, *Mylodon*, and *Nothrotheriops*, plus the smaller endemic Caribbean sloths *Parocnus* and *Acratocnus*. Phylogenetic analyses identify eight lineages of sloths grouped in three main clades and whose interrelationships are markedly incongruent with the currently accepted morphological picture. We show that extinct Caribbean sloths have a single origin but comprise two divergent lineages that are totally unrelated to living two-fingered sloths. Moreover, living three-fingered sloths do not represent the sister-group to all other sloth species but are nested within a clade of extinct ground sloths. Molecular dating also reveals that the eight newly recognized sloth families all originated between 36 and 28 million years ago (Mya). The early divergence of recently extinct Caribbean sloths around 35 Mya provides support for the debated GAARlandia biogeographic hypothesis postulating the existence at that time of a land corridor joining South America and the Greater Antilles. This new molecular phylogeny has major implications for reinterpreting sloth morphological evolution, biogeography, and diversification history.

Using Ancient DNA to Study Natural Selction: New Models and New Data
SMBE-PO-776

The contribution of Neandertal admixture to modern human disease phenotypes

M. Dannemann [1,*], J. Kelso [1]
[1]Max Planck Institute For Evolutionary Anthropology, Leipzig, Germany

Abstract: The admixture of modern humans with Neandertals ~55,000 years ago has resulted in ~2% of the genomes of present-day non-Africans still being composed of Neandertal DNA. And while negative selection has played an active role in removing parts of the introgressed DNA, still ~40% of the Neandertal genome can be found in present-day non-Africans, some of which has been shown to be adaptive. Association studies have shown that Neandertal DNA significantly influences gene expression and various disease and non-disease phenotypes, including immunity, skin and hair color and behavioral traits. However, assessment of the set of disease phenotypes to which Neandertal variants contribute has been limited due to the low prevalence of many of the diseases in the cohorts.
We have used the full release of the UK biobank and revisited the impact of Neandertal DNA on disease phenotypes in 360,000 individuals. We found several strong associations between introgressed variants and multiple diseases, including iron metabolism, bone disorders and major depression. Additionally, to quantify whether Neandertal variants affect specific disease phenotypes, we estimated the contribution of Neandertal DNA to specific groups of diseases. While some of our results replicate previous findings, we provide several new associations. Our results further complete our picture of how admixture with Neandertals contributes to disease phenotypes in present-day humans.

Using Ancient DNA to Study Natural Selction: New Models and New Data
SMBE-PO-769
Unravelling the evolutionary history of the vaccinia virus, the vaccine for smallpox

A. F. Porter [1,*], A. T. Duggan [2], J. Klunk [3], E. C. Holmes [1 4 5 6], H. Poinar [2 3 7 8 9 10], A. N. Dhody [11 12], R. Hicks [11 12], G. Smith [13], M. Humpherys [14], A. McCollum [15], W. Davidson [15], K. Wilkins [15], Y. Li [15], A. Burke [12], H. Polasky [12], L. Flanders [12], D. Poinar [2], A. R. Raphenya [7 9 16], B. Alcock [7 9 16], T. T. Lau [7 9 16], A. McArthur [7 9 16], B. Golding [3]
[1]School of Life and Environmental Sciences, University of Sydney, Sydney, Australia, [2]Department of Anthropology, [3]Department of Biology, McMaster University, Hamilton, Canada, [4]Marie Bashir Institute for Infectious Diseases and Biosecurity, [5]Charles Perkins Centre, [6]Sydney Medical School, University of Sydney, Sydney, Australia, [7]Department of Biochemistry and Biomedical Sciences, McMaster University, Hamilton, Canada, [8]M.G. DeGroote Institute for Infectious Disease Research, McMaster University, Hamilton, United States, [9]DeGroote School of Medicine, McMaster University, Hamilton, [10]Humans & the Microbiome Program, Canadian Institute for Advances Research, Toronto, Canada, [11]Mütter Research Institute, [12]Mütter Museum of The College of Physicians of Philadelphia, Philadelphia, United States, [13]Department of Pathology, University of Cambridge, Cambridge, United Kingdom, [14]Department of History, Duke University, Durham, [15]Division of High-Consequence Pathogens and Pathology, U.S. Centers for Disease Control and Prevention, Atlanta, United States, [16]M.G. DeGroote Institute for Infectious Disease Research, McMaster University, Hamilton, Canada

Abstract: For over two centuries, vaccinia virus (VACV), has been used for vaccination against variola virus, the causative agent of smallpox. Smallpox was one of the most devastating diseases in human history, and VACV played an essential role in its eradication program. During the 20th century, smallpox vaccination relied on mass-produced VACV, however, pre-20th century, the source of vaccines was unregulated and often unknown. Dr. Edward Jenner initiated the process of vaccination through his observation that milkmaids exposed to cowpox were immune to smallpox. However, modern cowpox is phylogenetically distinct from VACV. As VACV can infect many hosts, and we are unsure of its natural host, its origin remains a mystery. Additionally, as VACV has been interacting with the human host for centuries, it is likely we have influenced its evolution. This, along with the convoluted history of VACV and other orthopoxviruses, makes it difficult to define host-virus relationships, origins and evolutionary history.
Accordingly, in this study we aimed to use phylogenetic inference to estimate the evolutionary history of orthopoxviruses, using ancient DNA sourced from historical samples, which provide a greater time-scale to estimate evolutionary patterns. We isolated near-complete virus genomes from historical samples, including American Civil War era medical artefacts, which are closely related to “horsepox” virus, and lay within the modern VACV clade. To investigate further, we have begun work on 20th century Australian smallpox vaccination kits. By reconstructing and analyzing orthopoxvirus genomes used throughout smallpox vaccination history, we can understand the evolutionary history of VACV and other orthopoxviruses.

Using Ancient DNA to Study Natural Selction: New Models and New Data

SMBE-PO-771

The Population History and Cultural Dispersal Pattern of Hanging Coffins: A Matrilineal Genetic Perspective both in Southern China and Northern Thailand

X. M. Zhang [1,*], C. M. Li [1], X. P. Ji [2], B. Su [1]

[1]Kunming Institute of Zoology, Chinese Academy of Sciences, [2]Yunnan Institute of Cultural Relics and Archaeology, Kunming, China

Abstract: Historically, from a couple of thousand years ago, The Hanging Coffin custom origin, dispersed and maintained a very long time, in a wide region covering southern China, Mainland Southeast Asia, Island Southeast Asia (including Taiwan) and near pacific regions. Along history, there is acutely controversy regarding the population histories and Hanging Coffin cultural dispersal pattern among different regions. Here, by genetic analyzing 46 individual human remains from 4 and 9 Hanging Coffin archaeological sites both in China and Thailand respectively, we reveal that the Hang coffins populations from different regions have a very recent common ancestry and some even high genetic homogeneity & very close genetic relationship each other, which reflect their consanguinity.

Both population and individual level deep dissection indicate the most genetic close present day populations with Hanging Coffin conductors, is the Daic language speakers both in southern China and Northern Thailand, we further found the Hanging Coffin populations also share some mtDNA diversities with other present day surrounding populations, which conforms recent admixture, or the possible of performed Hanging Coffin by the early ancestors of these peoples as well. Finally, by surprising, Comparing the others, we found that a Thai branch, Khon Mueang, present day Daic language speakers in Northern Thailand, have the most close genetic affinity with the prehistoric Hanging Coffin conductors in Thailand, who have lived there ~2,000 years ago.

Overall, from the matrilineal pattern, we reveal that the Hanging Coffin culture dispersal initially happened by demic diffusion in southern China from eastern coastline region, but after arrival to Southeast Asia, the major dispersal pattern have shift to cultural diffusion and widely spread to other surrounding local settlers.

Using Ancient DNA to Study Natural Selction: New Models and New Data
SMBE-PO-770

The evolution of skin pigmentation associated variation in Europe

D. Ju [1,*], I. Mathieson [1]
[1]Genetics, Perelman School of Medicine, Philadelphia, United States

Abstract: Differences in selection and demography among human populations underlie present-day variation in skin pigmentation. In particular, light pigmentation in Europe is widely believed to be an adaptation to low levels of UV radiation. However, the timing and nature of selection on pigmentation associated variants is poorly understood. Previous studies have looked at a small number of variants, but little is known about how polygenic selection operated in the context of European demographic history, which is marked by complex admixture of structured populations. We used ancient DNA from 1030 ancient West Eurasians to examine changes in allele frequencies over the past 45,000 years of 43 SNPs associated with skin pigmentation curated from the literature and 193 SNPs from the UK Biobank genome-wide association study.

Variants associated with decreased pigmentation increased in frequency over the past 45,000 years (p=2.2E-3) as well as over the past 10,000 years (p=9.0E-3), even when controlling for changes in ancestry. However, this recent trend is driven by a subset of strongly selected SNPs. Changes in frequency of many individual variants can be explained by changes in ancestry. We hypothesize that this observation can be explained by convergent evolution of light skin pigmentation in the ancestral populations, followed by recent admixture and ongoing selection. Using the population branch statistic, we identified pigmentation variants under independent selection in Mesolithic Hunter-Gatherers and Neolithic Early Farmers, including a well-known *SLC24A5* variant and a SNP in *BNC2*, both of which were selected on the Farmer lineage.

Using Ancient DNA to Study Natural Selction: New Models and New Data

SMBE-PO-754

Ancient genomics and the evolution of human height in Europe

S. L. Cox [1], C. B. Ruff [2], R. M. Maier [3 4 5], I. Mathieson [1,*]

[1]Department of Genetics, University of Pennsylvania, Philadelphia, [2]Center for Functional Anatomy and Evolution, Johns Hopkins University, Baltimore, [3]Program in Medical and Population Genetics, [4]Stanley Center for Psychiatric Research, Broad Institute, Cambridge, [5]Analytical and Translational Genetics Unit, Massachusetts General Hospital, Boston, United States

Abstract: Human height is an ideal trait to study as a model for the evolution of complex traits. Genome-wide association studies (GWAS) provide detailed information about the genetic basis of variation in the phenotype, while skeletal measurements reveal changes over time. In Europe, there is an excellent record of prehistoric changes in stature, and an increasingly complete ancient DNA record of genetic changes. We therefore set out to test whether observed changes in stature, and other anthropometric traits, could be predicted from ancient genetic data.

We collected published ancient DNA data from 1122 ancient West Eurasians dating to between 45 and 1 thousand years before present (kBP) and published skeletal measurements from 1159 individuals from comparable populations. We computed polygenic risk scores for height and other phenotypes for the ancient individuals using summary statistics from GWAS and within-family analysis of the UK Biobank.

We show that the ~10cm decrease in height from the Early Upper Paleolithic (45-29 kBP) to the Neolithic (8-4 kbp) is qualitatively consistent with changes in predicted genetic height due to admixture and migration. The ~3cm increase between the Neolithic and Bronze Age is also consistent, and likely driven by polygenic adaptation. In contrast, sitting height changes relatively little through time in both the skeletal measurements and PRS, indicating that changes in height were driven largely by changes in leg length. Geographic variation–particularly a post-Neolithic latitudinal gradient in height–is also reflected in the PRS. Finally, we show that changes in genetic bone mineral density are consistent with changes in femoral bending strength–demonstrating both genetic and plastic responses to changes in mobility in the Mesolithic and Neolithic.

Using Ancient DNA to Study Natural Selction: New Models and New Data

SMBE-PO-775

Mitochondrial DNA from Iron Age to present in Eastern Fennoscandia

S. Översti [1,*], K. Majander [1 2 3], E. Salmela [1 2], K. Salo [1], L. Arppe [1], S. Belskiy [4], H. Etu-Sihvola [1], V. Laakso [5], E. Mikkola [6], S. Pfrengle [3], M. Putkonen [1], J.-P. Taavitsainen [5], K. Vuoristo [6], A. Wessman [1], A. Sajantila [1], M. Oinonen [1], W. Haak [2], V. J. Schuenemann [3 7], J. Krause [2], J. U. Palo [1 8], P. Onkamo [1 5]

[1]University of Helsinki, Helsinki, Finland, [2]Max Planck Institute, Jena, [3]University of Tübingen, Tübingen, Germany, [4]Russian Academy of Science, St. Petersburg, Russian Federation, [5]University of Turku, Turku, [6]Finnish Heritage Agency, Helsinki, Finland, [7]University of Zürich, Zürich, Switzerland, [8]National Institute for Health and Welfare, Helsinki, Finland

Abstract: Ancient DNA has revealed that the process of Neolithization involved a turnover of mitochondrial DNA (mtDNA) lineages in Europe: haplogroup (hg) U, dominating in hunter-gatherers, was widely replaced by farmer-associated hgs such as H. However, the Neolithic contribution in contemporary populations varies. Although the mtDNA diversity among European populations is relatively uniform, Finns show a slightly different haplogroup composition. Present-day Finns exhibit a relatively high frequency of U and an internal substructure: U is more common in the north-east (NE) Finland and farmer-associated hgs in the south-west (SW). This spatial pattern could reflect the arrival of agriculture from the south-west, generally associated with the spread of the Corded Ware Culture c. 4800ya. Nevertheless, direct archaeological evidence for the timing and direction of the spread of agriculture is scarce. Based on pollen analysis it has been proposed that farming has been practiced only on a local scale until cultivation reached many areas of Finland as late as during early Iron Age, *c.* 100 AD. Exploring the past in Finland with ancient genetic data is hampered by poor survival of bone material in the acidic soil.

To provide insight into the past of Eastern Fennoscandia, we have analyzed mtDNA diversity from bone finds from Late Roman Iron Age to Middle Ages (300-1400 AD) yielding 70 complete mtDNA genomes. Sites Hiitola and Tuukkala are located in eastern Finland and sites Levänluhta, Luistari and Hollola in SW. In addition, 33 mitochondrial genomes from mainly historical burials (1400-1800 AD) from southern parts of country were included in analysis.

The 103 haplotypes obtained belonged to hgs observed in contemporary Finns, but the frequencies differed both geographically and temporally. Within the Iron Age and Medieval sites, the SW sites showed a higher frequency of U (58%) than the eastern sites (20%) whereas H showed an opposite trend: 28% in SW and 53% in the east. In addition, these older sites were distinct from historical burial sites (U 18% and H 46%) and from contemporary Finns (U 24% and H 33%). Furthermore, the SW Iron Age and Medieval sites showed uneven distribution of U subhaplogroups: Levänluhta had high frequency of U5a and modern Saami-related hg U5b1b1a whereas other, younger SW sites showed relatively high frequencies of U4. In a sequence-level analysis, Iron Age and Medieval SW sites had higher affinity to the contemporary NE Finland, while Hiitola and Tuukkala clustered with modern SW Finns.

Our results suggest that among the studied sites and contemporary Finns, there are varying levels of admixture in mitochondrial genepool. These ancestries consist of lineages related to hunter-gatherer populations (U4, U5), with some resembling the current Saami (specifically U5b1b1) and haplogroups possibly introduced with farming (H, J, K, T). Assuming that the hg composition has correlated with the mode of subsistence, the high prevalence of H in the eastern sites might reflect a bidirectional arrival of the farming-associated populations into Finland.

Using Ancient DNA to Study Natural Selction: New Models and New Data

SMBE-PO-772

Bayesian inference of evolutionary histories under time-dependent substitution rates

J. V. Membrebe [1,*], M. Suchard [2 3 4], A. Rambaut [5 6], G. Baele [1], P. Lemey [1]

[1]Department of Microbiology and Immunology, Rega Institute, KU Leuven, Leuven, Belgium, [2]Department of Biomathematics, David Geffen School of Medicine, University of California, [3]Department of Biostatistics, Fielding School of Public Health, University of California, [4]Department of Human Genetics, David Geffen School of Medicine, University of California, Los Angeles, United States, [5]Institute of Evolutionary Biology, University of Edinburgh, Edinburgh, United Kingdom, [6]Fogarty International Center, National Institutes of Health, Bethesda, United States

Abstract: Many factors complicate the estimation of time-scales for phylogenetic histories, requiring increasingly complex evolutionary models and inference procedures. The widespread application of molecular clock dating has led to the insight that evolutionary rate estimates may vary with the time frame of measurement. This is particularly well-established for rapidly evolving viruses that can accumulate sequence divergence over years or even months. However, this rapid evolution stands at odds with a relatively high degree of conservation of viruses or endogenous virus elements over much longer time-scales. Building on recent insights into time-dependent evolutionary rates, we develop a formal and flexible Bayesian statistical inference approach that accommodates rate variation through time. We evaluate the novel molecular clock model on a foamy virus co-speciation history and a lentivirus evolutionary history, and compare the performance to other molecular clock models. For both virus examples, we estimate a similarly strong time-dependent effect that implies rates varying over four orders of magnitude. The application of an analogous codon substitution model does not implicate long term purifying selection as the cause of this effect. However, selection does appear to affect divergence time estimates for the less deep evolutionary history of the Ebola virus genus. Finally, we explore the application of our approach on an ancient DNA data set of woolly mammoths, which shows a much weaker, but still important, time-dependent rate effect that has a noticeable impact on node age estimates. Future developments aimed at incorporating more complex evolutionary processes will further add to the broad applicability of our approach.

Using Ancient DNA to Study Natural Selction: New Models and New Data

SMBE-PO-773

Ancient DNA from the skeletons of Roopkund Lake reveals migrants of Mediterranean origin in South Asia

É. Harney [1 2 3,*], A. Nayak [4], N. Patterson [5 6], V. Mushrif-Tripathy [7], P. Joglekar [7], S. Walimbe [7], S. Mallick [2 5 8], N. Rohland [2], J. Sedig [2], N. Adamski [2 8], R. Bernardos [2], N. Broomandkhoshbacht [2 8], B. J. Culleton [9 10], M. Ferry [2 8], T. Harper [10], M. Michel [2 8], J. Oppenheimer [2 8], K. Stewardson [2 8], Z. Zhang [2], V. K. Srivastava [11], .. Harashawaradhana [12], S. Kumar [13 14], S. C. Diyundi [15], P. Roberts [4], N. Boivin [4], D. J. Kennett [9 10], K. Thangaraj [13], D. Reich [2 3 5 8], N. Rai [14]

[1]Dept. of Organismic and Evolutionary Biology, Harvard University, Cambridge, [2]Dept. of Genetics, Harvard Medical School, Boston, United States, [3]The Max Planck-Harvard Research Center for the Archaeoscience of the Ancient Mediterranean, Jena Germany/Cambridge USA, -, [4]Dept. of Archaeology, Max Planck Institute for the Science of Human History, Jena, Germany, [5]Broad Institute of Harvard and MIT, [6]Dept. Human Evolutionary Biology, Harvard University, Cambridge, United States, [7]Deccan College, Pune, India, [8]Howard Hughes Medical Institute, Boston, [9]Institute for Energy and the Environment, [10]Dept. of Anthropology, Pennsylvania State University, University Park, United States, [11]Anthropological Survey of India, Kolkata, [12]Anthropological Survey of India, Dehradun, [13]CSIR Centre for Cellular and Molecular Biology, Hyderabad, [14]Birbal Sahni Institute of Palaeosciences, Lucknow, [15]Gautam Budh Health Care Foundation, Noida, Uttar Pradesh, India

Abstract: Roopkund Lake—a high-altitude site situated ~5000 meters above sea level in the Himalayan Mountains—is home to the scattered skeletal remains of several hundred ancient individuals. Little is known about the identity of these people or their reason for traveling to Roopkund Lake. We present ancient DNA data from 38 skeletons from Roopkund Lake and find that these individuals cluster into three distinct groups. Over half of the individuals possess ancestry that is within the range of variation of most present-day South Asians. A second large group is comprised of individuals with ancestry typical of inhabitants of the eastern Mediterranean, with particular affinity to present-day populations from the Greek island of Crete. Additionally, we identify one individual with Southeast Asian-related ancestry. We obtained radiocarbon dates from all but one of the individuals, finding that these remains were deposited during at least two distinct events, with the individuals that possess ancestry typical of present-day South Asia dating to ~800 CE, and the individuals belonging to the other two groups dating to ~1800 CE. These differences in ancestry and time are also reflected in stable isotope measurements, which reveal a distinct dietary profile for the two main groups. These findings contradict previous hypotheses that the skeletons of Roopkund Lake are the remains of a single traveling party that was struck down during the same catastrophic event and highlight the ability of biomolecular tools to shed light on the history of even the most mysterious sites when data on archaeological context is limited.

Using Ancient DNA to Study Natural Selction: New Models and New Data
SMBE-PO-777

Ancient human genomes unmask our evolutionary history

C. D. Huber [1,*], Y. Souilmi [1], R. Tobler [1], A. Johar [1], M. Williams [1], S. Grey [2], M. Cox [3], W. Haak [4], J. Teixeira [1], A. Cooper [1]
[1]Australian Centre for Ancient DNA, University of Adelaide, Adelaide, [2]Garvan Institute of Medical Research, Darlinghurst, Australia, [3]School of Fundamental Sciences, University of New Zealand, Palmerston North, New Zealand, [4]Max-Plank-Institut fuer Menschheitsgeschichte, Jena, Germany

Abstract: Our understanding of human demographic history has advanced rapidly in the past decade, primarily due to explosive growth in ancient human genomic datasets from western Eurasia. However, only a handful of studies have used ancient genomes for exploring human adaptation. Consequently, our present understanding is predominantly based on genomes from modern populations, where the apparent lack of marked genomic signatures of selection has led to views that recent phases of adaptation involved selection on polygenic traits or from standing genetic variation. To investigate signatures of selection in ancient human populations, we examined more than 1000 ancient western Eurasian genomes for signatures of selection. Our results suggest that adaptation via hard selective sweeps has played a more expansive role in recent human history than previously appreciated – but that genetic mixing during and after the Bronze Age has obscured these genomic signatures in modern populations. We find that selective sweeps are aggregated in pathways specific to the immune response and the metabolism of carbohydrate byproducts (i.e. free radicals) and proteins, and that selection often targeted interacting gene cohorts directly involved in responding to stressful stimuli. We created a detailed reconstruction of adaptation during the past 50,000 years, including the transition from hunter-gatherer to farming lifestyles over the past 10,000 years. This study highlights the unique potential of ancient genomes to unmask previously hidden evolutionary histories and reveals potentially relevant medical genomic loci.

Using Ancient DNA to Study Natural Selction: New Models and New Data

SMBE-PO-779

Imputing gene regulatory changes between archaic hominins and modern humans reveals barriers to introgression and phenotypic differences

L. L. Colbran [1,*], E. R. Gamazon [1 2], P. Evans [1], N. J. Cox [1], J. A. Capra [1 3]

[1]Vanderbilt Genetics Institute, Vanderbilt University, Nashville, United States, [2]Clare Hall, University of Cambridge, Cambridge, United Kingdom, [3]Departments of Biological Sciences and Biomedical Informatics, Vanderbilt University, Nashville, United States

Abstract: Most aspects of archaic hominin biology cannot be directly studied due to their lack of preservation in fossil material. However, the sequencing of DNA derived from archaic fossils has enabled the study of these groups at the genetic level and revealed that archaic hominins and anatomically modern humans (AMHs) interbred. We tested this by adapting the PrediXcan approach to impute the effects of absent Neanderthal DNA on modern human gene regulation and identified 766 genes with significant evidence for differential regulation by non-introgressed Neanderthal DNA. including candidates in all introgression deserts. These divergently regulated genes function in diverse traits, including genes essential to language and circadian regulation, and they are enriched for associations with spontaneous abortion, polycystic ovary syndrome, myocardial infarction, and several cancers, including melanoma. We then analyze ~23,000 patients' biobank data and demonstrate that modern human variation in these genes' regulation is associated with cardiovascular and immune phenotypes. Next, we expand our analysis and impute the gene regulatory landscapes of two Neanderthals and a Denisovan and compare them to modern humans. This identifies differences in gene regulation relevant to immune system variation and morphological differences between modern and archaic hominins, including differences in skeletal and dental morphology consistent with the fossil record. These results establish substantial differences in gene regulatory architecture between AMHs and archaic hominins, suggest that gene regulatory divergence was a barrier to introgression of Neanderthal DNA, and provide an avenue for exploring phenotypic differences between archaic groups from genomic information alone.

Late Breaking Abstracts

Barriers and drivers of evolutionary innovation by horizontal gene transfer
SMBE-LBA-016

HORIZONTALLY ACQUIRED CYSTATINS AND STEFINS ARE NOVEL VIRULENCE FACTORS IN PATHOGENIC BACTERIA

D. Kordis*

Poster Submission: Cysteine proteases are the largest groups of proteases and play multi-faceted roles in every aspect of physiology and development. Cystatins are natural tight–binding reversible inhibitors of cysteine proteases. During the large scale analysis of cystatin superfamily in prokaryotes, we have found that eukayotic cystatins and stefins have been acquired and co–opted by a few bacterial organisms (pathogenic or comensal). One of the major roles of horizontally acquired cystatins and stefins in bacteria could be to evade host immunity or to protect them when in close contact with diverse eukaryotic hosts.

We performed a detailed comparative and evolutionary genomic analysis of the cystatins and stefins in numerous prokaryotic genomes and functionally and structurally characterized two stefins from pathogenic bacteria. Structural alignment has shown that bacterial stefins and cystatins possess both the inhibitory motif QXVXG and conserved N-terminal Gly residue. They lack however the PW or PG motif that is conserved in eukaryotic cystatins. This indicates that their modified cystatin domain has been adapted to inhibit a broad spectrum of cysteine proteases. In order to demonstrate the biochemical activity of bacterial stefins we expressed *Vibrio cholerae* stefin (hypothetical protein VCA0935) and *Bacteroides fragilis* fusion inhibitor containing chagasin and cystatin domains (hypothetical protein BF1388). We explored the inhibitory properties of recombinant VCA0935 and BF1388 proteins and determined their interaction constants with diverse cysteine proteases, cathepsins L, S, K, V, B and papain. Both VCA0935 and BF1388 were found to act as fast and tight binding inhibitors of endopeptidases cathepsins K, S, V, L and papain, however their interaction with exopeptidase cathepsin B was several orders of magnitude weaker. Interestingly, the pathogen stefins inhibits the endopeptidase activity of cathepsins S, K, L and V, which are all important players in the host adaptive and innate immunity.

Three main conclusions can be drawn from our study. First, the acquisition of novel virulence factors for pathogenic bacteria by horizontal gene transfer from eukaryotes is very rare and was not analysed extensively. Second, while there is strong evidence that proteases are essential virulence factors for prokaryotic and eukaryotic parasites and pathogens during all stages of infection processes, there are very few cases where protease inhibitors have been shown to assist pathogens in invading the eukaryotic hosts by inhibiting their proteases. Third, bacterial stefins and cystatins with inhibitory spectra for diverse families of cysteine proteases are especially suited to inhibit the numerous host proteases during infection. Therefore, the bacterial stefins and cystatins are novel virulence factors that could function in the invasion and dissemination of the pathogens.

Barriers and drivers of evolutionary innovation by horizontal gene transfer

SMBE-LBA-014

THE GENOME SEQUENCE OF A PALM PEST, THE RED PALM WEEVIL (RHYNCHOPHORUS FERRUGINEUS) REVEALS KEY FUNCTIONAL GENES AT THE PLANT-BEETLE INTERFACE

K. M. Hazzouri* and Naganeeswaran Sudalaimuthuasari, David Nelson, Mohammad Ali Al-Deeb, Alain Le Mansour, J. Spencer Johston, Claude Desplan, Khaled M.A Amiri

Poster Submission: Red palm weevil, *Rhynchophorus ferrugineus*(Olivier 1790) is one the most dangerous insect (Coleoptera: Curculionoidea), with major environmental and economic effects on the date palm (*Phoenix dactylifera*) industry, which is a major fruit crop of the arid regions. Comprehensive studies aiming at addressing the pest-host interaction, genetic manipulations, and genetic diversity are very scarce because of the lack of a good assembly of the genome. We report the sequencing of the whole genome of red palm weevil (male and female) using a combination of Illumina sequencing (Hiseq2500) paired ends (150bp) and 10X genomics. The final pseudochromosomes assembly is about 781 Mb for male and 783 Mb for female with an average scaffold size N50 (60 Mb). Genome size was estimated using flow cytometry (female: 726+/- 12.8 Mb; male: 696.3 +/- 5.3 Mb). The flowcytometry results shows signs of endoreduplication in the nervous system, where copy number variation shows duplication of some insecticide resistance genes (e.g. Resistance to Dieldrin (Rdl), glutathione S-transferase (GST)) mapped to these locations. Furthermore, we highlight some important gene families (e.g., plant cell wall degradation (GH16), which are horizontally transferred into the weevil genome and are under strong positive selection in the species. The genome assembly of the red palm will be a valuable resource that illuminates insect evolution, behavior and will be a step toward genetic modification of key important genes that will help eradicate this pest.

Barriers and drivers of evolutionary innovation by horizontal gene transfer

SMBE-LBA-006

HORIZONTAL-GENE TRANSFER AS A SOURCE OF NOVEL SRNAS IN SALMONELLA TYPHIMURIUM

B. R. Jose [1,*], L. Barquist [2], P. P. Gardner [1]

[1]Department of Biochemistry, University of Otago, Dunedin, New Zealand, [2]Helmholtz Institute for RNA-based Infection Research, Helmholtz-Centre for Infection Research, Würzburg, Germany

Poster Submission: Bacterial sRNAs are small, highly structured non-coding RNA molecules, that may function through interactions with other biomolecules, forming complex regulatory networks. Many sRNA families are expressed under specific growth conditions, and have been found to regulate virulence factors in pathogenic bacteria, such as the *Salmonella* Typhimurium sRNA PinT which controls the expression of *Salmonella* virulence factors during macrophage infection1.

Studies of the evolutionary dynamics and origins of sRNA genes have been hindered by poor sequence conservation, which makes annotation via sequence homology challenging. The short length and relative simplicity of sRNA genes also make them interesting candidates for observing *de-novo* gene formation from transcriptional noise, or exaptation from existing elements2.

We have used a pipeline based on profile hidden Markov models (HMMs) to study the conservation patterns of sRNA genes from *Salmonella* Typhimurium. Our results show that sRNAs are both rapidly acquired and exhibit rapid sequence turnover. We found that horizontal gene transfer is the main driver of sRNA acquisition in *Salmonella*, and identified *Salmonella*-specific sRNAs that appear to be derived from phage control systems, and other mobile genetic elements, as well as Type I toxin-antitoxin systems.

1. Westermann, A. J. *et al.* Dual RNA-seq unveils noncoding RNA functions in host-pathogen interactions. *Nature* 529, 496–501 (2016).
2. Jose, B. R., Gardner, P. P. & Barquist, L. Transcriptional noise and exaptation as sources for bacterial sRNAs. *Biochem. Soc. Trans.* (2019). doi:10.1042/BST20180171

Biochemistry, epistasis and the evolutionary process

SMBE-LBA-012

A HYDROPHOBIC RATCHET ENTRENCHES MOLECULAR COMPLEXITY IN THE STEROID HORMONE RECEPTORS

G. K. A. Hochberg[*], Y. Liu[1], E. Marklund[2], A. Laganowsky[3], J. Thornton[4]

[1]Department of Chemisrty, University of Texas A&M, College Station, United States, [2]Uppsala University, Uppsala, Sweden, [3]University of Texas A&M, College Station, [4]Department of Ecology & Evolution, University of Chicago, Chicago, United States

Poster Submission: Many proteins assemble into multisubunit complexes , which are usually thought to exist because they enable or enhance molecular functions. But complexes could also be irreversibly maintained across evolutionary time, if their subunits acquire substitutions that can be tolerated only in the assembled form. Here we provide evidence and a mechanism for ancient entrenchment of multimeric interactions during the evolution of the steroid receptor family of transcription factors (SRs). Of the two major SR classes, the estrogen receptors (ERs), dimerize via a large hydrophobic interface in their ligand-binding domain (LBD), but the other members of the family -- the ketosteroid receptors (kSRs) -- lack this interaction. Using ancestral protein reconstruction and biochemical assays, we show that the dimeric form has been entrenched across hundreds of millions of years of ER evolution because exposing the ancient interface to solvent reduces protein stability and impairs transcriptional activation, even under conditions in which dimerization per se does not enhance function. In kSRs, the form of hydrophobic entrenchment shifted: a new C-terminal extension evolved, which binds to and shields the ancestral hydrophobic interface; this intramolecular interaction has been entrenched ever since, because compromising it causes misfolding, aggregation, and a loss of function. We identify dozens of other protein families in which dimer interfaces show similar signs of long-term hydrophobic entrenchment. Many protein complexes may therefore be maintained by purifying selection because the capacity of their subunits to exist as monomers was compromised long ago by hydrophobic mutations that are neutral only in the assembled state, even if the complexity itself is functionally inconsequential.

Biochemistry, epistasis and the evolutionary process

SMBE-LBA-062

CAN MULTIDIMENSIONAL EPISTASIS BE NOT SIGN EPISTASIS?

D. Ivankov [1,*]

[1]Center of Life Sciences, Skolkovo Institute of Science and Technology, Moscow, Russian Federation

Poster Submission: Genotype to phenotype relationship is one of the main unresolved problems of evolutionary biology. The main obstacle to a better prediction of phenotype from genotype is epistasis, a dependence of a mutation contribution on genetic context.

In this work we focus on the distinguishing unidimensional epistasis, where fitness or phenotype is a one-variable monotonic function of additive variable called fitness potential, from multidimensional epistasis where one fitness potential is not enough for existence of monotonic function, and two or more fitness potentials are needed for such a description. Up to now, the only cases of multidimensional epistasis are sign and reciprocal sign epistasis.

We show here that more complicated cases of multidimensional epistasis exist and are abundant in experimentally measured fitness landscapes. Therefore, the multidimensional epistasis is a more complicated phenomenon than just a sign epistasis. The biological meaning and evolutionary role of new types of multidimensional epistasis still remain unclear.

Contemporary Evolution

SMBE-LBA-015

ROBUST ESTIMATION OF RECENT EFFECTIVE POPULATION SIZE FROM NUMBER OF INDEPENDENT ORIGINS IN SOFT SWEEPS

B. S. Khatri*, A. Burt [1]

[1]Life Sciences, Imperial College, Ascot, United Kingdom

Poster Submission: Estimating recent effective population size is of great importance in characterising and predicting the evolution of natural populations. Methods based on nucleotide diversity may underestimate current day effective population sizes due to historical bottlenecks, whilst methods that reconstruct demographic history typically only detect long-term variations. However, soft selective sweeps, which leave a fingerprint of mutational history by recurrent mutations on independent haplotype backgrounds, holds promise of an estimate more representative of recent population history. Here we present a simple and robust method of estimation based only on knowledge of the number of independent recurrent origins and the current frequency of the beneficial allele in a population sample, independent of the strength of selection and age of the mutation. Using a forward time theoretical framework, we show the mean number of origins is a function of $\theta=2N\mu$ and current allele frequency, through a simple equation, and the distribution is approximately Poisson. This estimate is robust to whether mutants pre-existed before selection arose, and is equally accurate for diploid populations with incomplete dominance. For fast (e.g., seasonal) demographic changes compared to time scale for fixation of the mutant allele, and for moderate peak-to-trough ratios, we show our constant population size estimate can be used to bound the maximum and minimum population size. Applied to the Vgsc gene of \textit{Anopheles gambiae}, we estimate an effective population size of roughly 6×10^7, and including seasonal demographic oscillations, a minimum effective population size greater than 3×10^7 and a maximum less than 6×10^9, suggesting a mean ~10^9.

Contemporary Evolution

SMBE-LBA-040

COMPARATIVE INVASION GENOMICS OF A WIDE-SPREAD AVIAN INVADER, THE COMMON MYNA (ACRIDOTHERES TRISTIS)

T. Magory Cohen [1,*], K. Ewart [2], R. Major [2], R. S. Kumar [3], M. Nair [3], R. Dor [1]

[1]Tel Aviv University, Tel Aviv, Israel, [2]The Australian Museum, Sydney, Australia, [3]Wildlife Institute of India, Dehradun, India

Poster Submission: Invasive species are considered to be a primary threat to the natural environment. Understanding the mechanisms that enable successful biological invasions through microevolutionary adaptations is essential to managing the geographic spread of such species. Current theory suggests that the amount of genetic variation in the introduced populations affects invasion success by introducing potential adaptive phenotypes, and that it decreases from the invasion source to the range front. However, the rate in which new mutations accumulate (in invasive populations) compared with optimal genetic variation (in the native population) has been scarcely studied. The common myna (*Acridotheres tristis*), a wide-spread avian invader, is considered one of the 100 world worst invasive species due to its high adaptation capability to new environments, but its introductions included known demographic bottlenecks. We aimed to explore the genomic changes that occurred during biological invasions by studying native (India) and introduced (Israel – new invasion, Australia – old invasion) natural populations of the common myna. We extracted DNA from individuals captured in Israel (n=160), India (n=27) and Australia (n=462) and used Restriction-site-associated DNA markers (RAD-sequencing) to generate a robust, high-resolution array of single nucleotide polymorphism markers (DArTseq™). We applied population genomic analysis techniques to determine gene flow, genetic diversity and population structure in order to estimate the recovery from the genetic confounding event experienced by the introduced populations compared with the native population. We estimated genetic diversity in recently established populations (Israel) compared with populations in older invaded areas (Australia), and in native populations (India). We found evidence of significant genetic structure in Australian populations, with gene flow directionality that was higher from source to front and compared it to Indian populations and Israeli populations. We located the point of introduction in Israel and in Australia with evidence for subsequent secondary introductions in the latter. We also found significant signals for isolation by distance among front populations that derived from the same source population, with genetic diversity decreasing away from the invasion source. These findings suggest that high new mutation accumulation rate is a key property in facilitating invasion success in the common myna. Moreover, we show that local adaptation can occur despite relatively low genetic diversity in recently established populations, indicating possible involvement of additional mechanisms such as differential gene expression or epigenetic methylation processes.

Evolution of phenotypes: understanding diversity and the role of plasticity in adaptation to new environments
SMBE-LBA-013

A GENOME SCAN FOR ADAPTATION TO HIGH ALTITUDE IN WILD RHESUS MACAQUES

Z. Szpiech [1,*], L. Stevison [1]
[1]Biological Sciences, Auburn University, Auburn, United States

Poster Submission: When natural populations split and migrate to different environments, they may experience different selection pressures that can lead to localized adaptation. For aerobic life, the low atmospheric oxygen content of high altitude living presents a special challenge and a strong selection pressure. Searching for evidence of adaptation to high altitude, we analyze the whole genomes of 23 wild rhesus macaques captured at high altitude (>4000m above sea level) alongside 22 wild rhesus macaques captured at lower altitude (<500m above sea level). Using XP-EHH, a haplotype-based genomic scan for differential local adaptation, we find evidence of local adaptation in the high-altitude population at or near 66 known genes and several unannotated regions. As a set, gene ontology analysis shows these genes are overrepresented for cellular metabolic processes (FDR 3.73e-2), including a very strong signal at one gene encoding a critical protein involved in the citric acid cycle. Together, this may suggest a metabolic adaptation to cope with very low atmospheric oxygen.

Evolution of phenotypes: understanding diversity and the role of plasticity in adaptation to new environments
SMBE-LBA-019

NATURAL VARIATION IN RECOMBINATION RATE IS SHAPED BY ENVIRONMENTAL CONDITIONS IN WILD BARLEY

S. Dreissig [1,*], M. Mascher [2 3], S. Heckmann [1]
[1]Meiosis Research Group, [2]Domestication Genomics Research Group, [3]IPK Gatersleben, Gatersleben, Germany

Poster Submission: Meiotic recombination generates genetic diversity upon which selection can act. Recombination rates are highly variable between species, populations, individuals, sexes, chromosomes and chromosomal regions. The underlying mechanisms are controlled at the genetic and epigenetic level and show plasticity towards the environment. Environmental plasticity may be divided into short-term and long-term responses. We estimated recombination rates in natural populations of wild barley and domesticated landraces. In wild barley, high recombination rates are found in more interstitial chromosome regions compared to domesticated barley, which differ in gene context compared to the distal ends of the chromosome. Among sub-populations of wild barley, natural variation in recombination rate is correlated with temperature, isothermality, solar radiation, and precipitation in a reverse U-shape manner. Increased recombination in wild barley populations subjected to intermediate environmental conditions, rather than extremes, could be an adaptation of the recombination machinery as a mechanism to maintain fitness in a strictly inbreeding species.

Evolution of phenotypes: understanding diversity and the role of plasticity in adaptation to new environments
SMBE-LBA-011

JUNK DNA RECYCLED: INCOMPLETE ELIMINATION OF GERMLINE DNA SEQUENCES MIGHT FAVOR SWIFT ADAPTIVE RESPONSES TO ENVIRONMENTAL CHANGES IN THE CILIATED PROTOZOAN PARAMECIUM

V. Vitali [1,*], R. Hagen [1], F. Catania [1]
[1]Institute for Evolution and Biodiversity, University of Muenster, Muenster, Germany

Poster Submission: How do new evolutionary novelties arise? Can otherwise discarded DNA regions be used to generate functional somatic novelties? In *Paramecium*, functional nuclear differentiation requires the efficient and reproducible elimination of thousands of germline DNA sequences known as Internal Eliminated Sequences (IESs). This impressive feat of genome remodeling is accomplished via Programmed DNA Elimination (PDE), a small-RNA guided genome-editing program evolved to safeguard the integrity of the somatic genome against junk DNA accumulation. Previous studies suggest that PDE is not entirely accurate. Here we extend these findings and find that departure from the standard cultivation temperature of 25-27°C leads to elevated rates of incomplete IES excision. Further, many IESs that are retained in the polyploid somatic nucleus are under epigenetic control**,** and thus might be passed on to the sexual offspring. On this basis, we hypothesize that somatic IES retention may occasionally be co-opted to modulate gene expression and/or diversify coding sequences. This co-option should facilitate adaptation to a new environment and is under scrutiny in an experimental evolution study that is ongoing in our lab. Exploring the environmental sensitivity of PDE can further current understanding of how new functions emerge.

Evolution of phenotypes: understanding diversity and the role of plasticity in adaptation to new environments
SMBE-LBA-057

RAPID PHENOTYPIC EVOLUTION WITH SHALLOW GENOMIC DIFFERENTIATION DURING THE EARLY STAGE OF HIGH ELEVATION ADAPTATION IN EURASIAN TREE SPARROWS

Y. Qu[*], W. Zhai [1], Y. Zhang [1], S. DuBay [2], P. Ericson [3], F. Lei [1]
[1]Key Laboratory of Zoological Systematics and Evolution, Institute of Zoology, Chinese Academy of Sciences, Beijing, China, [2]University of Chicago, Chicago, United States, [3]Swedish Museum of Natural History, Stockholm, Sweden

Poster Submission: Known as the “third polar region”, the Qinghai-Tibet Plateau represents one of the harshest highland environments in the world, yet a number of organisms thrive there. Previous studies have focused on well differentiated populations in the later stage of phenotypic divergence. The adaptive processes during the initial phase of highland adaptation remain poorly understood. We studied a human commensal, the Eurasian Tree Sparrow, which has followed human beings to the Qinghai-Tibet Plateau. Despite strong phenotypic differentiation at multiple levels, in particular muscle related phenotypes, the highland and lowland populations show shallow genomic divergence and the colonization event is found to be around thousands years ago. In an acclimation experiment exposing lowland tree sparrows to a hypoxic environment, we did not observe muscle phenotypic changes, suggesting that phenotypic plasticity might not be responsible for reshaping the rapidly evolved muscle phenotypes. Through population genetic analyses, we identified a signature of polygenic adaptation, whereby shifts in allele frequencies spread across multiple loci, many of which are strongly associated with muscle related processes. Our results reveal an exceptional case of positive selection in which polygenic adaptation appears to drive rapid phenotypic evolution, shedding light on early stages of adaptive evolution to a novel environment.

Evolution of phenotypes: understanding diversity and the role of plasticity in adaptation to new environments
SMBE-LBA-025

DIRECT EVIDENCE FOR SOMATIC CELL INHERITANCE

E. O'Brien [1], K. Ensbey [1], B. Day [1], P. Baldock [2], G. Barry [1,*]
[1]QIMR Berghofer, Brisbane, [2]Garvan Institute of Medical Research, Darlinghurst, Australia

Poster Submission: The traditional concept that heritability occurs exclusively from the transfer of germline-restricted genetics is being challenged by the increasing accumulation of evidence confirming the existence of experience-dependent transgenerational inheritance. Transgenerational inheritance is emerging as a powerful mechanism for robustly transmitting phenotypic adaptations to offspring. However, questions remain unanswered as to how this heritable information is passed from somatic cells. Previous studies have implicated the critical involvement of RNA in heritable transgenerational effects and the high degree of mobility and genomic impact of RNAs in all organisms is an attractive model for the efficient transfer of genetic information. Here we show, for the first time, transport of RNA from the brain of an adult male mouse to sperm, and subsequently to offspring. Our observation of heritable genetic information originating from a somatic tissue may reveal a mechanism for how transgenerational effects are transmitted to offspring.

Evolution of phenotypes: understanding diversity and the role of plasticity in adaptation to new environments
SMBE-LBA-008

NEUTRAL EVOLUTION OF CELLULAR PHENOTYPES

J. G. Wideman*, A. Novick, W. F. Doolittle

Poster Submission: The intracellular and extracellular morphological features of eukaryotes are extremely diverse. Single-celled eukaryotes can be as small as some bacteria and as large as some multicellular animals. Even with this structural diversity, the basic architecture of eukaryotic cells is fairly constant. They have a nucleus, Golgi, cytoskeleton, plasma membrane, vesicles, ribosomes, and all but one known lineage has mitochondria-related organelles. Eukaryotes undergo processes like mitosis, meiosis, recombination, and often perform feats like phagocytosis, amoeboid movement, and flagellar motility. With all of these commonalities it is obvious that eukaryotes evolved from a common ancestor, but it is not obvious how eukaryotes came to have their diverse phenotypes. Are diverse eukaryotic phenotypes adaptations due to positive natural selection, adapting each lineage to its particular niche? Or is eukaryotic cellular diversity substantially derived from neutral evolutionary processes, with niche adaptation either illusory or a secondary consequence? Here, we outline how a hierarchical view of phenotype can be used to articulate a neutral theory of phenotypic evolution through processes such as gene loss, gene replacement by homo/xeno/analogues, gene duplication followed by subfunctionalization, and constructive neutral evolution. We suggest that iterations of these contingent processes followed by entrenchment of their products can explain much of the diversity of cellular, developmental, and biochemical phenotypes of eukaryotes and should be explored in addition to adaptive explanations.

Evolution of phenotypes: understanding diversity and the role of plasticity in adaptation to new environments
SMBE-LBA-038

PRE-ADAPTATION TO CLIMATE CHANGE THROUGH TOPOGRAPHY-DRIVEN PHENOTYPIC PLASTICITY

H. De Kort [1,*], K. Helsen [1], B. Panis [2], O. Honnay [1]
[1]Biology, [2]Biosystems, Leuven University, Leuven, Belgium

Poster Submission: Climate change will increase the level of drought stress experienced by plant communities, yet the spatial distribution of projected changes in dryness remains highly uncertain. Species can, to some extent, deal with climate uncertainty through natural variation in adaptive responses to environmental heterogeneity and predictability. Biodiversity conservation could thus target populations pre-adapted to climatic heterogeneity to anticipate climate uncertainty. Disentangling evolution of trait means vs. trait plasticity, however, requires a sampling design with genetic replicates grown under distinct environmental conditions.

Here, we applied three soil moisture treatments to genetic replicates of *Fragaria vesca* plants raised from seeds that were sampled in distinct topographical settings, to study adaptive trait and plasticity divergence in response to drought. We demonstrate that various fitness traits evolved along topographical gradients, including increased specific leaf area (SLA) with increasing slope, and increased growth plasticity with increasing altitude.

Our results indicate that traits and their plasticity can evolve independently in response to distinct topographical stressors. We further show that trait heritability varies considerably among traits and topographical settings. Heritability of phenotypic plasticity tended to increase with altitude for all traits, with populations from high altitudes harboring more than twice the heritability for growth and SLA plasticity compared to populations from low altitudes.

We conclude that (i) low altitudinal populations, which are expected to be most vulnerable to climate change, may only withstand limited increases in drought stress, while (ii) populations that evolved to thrive in heterogeneous conditions are likely pre-adapted to climate change through high plasticity and heritability. Heterogeneous landscapes thus represent invaluable sources of quantitative genetic variation that could support conservation where climate projections are inconclusive.

Evolution of proteins and molecular pathways/networks under higher level selective effects
SMBE-LBA-065
ADAPTIVE EVOLUTION AND FUNCTIONAL DIFFERENTIATION OF TESTIS EXPRESSION GENES IN THERIA
Y. Katsura*

Poster Submission: An expression pattern of genes differs in different tissues, and it is known the expression pattern of genes in testes is extremely variable in different species. In order to clarify how the testis transcriptomic pattern has been changed in species, we examine the evolution of testis transcriptome in Theria using 10 species. The species used are 2 marsupials (opossum and Tasmanian devil), 6 eutherian (placental) mammals (human, chimpanzee, bonobo, gorilla, rhesus macaque and mouse) and 2 outgroup species (platypus and chicken), and the Tasmanian devil transcriptome was obtained by RNA sequencing in this study. Here I show that 22 testis expression genes are marsupial-specific suggesting acquisition in the stem lineage of marsupials after the divergence from eutherians, and of them *PR/SET Domain 1*(*PRDM1*) is highly expressed in opossum newborn testes during sex determination. Despite the time length of the eutherian stem lineage being similar to the marsupial one, testis expression genes were not found in the stem lineage of eutherians suggesting the rapid gene evolution in eutherians. Thus the therians have shared 15 testis expression genes, and the evolutionary tempo of three testis genes is faster in eutherians from that of marsupials. The adaptive evolution of *Rho GTPase activating protein 28*(*ARHGAP28*) is suggested to be specific to the common ancestor of eutherians.

Evolution of proteins and molecular pathways/networks under higher level selective effects
SMBE-LBA-068

FISH AND CHIPS: USING OUR UNDERWATER ANCESTORS TO BETTER UNDERSTAND POLYMORPHISM IN THE HUMAN GENOME

S. V. Good [1,*], R. Gotesman, S. Yegorov [2], A. D. Patterson [1]
[1]Genetics and Genomics, The Hospital for Sick Children, Toronto, Canada, [2]Department of Natural Sciences Education, Suleyman Demirel University, Almaty, Kazakhstan

Poster Submission: FISH AND CHIPS: USING OUR UNDERWATER ANCESTORS TO BETTER UNDERSTAND POLYMORPHISM IN THE HUMAN GENOME

Ryan Gotesman, Sergey Yegorov, Andrew D. Paterson and Sara V. Good

Genome-wide association studies (GWAS) have identified thousands of single nucleotide polymorphisms (SNPs) that influence complex traits. The distribution of GWAS-significant SNPs in the human genome is non-random. We hypothesized that the evolutionary duplication history and linkage relationship of regions across the human genome influences the number of GWAS-significant SNPs, levels of gene constraint and gene ontology. To test this hypothesis, we mapped the start and end base-pair positions of each vertebrate ancestral chromosomes (VAC) segment in the human genome, calculated the number of GWAS-significant SNPs in them and compared it to the expected frequency of SNPs in the region based on HapMap. Next we assessed whether there was gene function enrichment in gene sets found in each VAC. Lastly, we examined whether the gene sets in different VACs harbour genes with different levels of evolutionary constraint (pLI index). We find that the number of trait-associated SNPs in a region was dependent on the ancestral chromosome from which the SNP originated and detect differences in ontological enrichment in regions with more GWAS-significant SNPs.

Evolution of proteins and molecular pathways/networks under higher level selective effects
SMBE-LBA-026

PROBABILISTIC MODELLING OF PROTEIN STRUCTURE EVOLUTION

M. T. Golden*, O. G. Pybus[1]
[1]Department of Zoology, University of Oxford, Oxford, United Kingdom

Poster Submission: We present a probabilistic model of protein evolution that captures several important features of protein sequence and local structure. The key feature being dependencies between neighbouring amino acid positions that are temporal in nature due to sequence mutations that occur during evolution. The model is trained on a large number of protein alignments and corresponding phylogenetic trees that represent the evolutionary history of the aligned proteins. This yields a model that acts as a rich prior distribution over protein evolution. Our model provides a complete probabilistic description of protein backbone structures using an angle and bond length representation. Structure evolution is modelled jointly with sequence, permitting ancestral structures and sequences to be reconstructed in a phylogenetically rigorous manner. Likewise, the model can perform homology modelling to predict the unknown local backbone structure of a known protein sequence using additional information from potentially large numbers of homologous proteins. The model is highly flexible with respect to input, implying that arbitrary combinations of protein sequences and structures can be used when performing various inference tasks. Our current model does not capture global features of protein structure that are necessary for accurate homology modelling or reconstruction of ancestral three-dimensional structures. However, it is ultimately expected to be combined with protein structure prediction models that account for such long-range dependencies, but that do not account for evolutionary information.

Evolutionary genetics and genomics of metabolic networks
SMBE-LBA-066

LATERAL GENE TRANSFERS REVEAL AND CONTEXTUALIZE ANCESTRY IN THE ARCHAEPLASTIDA

E. S. Morrison [1,*], J. A. Burns [2]
[1]Sackler Institute for Comparative Biology, [2]Sackler Insitute for Comparative Biology, American Museum of Natural History, New York, United States

Poster Submission: The information contained in a genome is the outcome of varying rates of ancestral losses, duplications, changes and exchanges across biological scales ranging from single nucleotides all the way to metabolic pathways. While no two genomes are identical, those from organisms with shared ancestry should contain similar hallmarks of ancient evolutionary changes. In plastid-bearing organisms, the origin of many metabolic genes can be traced to lateral gene transfers from a prokaryote, most often from a cyanobacterium, the presumed ancestor of modern plastids. There are currently three major extant lineages of photosynthetic eukaryotes that contain ancient "primary" plastids (chloroplasts) derived from endosymbiosis with a cyanobacterium. They are the Rhodophyta (red algae), Chloroplastida (green algae and land plants), and Glaucophyta, together forming the taxonomic supergroup Archaeplastida. Although it is clear that the three lineages are related, traditional phylogenomics methods have not been able to definitively resolve their evolutionary relationships.
In this study, we reconstruct the history of gene losses and gains via lateral transfer in the evolution of metabolic pathways in plastid-bearing eukaryotes to infer the ancestral relationships among the Archaeplastida lineages. We have found evidence that instead of a one-gene-at-a-time ratchet-like process of lateral gene transfer, the ancient cyanobacterium may have donated all of its genes to the Archaeplastida ancestor at once; lateral genome transfer rather than lateral gene transfer. Then, over time, the Archaeplastida ancestor genes and the symbiotic cyanobacteria genes equilibrated, retaining duplicate copies in some cases, and losing redundant copies in others. These lineage-specific patterns of lateral transfers and gene losses follow known evolutionary groupings of the three modern Archaeplastida lineages and contextualize their relationships in a new way.

Genetic conflicts in molecular evolution
SMBE-LBA-007

POP-CON : A TOOL TO DETECT POPULATION GENETIC CONFLICTS

Y. Anselmetti [1,*], P. Simion [1], P. Tran Van [2], K. Jaron [2], M. Robinson-Rechavi [2], T. Schwander [2], N. Galtier [1]
[1]Institut des Sciences de l'Évolution de Montpellier, Montpellier, France, [2]Department of Ecology and Evolution, University of Lausanne, Lausanne, Switzerland

Poster Submission: We propose a new tool called Pop-Con (detection of **Pop**ulation genetic **Con**flicts) that allows to detect unexpected genotype profiles under the Hardy-Weinberg equilibrium. Here, we define a genotype profile as the ordered set of individuals genotypes at a given site.
Pop-Con takes as input a variant calling VCF file produced with the variant callers GATK [1] or read2snp [2]. Pop-Con has been designed for diploid species with a maximum of 2 alleles per site.
The Pop-Con approach plot the observed Site Frequency Spectrum (SFS) with the proportion of the genotype profiles for each pic of the plot. In complement, Pop-Con draws a mirror SFS plot with the expected proportion of genotype profiles under the Hardy-Weinberg equilibrium.
In this communication, we'll highlight the usage of Pop-Con, with two case studies:

- First case study will show contrast between a pair of close related sexual and asexual species in population genomic signatures. For this case study, we'll present two pairs of sexual/asexual species illustrating two different modes of asexual reproduction.
- Second case study will highlight an example of quick detection of an unexpected population genomic signature under the Hardy-Weindberg theorem in sexual species. Detection and localization of such unexpected genomic to get further analysis of genomic regions for biological process.

[1] McKenna, A. et al. The Genome Analysis Toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res. 20, 1297–1303 (2010).
[2] Gayral, P. et al. Reference-Free Population Genomics from Next-Generation Transcriptome Data and the Vertebrate–Invertebrate Gap. PLOS Genetics 9, e1003457 (2013).

Genetic conflicts in molecular evolution

SMBE-LBA-020

THE IDENTIFICATION OF NEW PROTEIN METHYLATION SITE ACQUISITION IN CONSERVED PROTEINS IN HUMAN LINEAGE.

D.-S. Kim, S. Lee, J. KIM [1,*]

[1]Department of Environmental Disease Research Centers, Kribb, Daejeon, Korea, Republic Of

Poster Submission: Protein methylation plays an important role in various biological processes, including chromatin structure remodeling, gene transcription, DNA repair, protein synthesis, RNA metabolism, cell cycle progression, apoptosis and signal transduction. We analyzed the human methylproteome data set(2,227 protein methylation site in 952 proteins) and identified 21 novel protein methylation site in 20 protein that arose in the human lineage since the last common ancestor of Euarchonta(primates and treeshrews). Among them, Arg-320 of Proline and serine-rich protein 2(PROSER2) and Arg-513 of Sodium channel protein type 5 subunit alpha (SCN5A) are human-specific. The results of molecular evolutionary analysis suggest that these sites were under positive selection during human evolution. We suggest that the novel methylation sites identified in this study may be useful for functional analysis to identify progressive genetic modifications to advantageous phenotypes acquired in the human lineage

Genome-wide methods for detecting selection
SMBE-LBA-061
EXPECTATION PROPAGATION ABC FOR POPULATION GENETIC INFERENCE
R. Rasteiro [1,*], J. Bridle [1], M. A. Beaumont [1]
[1]School of Biological Sciences, University of Bristol, Bristol, United Kingdom

Poster Submission: In recent years, several computational approaches have been introduced into Population Genetics aiming to provide improved inference of demographic history from genome-level samples. The use of approximate Bayesian computation (ABC) for whole-genome analysis can be challenging, particularly when considering the scale of simulation needed. However, there are potential advantages in using ABC when complex models are considered, including those that model the interplay of natural selection and demographic inference.
We developed a new model-based simulation approach to study evolutionary adaptation and demography from whole-genome data. It incorporates both coalescent and forward simulators and parallel Expectation Propagation ABC (EP-ABC). EP-ABC makes use of a "divide to conquer" approach together with recent developments in machine-learning, to enable efficient distributed computation in genomic analysis. Finally, we apply our approach to study the impact of demography and recent environmental changes in UK *Lepidoptera* species.

Genome-wide methods for detecting selection
SMBE-LBA-042
SIGNATURES OF SELECTION IN SNAPDRAGONS (ANTIRRHINUM MAJUS)

D. Shipilina [1,*], A. Estandía [1], A. Whibley [2], E. Coen [3], N. Barton [1], D. Field [4], M. Pickup [1]
[1]IST Austria, Klosterneuburg, Austria, [2]School of Biological Sciences, University of Auckland, Auckland, New Zealand, [3]1Department of Cell and Developmental Biology, John Innes Centre, Norwich, United Kingdom, [4]Department of Botany and Biodiversity Research, University of Vienna, Vienna, Austria

Poster Submission: The question of how natural selection shapes the genomes of organisms is a central question in evolutionary biology. Traditionally, most attention has focused on divergent selection – a form of selection, where genetic differences accumulate and lead to the formation of new species. Yet different forms of selection may act simultaneously, creating a tension between different parts of the genome. Balancing selection – where multiple alleles are maintained in populations – is an alternative form of selection that may interact with divergent selection to alter genomic patterns of diversity and divergence. In this study, we aim to examine how both these forms of selection operate within the model plant, *Antirrhinum majus*.
Detecting genomic signatures of different types of selection is an ambitious task that has three requirements: (i) a large sample size from two closely related subspecies, (ii) a deeply sequenced genome, and perhaps most importantly, (iii) identified and functionally described candidate loci.
The Snapdragon system (*Antirrhinum majus*) meets all the above-mentioned criteria as it contains well-defined target regions with confirmed functions – flower colour loci (*ROSEA*, *ELUTA* - further RosEl) and a self-incompatibility locus (S-locus). Previous studies suggest that the genomic region responsible for the magenta colouration of flowers (RosEl) is subject to divergent selection. While the well-defined gametophytic self-incompatibility system provides an example of negative-frequency dependent selection, a form of balancing selection where rare alleles have a selective advantage. Here we use an extensive data set containing both pooled (*A. m. pseudomajus* pool 50 plants, *A. m. striatum* 52 plants) and individual (6 samples from each subspecies) data.
We observed genomics signals of divergent and balancing selection by analyzing sequence variation. By comparing genomic regions, we found that the self-incompatibility locus showed higher nucleotide diversity (π_w) both for pooled and individual data, while for RosEL(divergent selection) diversity was slightly lower than the genomic average. Within two populations we performed F_{ST} scans, which showed lower differentiation in S-locus, while, in contrast, the genes responsible for flower colour displayed sharp peaks in $F_{ST.}$ Another signature of balancing selection is an elevated number of rare alleles. Our analysis of Tajima's D and the joint allele frequency spectrum demonstrate a greater frequency of rare alleles in the S-locus. Finally, we applied two methods of demographic inference to obtain migration rate values in both regions. We observed a slightly higher effective migration rate in S-locus, that is likely due to negative frequency-dependent selection. These results demonstrate the genomic signatures of two different forms of selection, providing a unique example of how natural selection shapes the genome and how this may vary between genomic regions subject to different selective pressures.

Genome-wide methods for detecting selection
SMBE-LBA-049
GENOMEGAMAP: WITHIN-SPECIES GENOME-WIDE DN/DS ESTIMATION FROM OVER 10,000 GENOMES
D. J. Wilson* and The CRyPTIC Consortium

Poster Submission: The *dN*/*dS* ratio provides evidence of adaptation or functional constraint in protein-coding genes by quantifying the relative excess or deficit of amino acid-replacing versus silent nucleotide variation. Inexpensive sequencing promises a better understanding of parameters such as *dN*/*dS*, but analysing very large datasets poses a major statistical challenge. Here I introduce genomegaMap for estimating within-species genome-wide variation in *dN*/*dS*, and I apply it to 3,979 genes across 10,209 tuberculosis genomes to characterize the selection pressures shaping this global pathogen. GenomegaMap is a phylogeny-free method that addresses two major problems with existing approaches: (i) it is fast no matter how large the sample size and (ii) it is robust to recombination, which causes phylogenetic methods to report artefactual signals of adaptation. GenomegaMap uses population genetics theory to approximate the distribution of allele frequencies under general, parent-dependent mutation models. Coalescent simulations show that substitution parameters are well-estimated even when genomegaMap's simplifying assumption of independence among sites is violated. I demonstrate the ability of genomegaMap to detect genuine signatures of selection at antimicrobial resistance-conferring substitutions in *M. tuberculosis* and describe a novel signature of selection in the cold-shock DEAD-box protein A gene *deaD/csdA*. The genomegaMap approach helps accelerate the exploitation of big data for gaining new insights into evolution within species.

Genome-wide methods for detecting selection

SMBE-LBA-048

GENETIC EVIDENCE OF POLYGENIC ADAPTATION DURING BERINGIA STANDSTILL

T. Hünemeier[*], K. Nunes [1], C. M. Caouto-Silva [2], C. E. G. Amorim [3]

[1]University of São Paulo, [2]University of São Paulo, São Paulo, Brazil, [3]University of California, Los Angeles, United States

Poster Submission: The Beringian environment, although hostile, was an important population refuge during the Last Glacial Maximum. The low UV radiation and consequent reduction in the synthesis of vitamin D, acted as a strong local selective pressure. Recently, we have shown that variants of *FADS* (Fatty Acid Desaturase) genes were selected during Beringia Standstill. These variants perform better metabolism of fatty acids, favoring the diet rich in animal fat. A recent study suggests that not only the *FADS* genes, but also the *EDAR* V370A variant would have been selected in Beringia. According to this hypothesis, the *EDAR* V370A variant increases the number of mammary ductal branching, which would allow a greater distribution of nutrients via mothers' milk. The *FADS* genes in turn modulate the milk lipid profile, balancing vitamin D deficiency through an Omega-3 rich diet. To test the hypothesis that *EDAR* and *FADS* were selected together in Beringia, we analyzed 53 Native American populations, 19 Asian and 09 European populations. First, we verified the allelic interaction of SNPs of the *EDAR* and *FADS* genes, through the Custom Correlation Coefficient approach implemented in the BlocBuster software. We identified the interaction of *EDAR* (21 alleles) with *FADS* (12 alleles) only in Native American populations (p-value<0.05). Next, we simulate genetic data with *ms* under different possible demographic scenarios that mimic the prehistoric settlement of the American continent. Our results showed that in 14% of the simulations the allelic frequencies distribution is similar to that observed for the *EDAR V370A* (>90% in Native Americans, at least 15% higher than other continents), being therefore consistent with genetic drift. These results suggest that selection in the *EDAR* variant occurred prior to arrival in Beringia. However, its interaction with the *FADS* cluster, which was selected in Beringia, may have contributed to a local adaptive advantage. In this context, the strong drift on arrival at Beringia led to the fixation of the *EDAR* allele, which created a genetic background that favored that it increased the adaptive value of the individuals carrying the variants for the *FADS* genes.

Genomic perspectives on plant and animal domestication
SMBE-LBA-028
GENOMIC INSIGHTS INTO POST-DOMESTICATION DIVERSIFICATION OF DATE PALMS
J. Flowers*

Poster Submission: Date palms are an ancient dioecious perennial crop whose diversification history remains poorly understood. We present evidence that hybridization with a distant crop wild relative has been central to the diversification of date palm and find evidence that alleles for key diversification traits are introgressed from outside the date palm gene pool. These results are consistent with crop wild relatives of date palm being a source of novel alleles for diversification traits and suggest that introgressive hybridization may have been a key factor into the domestication history of date palm.

Inside Africa: Uncovering patterns of human genetic diversity

SMBE-LBA-036

GENETIC STRUCTURE AND ADMIXTURE OF THE COMORIAN POPULATIONS REDRAW HISTORICAL MIGRATION ROUTES INTO THE COMORO ARCHIPELAGO

H. M. A. Babiker [1,*], I. Walker [2], J. Krause [3], N. Boivin [4]

[1]Linguistic and Cultural Evolution, Max-Planck Institute for the Science of Human History, Jena, [2]Center for Interdisciplinary Area Studies (ZIRS), Martin Luther University Halle-Wittenberg, Halle, [3]Archaeogenetics, [4]Archaeology, Max-Planck Institute for the Science of Human History, Jena, Germany

Poster Submission: The Indian Ocean had been a scene for human trading, movements and intermixing of populations from East Africa, Arabia, the Middle East, and Southeast Asia. The Comoro islands situated off the Southeast coast of Africa at the northern end of the Mozambique Channel are of great importance to study past events that have shaped the linguistic, genetic and cultural facets of the modern cosmopolitan society. In this study, we use genome-wide genotyping data to investigate the population structure of the Comorian populations across 31 villages from the three major islands of the Comoro Archipelago: Ngazidja, Ndzuani, and Mwali. Our results show that the genetic structure of the Comoros represents a sub-Saharan African genetic pool with gene flow from non-African populations. Further, we find significant signals of admixture from Southeast Asia estimated at ~800 - 900 y ago. In addition, we report Y-chromosome haplogroups distributions in the Comoros, which are indicative of African and non-African male migrations. These findings support previous evidence of intercontinental seafaring during the first centuries of the 2ndmillennium AD. Moreover, they show the impact of migration waves in framing the genetic and cultural portray of the Comoro islands and highlight their complex population dynamics.

Microbial Evolution in Complex Environments

SMBE-LBA-021

EXPERIMENTAL EVOLUTION OF SPECIES INTERACTIONS IN A MODEL MICROBIAL COMMUNITY FROM THE OCEAN

F. Gorter*, M. Ackermann [1 1]

[1]Environmental Microbiology, Eawag, Duebendorf, Switzerland

Poster Submission: Microbial communities, like all complex systems, are more than the sum of their parts: they are characterized by a multitude of interspecific interactions, which can range from mutualism to competition. The overall sign and strength of interspecific interactions have important consequences for the emergent properties at the community level, but it is currently unclear whether and how these interactions change over evolutionary time scales. I study the evolution of species interactions in a model microbial community from the ocean. This community plays an important role in global carbon cycling, and as such has broad direct relevance. Extending from previous theory, I developed a set of specific hypotheses for the evolution of interspecies interactions in such complex systems. To address these hypotheses, I will experimentally evolve communities of five bacterial species in either well-mixed or spatially structured environments, and investigate how each of the constituent species, as well as their interactions, changes over time using a combination of phenotypic assays, whole-genome sequencing, and metabolic profiling. Moreover, I will investigate whether adaptations in a community context are typically selected because of their direct or indirect fitness effects by studying the growth of single cells in precisely controlled environments, where their effect on the—local and global—environment can be manipulated. I anticipate that evolution in well-mixed environments will proceed exclusively via the selection of directly beneficial traits. Conversely, evolution in spatially structured environments may proceed via the selection of both directly and indirectly beneficial traits, and which of these two predominates should depend on the availability and effect size of mutations affecting both types of traits. Selection of indirectly beneficial traits is predicted to result in an increase in interaction strength over time, while selection of directly beneficial traits is not predicted to have such a systematic effect. My work will be the first to directly and systematically address these novel hypotheses, and as such, is expected to provide important general insights into how microbial communities evolve.

Microbial Evolution in Complex Environments
SMBE-LBA-056
THE EFFECTS OF HOST GENETIC DIVERSITY ON PARASITE EVOLUTION
A. Ekroth*

Poster Submission: The 'monoculture effect' is when genetically homogenous populations are more vulnerable to outbreaks of disease. Such a phenomenon could be predicted to shape parasite evolution such that genetically heterogenous host populations would impose stronger selection on parasites to adapt. Wetest this idea by experimentally evolving the parasitic bacteria, *Staphylococcous aureus,* in genetically heterogenous and homogenous host populations of wild isolates of the nematode *Caenorhabditis elegans.* When comparing gut colonizations of ancestral to evolved *S. aureus*, we find great variation in infectivity between replicate homogenous host populations to heterogenous ones. We also conduct a meta-analysis to directly test the biological conditions under which host genetic diversity limits disease spread. Overall, we find broad support for the monoculture effect across host species. The effect was independent of host-parasite specialisation (genotypic or species-level), parasite type and diversity, virulence, and experimental environment. Together, these studies highlight the impact of host genetic diversity as a driver of parasite evolution and ecology with relevance to small and at-risk host populations.

Microbial Evolution in Complex Environments

SMBE-LBA-043

PLANT PATHOGENIC BACTERIA CAN RAPIDLY EVOLVE TOLERANCE TO PLANT CHEMICAL-BASED BIOFUMIGATION BIOCONTROL STRATEGY

C. Alderley*, V. Friman [1]

[1]Biology, University of York, York, United Kingdom

Poster Submission: Crop losses to pesticide-resistant pathogens are a growing threat for global food security and novel, more effective control strategies are urgently required. Biofumigation, an agricultural technique involving the mulching of *Brassica* plant tissues into soils to release toxic isothiocyanates (ITCs), has been proposed as an environmentally-sustainable alternative to agrochemicals. While biofumigation has been shown to be effective against a range of different pests, its effects against pathogenic bacteria, and the ability for these bacteria to evolve tolerance to biofumigation, have been largely unexplored. Here we used a laboratory model system to compare the efficacy of different types of ITCs and potential ITC-tolerance evolution of *Ralstonia solanacearum*, a plant pathogenic bacterium. To explore potential resistance evolution to allyl-ITC we conducted a separate three-week long selection experiment under different ITC exposure regimes (24h, 48h and 72h exposure periods). We found that of all tested ITCs, only allyl-ITC was efficient at suppressing *R. solanacearum*. We found that ITCs were effective at suppressing pathogenic growth in all exposure regimes, but that increased ITC tolerance was observed only in the 72h exposure regime. These results could be explained by relatively higher mutation supply rates. Alternatively, bacteria might have been able to acquire certain compensatory mutations more easily in 72h exposure treatment due to relatively longer recovery periods as ITCs were volatilised in less than 48 hours. Together, these results suggest that repeated exposure to allyl ITC during biofumigation could select for more tolerant bacterial pathogens, potentially weakening the long-term efficiency of this biocontrol strategy.

Mitochondrial-Nuclear Interactions

SMBE-LBA-018

MITONUCLEAR COEVOLUTION: IMPLICATIONS (OR LACK THEREOF) FOR MOTHER'S CURSE, DOUBLY UNIPARENTAL INHERITANCE, AND SPECIATION

J. Havird*, H. McConie, J. Han, S. Rabinowitz, K. Thueson, N. Crouch, F. Ghiselli

Poster Submission: Eukaryotes are a chimeric lineage and as such, most eukaryotes possess multiple genomes within their cells. At a minimum, nuclear and mitochondrial genomes must function properly together to maintain organismal integrity in most eukaryotes. Mitonuclear coevolution, the notion that sequence evolution within one genome exerts selection for complementary changes in the other genome, has been proposed to be ubiquitous in eukaryotes. Such coevolution has implications for diverse biological fields. However, hypotheses stemming from these ideas are just beginning to be tested using the tools of molecular evolution. We examined several such hypotheses related to three broad topics: 1) mother's curse – the hypothesis that male-harming mutations will accumulate in maternally-inherited mitochondrial DNA due to a sex-specific selective "sieve", 2) doubly uniparental inheritance – the observation that in many bivalves mtDNA is inherited from both parents, but remains sequestered in separate cell lineages, and 3) speciation – specifically, that coadapted mitonuclear genotypes within lineages leads to reproductive isolation among lineages. Using methods to detect selection (e.g., d_N/d_S ratios), spatial information from enzyme structures, and correlations between rates of speciation and mitochondrial molecular evolution, we find that many of the predictions derived from mitonuclear coevolution were not supported. This suggests that despite many case studies supporting the importance of mitonuclear coevolution, caution should be used when extending these observations across biology.

Mitochondrial-Nuclear Interactions

SMBE-LBA-070

IMPACT OF MITOCHONDRIAL PROTEOME ON THE EVOLUTION OF YEAST HYBRIDS

A. Manousaki [1,*], J. Boyd [2], S. Hubbard [1], M. White [1], D. Delneri [1]

[1]University of Manchester, [2]University o fManchester, Manchester, United Kingdom

Poster Submission: The interspecific hybrids of the Saccharomyces sensu stricto species inherit a bi-parental genome but retain the mitochondrial DNA of only one parent. In this environment, where two homologous proteomes co-exist, the potential to form chimeric protein complexes may confer improved properties to the hybrid strain. However, cases of hybrid defects and lethality have been previously attributed to cytonuclear incompatibilities, a phenomenon also associated with degenerative diseases in humans. In this study, we aim to investigate the role of mitochondria in the evolution of hybrid genomes by exploring how the different abundance of mitochondrial-associated proteins may influence the dynamics of the mitochondrial proteome in hybrids of S. cerevisiae with S. uvarum. To address this, we used the Fluorescence Correlation Spectroscopy (FCS) microscopy approach to first generate a reference set of absolute quantitative data for the low-abundant fission proteins, Fis1p and Mdv1p, in the S. cerevisiae parental strains growing under respiratory conditions. Analysis of our FCS data in the S. cerevisiae strains revealed the absolute concentrations of GFP-labelled Fis1p and Mdv1p to be of ~88-110 molecules per cell. Moreover, FCS measurements on the Fis1p mobility provide evidence that the protein is mainly found free in the cytoplasm of S. cerevisiae x S. uvarum hybrid cells growing in respiratory conditions. Further comparison of the FCS-based molecule concentrations to the data generated in the hybrid environment with either S. cerevisiae or S. uvarum mitotype will provide us an absolute quantification of single-cell protein-protein interactions. This will help us better understand how protein dynamics change throughout cellular functions and further enhance ongoing attempts to model protein pathways and define global protein abundance in yeast.

Mitochondrial-Nuclear Interactions

SMBE-LBA-027

DNA REPAIR AND MUTATION RATES IN PLANT MITOCHONDRIA

A. C. Christensen [1,*], E. L. Wynn [1]

[1]Biological Sciences, University of Nebraska, Lincoln, NE, United States

Poster Submission: Synonymous substitution rates in plant mitochondrial genes are very low, in spite of evidence for their near-neutrality. Efficient and extensive double-strand break repair (DSBR) has been proposed as an explanation for both the low substitution rates (because repair is template-directed) and the abundant rearrangements and expansions (because DSBR can lead to crossovers and break-induced replication). Interestingly, the DNA replication and repair enzymes are all encoded in the nucleus. Plant mitochondria possess the base-excision repair pathway for removal of uracil. In order to assess how important that pathway is, we maintained by single-seed descent several independent lines of *Arabidopsis thaliana* deficient in uracil-N-glycosylase (UNG) for 10 generations to determine the repair outcomes when the pathway is missing. Surprisingly, no single nucleotide polymorphisms (SNPs) were fixed in any line by generation 10. There were heteroplasmic SNPs in each line, but no allele frequencies were greater than 20%, and none were shared between lines. Plants from generation 11 were examined to see if the heteroplasmies were inherited. Clearly DNA maintenance in meristem mitochondria is very effective in the absence of UNG. These results indicate that double strand break repair is a general system of repair in plant mitochondria, and that under laboratory growth conditions the UNG pathway is dispensable.

Molecular basis of neural circuit and behavioral evolution

SMBE-LBA-058

GENOMIC VARIABILITY AND NEUROLOGICAL DISEASES: A FOCUS ON SOUTHERN ITALY

P. Abondio [1,*], A. De Giovanni [1], S. Sarno [1], C. Giuliani [1 2], D. Luiselli [3]

[1]Department of Biological Geological and Environmental Sciences, University of Bologna, Bologna, Italy, [2]School of Anthropology and Museum Ethnography, University of Oxford, Oxford, United Kingdom, [3]Department of Cultural Heritage, University of Bologna - Ravenna Campus, Ravenna, Italy

Poster Submission: Recent studies on more than 700 individuals (Sazzini *et al*. 2016) have highlighted an internal structure to the Italian population, with four groups (Northern Italy, Central Italy, Southern Italy and Sardinia) clearly characterized by genetic differences presumably related to diverse selective pressures (in particular, colder climate and diet in Northern Italy as opposed to a long-term exposure to pathogens in Southern Italy), as well as different demographic histories. Furthermore, a Gene Ontology enrichment analysis on the most differentiated genes among Northern and Southern Italy has emphasized processes of cell/neuron recognition, as well as cellular compartments such as lipid rafts, that are involved in the development of neurological and cardiovascular diseases (Michel and Bakovic, 2007).

A recent review on the role of apolipoprotein E in longevity (Abondio *et al*. 2019) has presented previously unpublished data on the Italian population that again highlight a clear gradient of the *APOE* variants along the peninsula, with Southern Italy clearly distinguished from Northern Italy.

Moreover, characteristic variants of several genes already known to be involved in neurodegenerative diseases have been found in individuals of Southern Italian origin, but not in Northern Italy (Conidi *et al*. 2015; Capozzo *et al*. 2017; Gagliardi *et al*. 2018).

These observations suggest that it could be worthwhile to investigate the diffusion, differentiation and potential differences in genotype association and haplotype distribution along the Italian peninsula for several genes related to neurodegenerative diseases like Alzheimer's and frontotemporal dementia, which are becoming more and more impactful in the ever aging Italian population, particularly in relation to their medical and social implications.

We plan to conduct this investigation through a Northern Italy-Southern Italy comparative approach that includes population structure analysis, as well as statistical analysis to detect potential signatures of selection and a network analysis approach to evaluate the involvement and influence of these genes on the pathways they belong to.

We also aim at evaluating instances of co-occurrence of genotypes and haplotypes in different genes involved in neurological disorders, in order to assess the potential influence of multiple disease-enhancing or disease-protecting variants on the incidence of neurodegenerative illnesses.

Mutation Rate Evolution
SMBE-LBA-037
RATE HETEROGENEITY AND DATA QUALITY INFLUENCE ON MOLECULAR DATING.
G. Louvel [1,*], H. Roest Crollius [1]
[1]Institut de Biologie de l'ENS (IBENS), Paris, France

Poster Submission: DNA sequences of modern day organisms are an insightful document of evolutionary history, to quote Zuckerkandl and Pauling, and this seminal observation has led to a broad variety of inference methods: computational models allow us to reconstruct phylogenetic species trees, infer selective pressures, or date past events. The field of molecular dating tackles questions regarding the divergence dates of species, or even genes. In return, it provides data required for subsequent analyses, such as studying the rates of species birth and death, or the timescale of gene duplications in a lineage.

Such projects are challenging because of the uncertainty in age estimates. Therefore, our first objective is to develop an empirical analysis of the confidence intervals of age estimates in gene trees: using our dating pipeline, we estimated the ages of known speciation events, in each primate gene tree. The speciation control dates are taken from a consensus of fossil-calibrated molecular studies, available in the TimeTree database. With these control ages, we compute a measure of the dating error in each gene tree, and therefore obtain its empirical distribution. As a direct follow-up, we investigated which features of the input data degrade the estimation. This is of interest to disentangle biological and methodological pitfalls in molecular dating, and might allow us to predict the confidence on new data.

The dating pipeline can be summarised into these main steps: we first take reconstructed gene trees and coding sequence alignments from the Ensembl database. Those trees are "reconciled" with the species tree, meaning we can distinguish duplication nodes from speciation nodes. Prior to anything, this input data is filtered to eliminate obvious errors such as gene splits or aberrant branch lengths. We optionally applied alignment cleaning methods such as Gblocks. We estimate the amount of synonymous substitutions, by using codeml from PAML (Yang 2007). Then we apply molecular clock models to convert those molecular distances into absolute durations in million years. The first of these methods is the fast and scalable Mean-Path-Length (Britton et al. 2002) which smoothes clock rates by averaging sister branch lengths. We also apply penalised likehood estimation of constant rate, correlated rate, and relaxed rate models (Sanderson 2002).

To correlate with these errors, we documented many features from the input data, such as alignment properties (length, percentage of gaps, etc.) or tree properties (total tree length, heterogeneity of rates, etc.). We find the heterogeneity of rates and the alignment length to be major factors of dating error. Next we determine what characteristics of the data are a good proxy for rate heterogeneity. We also currently work on incorporating features specific to the studied node position in the tree, such as tree imbalance.

Mutation Rate Evolution

SMBE-LBA-069

PLASTICITY OF GENETIC INTERACTIONS BETWEEN NON-CODING RNA MUTANTS

M. Fraczek [1,*], S. Parker [1], K. Dungrattanalert [1], D. Estranda-Rivadeneyra [1], R. Alves de Almeida [1], S. Griffiths-Jones [1], R. O'Keefe [2], D. Delneri [1]

[1]University of Manchester, [2]Univesrity of Manchester, Manchester, United Kingdom

Poster Submission: Background: Previously, we have performed a large scale phenotypic screening of the haploid non-coding RNA (ncRNA) deletion collection in a single condition (YPD, 30) and observed significant variations in fitness. Here, we analyse the phenotypic variation in different environments of double ncRNA mutants generated using Synthetic Genetic Array (SGA).Aims: Screening of the double knock-out (KO) haploid library in various stress conditions to infer genetic interactions between ncRNAs and their functional plasticity.Methods: Double KO mutants were generated using SGA by crossing a query strain (carrying single gene deletion) with the ncRNA KO collection. Generated double KO mutants have been screened in various conditions, including for example oxidative, temperature and respiratory stresses, and compared to the standard condition (YPD, 30), and to the WT.Results: In total, 27 query strains have been crossed with the library of ~450 single ncRNA mutants. Several gene interactions have been discovered in the standard condition showing either loss or gain in fitness, such as for example ΔSUT193-ΔSUT055 or ΔSNR13-ΔSUT347. Phenotypic analysis of double KOs in stress conditions revealed significant variations in fitness, such as for example ΔtI(AAU)P2-ΔSUT211 being lethal in oxidative stress but not in other tested conditions.Conclusions: The SGA analysis on a small number of query strains showed a number of positive and negative genetic interactions, with a few being lethal. Phenotypic analysis of double mutants will help us draw a map of genetic interactions between ncRNAs and understand their function in different conditions.

Mutation Rate Evolution

SMBE-LBA-003

SHORT-RANGE TEMPLATE SWITCHING EXPLAINS MUTATION CLUSTERS IN THE HUMAN GENOME

C. Walker [1 2,*], N. De Maio [2], N. Goldman [2]

[1]Department of Genetics, University of Cambridge, Cambridge, [2]European Molecular Biology Laboratory, European Bioinformatics Institute (EMBL-EBI), Wellcome Genome Campus, Hinxton, United Kingdom

Poster Submission: A recurring feature observed in human genome sequencing data is a surprisingly high frequency of complex mutations composed of multiple nearby base substitutions, insertions and deletions. We use a generalized model of mutation, capturing local template-switching which occurs during DNA replication, to explain the role of template switches in the formation of mutation clusters. Under this model, short genome regions are replaced during replication by similar length fragments copied from a nearby location on the complementary strand.

Our model successfully detects thousands of unique template-switch events during the evolution of human, chimpanzee and gorilla from their most recent common ancestor. Thousands of events are also detectable in 2,504 de novo assembled human genomes produced as part of the 1,000 Genomes Project. These template-switch events appear to be associated with regions which can form complex DNA secondary structures and provide a single-step mechanism for the generation of more energetically stable structures. These events represent a prevalent class of small structural variants not currently well characterised or addressed by existing resequencing pipelines, which are therefore potentially prone to misassembly and misclassification.

Mutation Rate Evolution

SMBE-LBA-034

CORRELATED MOLECULAR AND DIVERSIFICATION RATES ACROSS THE PRIMATE RADIATION

L. Aristide [1,*], H. Morlon [1]

[1]Institut de Biologie de l'École Normale Supérieure (IBENS), École Normale Supérieure, Paris Sciences et Lettres (PSL) Research University, Paris, France

Poster Submission: According to modern evolutionary theory, changes occurring at the genetic, population, and biodiversity (formation of new species) levels are the result of a common evolutionary process. Consequently, it is expected that the rates at which these changes occur are coupled. In fact, a positive correlation between molecular evolution and diversification rates has been observed in some clades at broad phylogenetic scales. However, whether this is a common feature of macroevolutionary diversification is currently unknown.

Here, we studied the fine-scale (i.e. lineage level) phylogenetic association between diversification and molecular rates across the Primate radiation, a large mammalian order with more than 400 species. We estimated per-branch substitution rates using fossil calibrations for nuclear and mitochondrial alignments with a total length of ~60.000bp for 367 species. Additionally, we estimated per-branch speciation rates using a recently developed phylogenetic diversification model (ClaDS). Phylogenetic correlations among per-branch estimates of molecular and speciation rates revealed a complex pattern of associations. Intriguingly, while mitochondrial substitution rates showed a positive correlation with speciation, correlations were negative for nuclear rates. Moreover, within the mitochondrial alignment, the association with speciation was stronger in less neutral partitions (1st and 2nd codon positions and mtRNA) than in a more neutral one (3rd codon position). We propose possible mechanisms, involving the interplay between demographic and life history factors, that might help explain these contrasting patterns of associations between the diversification and molecular evolution rates. These results have interesting implications for understanding macroevolutionary diversification processes

Novel insights into evolutionary genetics from emerging technologies

SMBE-LBA-044

THE GENETIC AND DEVELOPMENTAL BASIS OF NATURAL PIGMENT PATTERN DIVERSITY IN DANIO KYATHIT

B. M. Mccluskey [1,*], D. M. Parichy [1], J. H. Postlethwait [2], S. Uji [3]

[1]Biology, University of Virginia, Charlottesville, VA, [2]Institute of Neuroscience, University of Oregon, Eugene, OR, United States, [3]Fisheries Research Agency, National Research Institute of Aquaculture, Minami-Ise, Japan

Poster Submission: Vertebrates display a striking diversity of pigment patterns, even between closely related species. In the cyprinid genus *Danio*, melanophores assemble into vertical bars, reticulated nets, arrayed spots, and the horizontal stripes for which zebrafish (*D. rerio*) is named. To understand the evolution of these patterns, we investigated the closely related species, *Danio kyathit*, which has both spotted and striped morphs.

To understand the genetic basis of this phenotypic diversity, we performed two mapping crosses segregating alleles from the spotted and striped morphs of *D. kyathit*. We developed an image quantification pipeline to robustly phenotype 648 individuals from these crosses. Using RAD-seq for a subset of these individuals, we genotyped over 30,000 loci and identified quantitative trait loci (QTL) for phenotypes at the cellular and organismal level. QTL analysis of the first cross identified a single pleiotropic locus associated with body size and pigment pattern (as measured by several metrics). QTL analysis of the second cross excluded individuals with intermediate pigment phenotypes and recovered several QTL explaining a combined 40% of the phenotypic variation between progeny closely resembling the spotted or striped pigment morph. The inferred QTL contain promising candidate genes known to affect pigmentation in zebrafish mutants. We also observed segregation distortion across several chromosomes, strongly suggesting that meiotic drive or hybrid incompatibilities are pushing the spotted and striped morphs of *Danio kyathit* to become distinct species.

Danios offer a remarkably tractable system for understanding the cellular and developmental basis of the pigment phenotypes from the QTL analyses. We leveraged these strengths to better understand how pigment pattern differences arise by imaging and quantifying pigment pattern development in striped and spotted *D. kyathit.* Comparisons between the two morphs revealed that they differ not only in their overall pigment pattern, but also in pigment cell size, number, density, and motility.

Novel insights into evolutionary genetics from emerging technologies
SMBE-LBA-063

APPLICATION OF THE HARMONIC MEAN P-VALUE TO MULTIPLE SEQUENCE ALIGNMENTS

J. Armstrong [1,*], S. Earle [1], D. Wilson [1]
[1]Big Data Institute, University of Oxford, Oxford, United Kingdom

Poster Submission: Genome-wide association studies (GWAS) investigate the association between population genetic variation and phenotype (for example, antimicrobial resistance or virulence in bacteria). Traditionally, Bonferroni correction is used in GWAS to control against false positives due to the large number of tests. However, the more conservative p-value correction reduces the power of the test and discourages the testing of novel variants. The Harmonic Mean P-Value is a recently proposed alternative method to Bonferroni correction which improves statistical power while maintaining protection against false positives. We detail a novel method for applying the Harmonic Mean P-Value to multiple sequence alignments, combining p-values for all k-mers within a sliding window (of variable size) across the alignment, and calculating a single p-value for each region in a computationally efficient manner, while gracefully handling sequence gaps and disparate alignment depths. We verify the method in a secondary analysis of a case-control GWAS for fusidic-acid resistance in *Staphylococcus aureus*. This method is general and can be applied to a multi-sequence alignment of any depth and any window resolution, producing Bonferroni-like adjusted p-values without the associated penalties.

Novel insights into evolutionary genetics from emerging technologies
SMBE-LBA-029
TRACKING THE SOURCE OF GASTROENTERITIS USING MACHINE LEARNING
N. Arning[*], D. Wilson[1]
[1]Big Data Institute, University of Oxford, Oxford, United Kingdom

Poster Submission: Gastroenteritis is a food-borne disease accounting for an estimated 2.5 million cases each year in the United States, which is predominantly caused by the bacterium *Campylobacter jejuni*. Sources of infection for humans can be under-cooked meat, contact with animal faeces or environmental sources like contaminated drinking water. Attributing the source of a gastroenteritis outbreak is vital for public health regulations. Current methods depend on statistical inference through the comparison of Multi locus sequence typing between human samples and source samples. We have developed a Machine Learning based application for source attribution to outperform the existing methods and broaden the spectrum of viable input to whole genomes. We report an increase in accuracy and speed and present a protocol which can be used for whole genomes or parts thereof. Our results confirms the applicability of Machine learning in source attibution of infections using genomic data.

Open Symposium
SMBE-LBA-009
MODELLING STRUCTURAL CONSTRAINTS ON PROTEIN EVOLUTION1VIA SIDE-CHAIN CONFORMATIONAL STATES
U. Perron*

Poster Submission: Few models of sequence evolution incorporate parameters describing protein structure, despite its14high conservation, essential functional role and increasing availability. We present a structurally-aware15empirical substitution model for amino acid sequence evolution in which proteins are expressed using16an expanded alphabet that relays both amino acid identity and structural information. Each character17specifies an amino acid as well a rotamer state: the discrete geometric pattern of permitted side-chain18atomic positions. By assigning rotamer states in 251,194 protein structures and identifying 4,508,39019substitutions between closely related sequences, we generate a 55-state model that shows that the20evolutionary properties of amino acids depend strongly upon side-chain geometry. The model performs21as well as or better than traditional 20-state models for divergence time estimation, tree inference and22ancestral state reconstruction. We conclude that the concomitant evolution of sequence and structure is23a valuable source of phylogenetic information.

Open Symposium
SMBE-LBA-059
EVOLUTION INSIGHTS INTO A NON-CLASSICAL TYPE OF WING POLYPHENISM IN APHIDS
Z. Lu [1], S. Zhan[*]
[1]Institute of Plant Physiology and Ecology, Chinese Academy of Sciences, Shanghai, China

Poster Submission: The wing polyphenism of aphid is a classic example of environmental adaptation. Generally, environmental signals induce adult female aphids to produce two types of offsprings, i.e., wingless and winged aphids. Here, we found a rare kind of polyphenism in *Aphis spiraecola*, which reproduces abnormal wings for a considerable proportion at optimal conditions. We used a combined approach, including transcriptome sequencing, behavior assays, and in vivo genetics, to characterize the molecular basis of this rare morphs in aphids.

Open Symposium

SMBE-LBA-064

DETERMINING THE EXTENT TO WHICH THE MOLECULAR BASIS OF IMPLANTATION HAS BEEN CONSERVED ACROSS DIFFERENT EUTHERIAN MAMMALS

N. Forde [1,*], H. Tinning [2], A. Taylor [3], P. Mulhair [4], B. Constantinides [4], R. Sutton [2], G. Oikonomou [5], M. A. Velazquez [6], A. Treumann [7], M. J. O'Connell [8]

[1]University of Leeds, Leeds, United Kingdom, [2]LICAMM, [3]LICAMM, School of Biology, [4]School of Biology, University of Leeds, Leeds, [5]Institute of Infection and Global Health, University of Liverpool, Liverpool, [6]School of Natural and Environmental Sciences, [7]NUPPA, School of Medicine, Newcastle University, Newcastle , [8]School of Life Sciences, University of Nottingham, Nottingham, United Kingdom

Poster Submission: During the pre-implantation period of pregnancy in mammals, changes to the uterine endometrium are required (both at the transcriptional and protein level) to facilitate the endometrium becoming receptive to implantation. We know that the developing bovine conceptus (embryo and extraembryonic membranes) produce proteins during this developmental stage. As this is a fundamental and shared process in early pregnancy across all placental mammals, we hypothesised that some of these conceptus derived proteins (specifically 30 that are detectable in vitro) may be highly conserved between different mammal species, and that they may share functionality in altering the transcriptome of the endometrial epithelial cells to facilitate uterine receptivity to implantation. Homologs for all 30 bovine conceptus derived proteins were identified across of 15 placental mammals, 2 marsupial mammals, 1 monotreme, 1 bird and 1 fish. High levels of sequence identity were detected for all 30 proteins with an average identity score of 84.2%. To determine how these proteins may alter endometrial function a recombinant bovine form of one of these proteins, CAPG (91% sequence identity between bovine and human), was produced and bovine (bEECs) and human endometrial epithelial cells (hEECs) were cultured for 24 hours with and without rbCAPG. The transcriptional response was determined by RNA sequencing and quantitative real-time PCR analysis respectively (Control, vehicle, CAPG 10, 100, 1000 ng/ml: n=3 biological replicates per treatment per species). Treatment of bEECs with CAPG resulted in changes to 1052 transcripts (629 increased and 423 decreased) compared to vehicle controls including those previously only identified as regulated by the pregnancy recognition signal. Treatment of hEECs with bovine CAPG increased expression of transcripts previously known to interact with CAPG in different systems (*CAPZB*, *CAPZA2*, *ADD1*and *ADK*) compared with vehicle controls ($P<0.05$). In conclusion, we have demonstrated that some proteins produced by the conceptus during the peri-implantation period of pregnancy are conserved in a number of mammal species. Ones of these proteins, CAPG, elicits a transcriptional response in both bovine and human endometrial epithelial cells *in vitro*. Future work will investigate the common transcripts and pathways this protein alters in different species to determine to what extent the molecular basis of implantation has been conserved across different eutherian mammals.

Open Symposium

SMBE-LBA-032

STAIRWAY PLOT 2: A FAST AND CONVENIENT TOOL FOR INFERRING DEMOGRAPHIC HISTORY USING FOLDED OR UNFOLDED SNP FREQUENCY SPECTRA

X. Liu [1,*]

[1]USF Genomics, University of South Florida, Tampa, United States

Poster Submission: Inferring demographic history using genetic information can shed light on important evolutionary events such as population bottleneck, expansion, migration, and admixture, among others. Recently we developed a model-flexible method called stairway plot (Liu and Fu, Nature Genetics, 2015), which infers detailed population size changes over time using SNP frequency spectra (SFS). This method can be applied to low-coverage sequence data, pooled sequence data or RAD-seq data from non-model organisms. Since then we have made several important improvements and released the version 2 of stairway plot. First, we have extended this method to be applicable to folded SFS, therefore removing the potential bias caused by incorrected inferred ancestral alleles. Using extensive simulation we showed that the accuracy of stairway plot with folded SFS is similar to that with correct unfolded SFS. Second, we have used an ensemble approach to control model overfitting so that the accuracy of the inference can be further improved. Third, the efficiency of algorithm has also been significantly improved with a speed increase of 10-30X, which enables it to be applicable to thousands of individuals. Last but not least, we have improved the usage convenience by providing a pipeline builder which enables a user to run the whole analysis with one command.

Open Symposium
SMBE-LBA-047
QUANTIFYING THE EFFECTS OF SPLICE-ALTERING MUTATIONS IN HOMININ DIVERGENCE AND ADMIXTURE
A. Seyedian [1], R. C. McCoy [1,*]
[1]Department of Biology, Johns Hopkins University, Baltimore, United States

Poster Submission: Genetic variation influencing pre-mRNA splicing is ubiquitous in human populations and constitutes a primary link to phenotypic variation and disease. Despite striking anecdotal examples, the role of splice-altering mutations in hominin divergence and admixture remains poorly characterized. Seeking a genome-wide perspective on splicing evolution, we leveraged multiple functional genomic datasets to characterize the splicing effects of substitutions among hominin lineages.

Mutations disrupting essential splice sites are relatively easily predicted, and hominin single nucleotide substitutions at 470 such sites were catalogued with the sequencing of the Neandertal and Denisovan genomes. Non-coding variants occurring outside of this context can also contribute to splicing disruption via a phenomenon termed “cryptic splicing”. Though more challenging to predict, cryptic splice-altering mutations are prevalent and can be strongly deleterious. We scored archaic and modern human-specific substitutions for cryptic splice predictions. We identified 279 single nucleotide changes that are predicted to disrupt splicing outside the context of annotated splice sites (SpliceAI $\Delta > 0.2$). These include 153 derived alleles that are specific to the archaic lineages, as well as 126 derived alleles that are specific to the modern human lineage. Notable high-confidence examples include a fixed Neandertal and Denisovan cryptic splice acceptor gain (SpliceAI $\Delta = 0.89$) in *OPHN1*, mutations in which cause intellectual disability and cerebellar hypoplasia.

In addition to mutations that alter splice donor or acceptor sites, splicing effects may arise by mutations that alter the binding *cis*-regulatory sequences by *trans*-acting splicing factors. We thus broadened our analysis by testing for associations between putative introgressed Neandertal sequences and patterns of splicing measured in GTEx data. Among whole blood, cerebellum, and testis tissues, we identified a total of 53 splicing quantitative trait loci (sQTL) where a Neandertal-introgressed variant was the top-scoring variant in *cis* (10% FDR). Top associations among this set included introgressed SNPs in *SLC24A4* (rs61977313; $P = 1.21\times10^{-11}$), a gene involved in hair, eye, and skin pigmentation, as well as *OAS1* (rs11066451; $P = 2.46\times10^{-13}$), *AKAP13* (rs4843090; $P = 2.04\times10^{-10}$) and *TLR1* (rs3924113; $P = 1.68\times10^{-9}$), genes with well-characterized roles in innate immune response. Moreover, *TLR1* and *OAS1* are known candidate of adaptive introgression with demonstrated splicing effects in previous studies of human immunity, thereby providing a positive control in support of our genome-wide approach.

Our study highlights the underappreciated role of splice-altering mutations in the functional genomic basis of hominin phenotypic divergence. Persisting archaic introgressed sequences contribute to both isoform diversity and quantitative variation in the splicing landscape of modern human genomes.

Open Symposium
SMBE-LBA-041
MITOGENOMICS OF AN EXTINCT COLONY OF SOUTHERN ELEPHANT SEALS
A. Berg [1], S. Ho [1], N. Rawlence [2], P. Faulkner [1], M. De Bruyn [1,*]
[1]University of Sydney, Camperdown, Australia, [2]The University of Otago, Dunedin, New Zealand

Poster Submission: The West Point midden site in Tasmania represents the highest latitude known for a putative breeding colony of the southern elephant seal, *Mirounga leonina*, as well as being the best body of evidence for the predation of southern elephant seals by Indigenous Australians. West Point represents a 1000+ year history of predation, and a substantive contribution to the diet of the Indigenous Tasmanians. This project will document late Holocene dynamics of genetic and demographic change in the southern elephant seal population in the context of climatic and human impacts. Analysis of the full mitogenomes of individuals found in the West Point midden will play a significant role in determining major factors contributing to the extirpation of the population. Full mitogenome sequencing can reveal the demographic history of the population and its relationships and interactions with other southern elephant seal breeding colonies. This project will estimate when and by inference, under what environmental conditions, the Tasmanian colony of seals separated from the broader population, in the context of Holocene environmental change and human pressures.

Origins, evolution and function of novel genes

SMBE-LBA-004

EXPLOSIVE IMMUNOGENETIC DIVERSIFICATION COINCIDENT WITH THE RISE OF MAMMALS

A. Dornburg [1,*], L. Roupe-Abrams [1], D. Wcisel [2], J. Yoder [2]

[1]NC Museum of Natural Sciences, [2]North Carolina State University, Raleigh, United States

Poster Submission: The evolutionary pressure to maintain an effective defense against rapidly evolving pathogens has armed vertebrates with a range of tactics from which to distinguish self from non-self. At the front line of this evolutionary arms race are cluster of differentiation (CD) cell surface molecules that include a wide range of both ligands and receptors vital to immune function. Among these, CD300s have been implicated as serving critical roles in human ailments that span psoriasis, autoimmune disorders, and cancer. While clearly of high importance to human health, comparative studies of CD300s have been limited. This lack of an evolutionary perspective obstructs translation of findings between model organisms and humans as well as developing testable hypotheses concerning the mechanisms that facilitate immunoreceptor diversification. Integrating a phylogenetic comparative framework with functional genomic approaches, we establish an evolutionary origin of CD300s. Further we show major evolutionary transitions in gene synteny and copy number variation, with a surge of gene and gene copy diversification coincident with the rise of placental mammals. Our work provides a new perspective on the evolution of this important component of vertebrate immunity, illuminating a close relationship between mammal life history, genomic architecture, and novel molecular innovations to pathogen resistance.

Origins, evolution and function of novel genes

SMBE-LBA-051

REPEATED EVOLUTION OF STANDALONE ZP-C DOMAINS IN NEMATODES VIA ZP-N DOMAIN LOSS

C. Weadick [1,*]

[1]Biosciences, University of Exeter, Exeter, United Kingdom

Poster Submission: Protein domains frequently co-occur as modular units within functionally integrated supra-domains. Co-occurrence can be so tight that the constituent domains are rarely found on their own in nature. Comprised of two domains (ZP-N and ZP-C), the Zona Pellucida Domain module (ZPD) typically functions as a facilitator of polymerization in secreted proteins, including those involved in key biological processes such as syngamy and cellular morphogenesis. The current model of ZPD biology argues that polymerization is mediated by ZP-N, whereas ZP-C regulates ZP-N activity, preventing premature polymerization. Consistent with this proposed model, ZP-N has been found in isolation, while ZP-C has not, but efforts to test this pattern using comparative data have been limited. Through phylogenetic analysis of 1783 modules from 59 nematode species, I uncovered unexpected evidence for independent ZP-C domains in nature, finding standalone ZP-C across three nematode ZPD subfamilies (reflecting at least two instances of ancient ZP-N loss). Codon model analysis indicated that these standalone ZP-D domains have evolved under strong selective constraint, and through homology modelling I uncovered evidence for divergent disulfide connectivity patterns. This study provides a unique perspective on the origins of molecular novelty: the escape of anciently-evolved domains from otherwise conserved supra-domains. This work represents the largest-scale study of ZP modules yet undertaken and sets the stage for comparative analyses conducted at the Kingdom-wide (Animalia) scale, as well as experimental exploration of ZP-C function in lab-tractable nematodes.

Origins, evolution and function of novel genes

SMBE-LBA-023

DIGITAL INVESTIGATIONS ON THE EVOLUTION OF PROKARYOTE PHOTOSYNTHESIS REGULATION

A. Pontes [1,*], N. Panchy [2]

[1]Michigan State University, East Lansing, [2]National Institute for Mathematical and Biological Synthesis, Knoxville, United States

Poster Submission: We demonstrate a new digital evolution platform based on a mechanistic model of gene regulation and chemical signaling, and discuss how this platform is being used to investigate the evolution of homeostasis and circadian rhythms in photosynthetic prokaryotes.

Gene regulatory networks have unique properties that make them a powerful model for evolutionary computation. However, their full potential as an open-ended and biologically-relevant representation has remained largely unfulfilled. This is likely due to the complexity of such models and their high computational cost. We tackled both issues by using a new platform, ELFA (Evolutionary Lab for Flexible Agents), which is based on a mechanistic model of gene regulation that is realistic and fast, to explore questions related to the evolution of homeostasis and other types of robust control systems in which biological organisms excel.

For our first experiment, we created the digital equivalent to a simple cyanobacterium cell and placed it in a simulated underwater environment where it is free to grow and reproduce, under the selective pressures of a limited-size population. Our environment contains a daily light cycle that follows a sinusoidal curve, providing an incentive for cells to manage energy usage by evolving gene regulation of functions such as photosynthesis, growth, and reproduction to take advantage of the light energy when it is available.

The representation ELFA uses includes both cis- and trans-regulatory elements, in the form of enhancers, transcription factors and basal promoters. It also includes mechanisms for protein-protein and protein-ligand interaction. The strength of any interaction depends on the affinity level between molecules and permits different degrees of agonism. Therefore, ELFA allows for complex networks of interaction that replicate some interesting features of natural gene-regulatory networks, such as cross-talk, differential gene activation, and combinatory logic.

Phylogenomics under the multispecies coalescent

SMBE-LBA-045

MULTI-SPECIES DEMOGRAPHIC MODELLING UNCOVERS PAST INTROGRESSION BETWEEN CURRENTLY ALLOPATRIC FRESHWATER SPECIES

S. L. Mendes [1,*], M. M. Coelho [1], V. C. Sousa [1]

[1]cE3c - Centre for Ecology, Evolution and Environmental Changes, Lisboa, Portugal

Poster Submission: The isolation of populations due to the emergence of barriers to gene flow has been associated with divergence processes, ultimately leading to the formation of new species. This process of allopatric divergence is particularly relevant in freshwater fish, where changes in the conformation of rivers and lakes can isolate populations over long periods of time. Population genomics data and multi-species coalescent modelling allows reconstructing the divergence process. Here, we used a Genotyping by Sequencing (GBS) approach to infer the evolutionary history of four endemic freshwater fish species in the western Iberian region: *Squalius carolitertii*, *S. pyrenaicus*, *S. torgalensis* and *S. aradensis*. We inferred the species tree using a dataset of 25,353 SNPs. Our results support a species tree with two main evolutionary lineages: one comprising *S. torgalensis* and *S. aradensis* and a second lineage of *S. carolitertii* and *S. pyrenaicus*. Within this second lineage, we found lower levels of genetic differentiation between the northern populations of *S. pyrenaicus* and *S. carolitertii* than between the northern and southern *S. pyrenaicus* populations. Demographic modelling based on the multi-population two-dimensional site frequency spectrum (2D-SFS) indicates that this genome-wide pattern is the result of introgression, with northern *S. pyrenaicus* receiving a contribution of approximately 80% from *S. carolitertii* into its genome. This supports that the speciation process involved periods of gene flow and was more complex than simply allopatric divergence and illustrates the advantage of using coalescent-based methods to infer the evolutionary history of related species.

Repeats and mobile elements as drivers of innovations in protein coding genes
SMBE-LBA-050
STRESS-INDUCED TRANSPOSON MOBILIZATION AND HERITABILITY OF DNA METHYLATION PATTERNS IN STRAWBERRY (FRAGARIA VESCA)
M. E. Lopez*, E. Bucher

Poster Submission: Stress-induced transposon mobilization and heritability of DNA methylation patterns in strawberry (*Fragaria vesca*)

Plants require adequate environmental conditions such as humidity, soil minerals, temperature, and sunlight to develop successfully. However, due to climate change the instability of those essential growth factors has become a limitation for agricultural production. For these reasons, the induction of genetic and epigenetic changes has become an increasingly important tool to accelerate the process of obtaining relevant and interesting traits more rapidly compared to conventional breeding techniques. Likewise, analyzing how plants develop survival strategies through epigenetic mechanisms can have ecological relevance for understanding adaptation and may have a significant commercial value for farmers in agriculture. Fragaria vesca has been chosen as a model plant for rosaceous and fruit crops to examine the transmission of phenotypic and epigenetic changes induced by stresses. Furthermore, F. vesca has the advantage of having a rapid life cycle (3.5 months), compact plant architecture, multiplication processes, abundant seed production, and small diploid genome. Through this project we want to answer the following three main questions: 1) Are DNA methylation changes inherited by “clonal” daughters and/or seedlings? Not all epigenetic modifications are stably transmitted to the next generations as most of them are reset during meiosis which may result in the loss of acquired traits. Because of that, we want to know if in strawberry plants, clone-daughter plants (mitotic inheritance) and seedlings (meiotic inheritance), we can detect transcriptomic, genetic, and epigenomic changes compared to mother plants. For that we will expose plants to a selection of abiotic stresses including shading, drought, heat, and cold conditions. Once these differences will have been identified, the corresponding phenotypic changes will be observed and analyzed in the following generations. 2) Are there heritable stress-induced transcriptomic and epigenetic changes in *F. vesca*? One ecologically relevant adverse circumstance that strawberry plants must deal with in the field are heat, cold, lack of water, light intensity and quality, salinity, etc. The perception of each condition by different tissues of the plant induces several developmental responses that when they have had an early perception of the stress they are able to respond by priming. 3) Do transposable elements contribute to the adaptation of plants to stress conditions? Transposable elements are mobile DNA sequences that can move into different parts of the genome. Their stress-induced mobilization may promote gene diversification, evolution, adaptation, and speciation, making them a powerful tool for plant breeding.

The Causes of parallel molecular evolution

SMBE-LBA-035

CONTRASTING SIGNATURES OF GENOMIC DIVERGENCE IN RAPIDLY SPECIATING NICARAGUAN CICHLID FISHES

A. Meyer [1,*], A. Kautt [2], C. Kratochwil [1], A. Nater [1]

[1]Biology, University of Konstanz, Konstanz, Germany, [2]Organismal and Evolutionary Biology, Harvard University, Cambridge, United States

Poster Submission: In the crater lakes in Nicaragua several small adaptive radiations of cichlid fishes have formed within a few hundred genearations in parallel. Similar phenotypes evolved repeatedly. We investigated the genomic basis of these parallel adaptations and the genomics of spciation. In the absence of geographic barriers, the dynamics of gene flow and recombination versus divergent selection determine if speciation can happen. The genetic architecture of traits under divergent selection has been proposed to affect these dynamics. Using 456 re-sequenced genomes from 19 populations within the young Nicaraguan Midas cichlid species complex (*Amphilophus cf. citrinellus*), we reveal the genomic bases of divergence along multiple phenotypic axes. Our results show that genomic differentiation associated with mono- or oligogenic traits exhibits very few prominent peaks but is otherwise shallow. Conversely, genomic differentiation is pronounced in divergence along multiple polygenic traits. Hence, we propose that simple trait architectures are not always as conducive to stable speciation-with-gene-flow as previously thought, whereas polygenic trait architectures can promote rapid speciation in sympatry.

The genetic architecture of polygenic adpatation: sweeps, small shifts and everything in between
SMBE-LBA-017

GENE NETWORK CHANGES ASSOCIATED WITH THE EVOLUTION OF PARASITOID RESISTANCE

B. Wertheim [1,*], L. Salazar-Jaramillo, K. M. Jalvingh, S. Gerritsma
[1]Groningen Institute for Evolutionary Life Sciences, University of Groningen, Groningen, Netherlands

Poster Submission: Complex traits are manifestations of intricate gene interaction networks, and evolution of complex traits revolves around the genetic variation in such networks. To better understand the evolution of complex traits, we study the immune response of *Drosophila* against parasitoids. Previously, we characterized the natural variation in immunity within and among *Drosophila* species, and mapped how the genome and transcriptome changed during the evolution of immunity in *Drosophila*. Here, we present the gene network changes associated with the gain, loss and increase of parasitoid resistance. Firstly, we present how an evolutionary innovation in the *melanogaster* species group, i.e. the acquisition of a novel type of differentiated hemocyte (i.e. insect blood cell), combines co-option of existing gene interaction networks, recent gene duplications and rapid divergence of 11 genes. We show that the secondary loss of the trait in *D. sechellia* is accompanied by loss-of-function mutations and drastic changes in expression throughout the gene network. We also show that the genetic basis of the evolution of increased parasitoid resistance within a population can differ substantially among populations. Moreover, the network changes that confer higher parasitoid resistance after experimental evolution may not be similar in different genetic backgrounds or in natural populations that evolve more gradually under a variety of selective forces. Combined, this emphasizes that short- and long-term evolutionary responses can differ markedly, which may reflect that evolution exploits genetic variation within different parts of the gene interaction network.

The molecular basis of major transitions in evolution
SMBE-LBA-060

THE METAZOAN MICRORNA COMPLEMENT

B. Fromm*, D. Domanska, E. Hoye, M. Hackenberg, A. Mathelier, W. Kang, E. Aparicio-Puerta, M. Johansen, E. Hovig, K. Flatmark, M. R. Friedländer, K. J. Peterson

Poster Submission: Non-coding RNAs (ncRNA), a significant part of the increasingly popular 'dark matter' of the human genome, have gained substantial attention due to their involvement in animal development, including diseases and cancer. Among the many different types of regulatory short noncoding RNAs, microRNAs (miRNAs) are the only class with individual gene sequences conserved across the animal kingdom. Accordingly, absence / presence matrices of miRNAs have been successfully applied as phylogenetic and taxonomic markers, and were also used to ascertain clade specificity of highly derived organisms. Utilizing a set of unique features, *bona fide* miRNAs can clearly be distinguished from the myriad of small RNAs generated in eukaryotic cells. Unfortunately, recognition and utilization of these clear and mechanistically well-understood criteria has not been a common practice and thus many published miRNA complements have considerable issues with respect to quality and completeness. For example, missing miRNA families have frequently been misinterpreted as secondary losses, and the conservation and the utility of miRNAs as phylogenetic markers has been questioned.

We have addressed this by manually reannotating the full miRNA complements of 43 animal species representing the majority of metazoan groups. We employed more than 500 small-RNAseq datasets of different organs, tissues and cell-types to arrive at more than 10,000 manually curated metazoan miRNA entries in total that are summarized in our database MirGeneDB2.0.

We show that metazoan miRNA complements are very homogenous between closely related species, with rare observed losses, and conduct evolutionary analyses that elucidate the phylogenetic origin of miRNA families as key determinant for the expression of miRNA genes in all animals.

MirGeneDB2.0 represents a robust platform for providing deeper and more significant insights into the biology of miRNAs, their roles in development and their contribution to major transitions in the evolution of metazoans.

The molecular basis of major transitions in evolution

SMBE-LBA-024

GENOME EVOLUTION AFTER WHOLE GENOME DUPLICATION IN 32 BRASSICALES SPECIES: GHOSTS OF POLYPLOIDY PAST

T. V. Kent [1,*], A. E. Platts [2], J. Y. Choi [2], P. P. Edger [3], T. Zhao [4], F. Maumus [5], J. C. Pires [6], M. Purugganan [2], R. A. Wing [7 8], M. E. Schranz [4], S. Shu [9 10], J. Schmutz [11], D. Weigel [12], T. Mitchell-Olds [13], E. A. Pilon-Smits [14], I. Al-Shehbaz [15], J. Hall [16 16], M. A. Koch [17], A. E. Pepper [18], S. I. Wright [1] and Brassicales Map Alignment Project (BMAP) Consortium

[1]Ecology and Evolutionary Biology, University of Toronto, Toronto, Canada, [2]Center for Genomics and Systems Biology, New York University, New York, [3]Department of Horticulture, Michigan State University, East Lansing, United States, [4]Biosystematics Group, Wageningen University and Research, Wageningen, Netherlands, [5]Institut National de la Recherche Agronomique, Paris, France, [6]Division of Biological Sciences, University of Missouri, Columbia, United States, [7]Biological and Environmental Science and Engineering Division, King Abdullah University of Science and Technology, Thuwal, Saudi Arabia, [8]School of Plant Sciences, Ecology & Evolutionary Biology, Arizona Genomics Institute, University of Arizona, Tucson, [9]Lawrence Berkeley National Laboratory, Berkeley, [10]Department of Energy, Joint Genome Institute, Walnut Creek, [11]HudsonAlpha Institute for Biotechnology, Huntsville, United States, [12]Department of Molecular Biology, Max Planck Institute for Developmental Biology, Tuebingen, Germany, [13]Department of Biology, Duke University, Durham, [14]Department of Biology, Colorado State University, Fort Collins, [15]Missouri Botanical Garden, St. Louis, United States, [16]Biological Sciences, University of Alberta, Edmonton, Canada, [17]Centre for Organismal Studies, Heidelberg University, Heidelberg, Germany, [18]Texas A&M University, College Station, United States

Poster Submission: In plants, a common mechanism through which genome size is drastically altered and genome evolution is rapidly altered is whole genome duplication (WGD), or polyploidization. WGDs are strikingly common transitions, having occurred sometime in the histories of all flowering plant species, and often re-occur within single lineages. While a WGD will result in an instant doubling of genome size, both the subsequent rediploidization process and the potential for an increase in transposable element (TE) activity after the event means that past WGDs can be cryptic in current diploid lineages, and can potentially cause substantial variation in genome size. Genomic evidence for the role of WGDs in driving genome size variation and an increase in TE activity, however, has been limited by the lack of comparative genomic datasets which include variation in WGD history. In order to better understand the role of WGD in genome evolution, we sequenced and assembled 19 new genomes in the Brassicales, doubling the current available genomes for the family, to provide 32 species containing 9 independent WGD events. Using this comparative genomic data, we show that 81% of the variation in genome size across the family can be explained by past WGD events and TE content. We further show that WGDs caused bursts of TE activity, and largely explain current TE abundance, with evidence for relaxed selection after WGD events leading to genome wide relaxed selection against TEs inserting near conserved regions. Finally, we provide evidence for both gain and loss of conserved noncoding elements following WGD, consistent with relaxed genome wide selection and functional redundancy driving the evolution of normally static sequence.

The molecular basis of major transitions in evolution

SMBE-LBA-002

MORE THAN ONE-TO-FOUR VIA 2R: EVIDENCE FOR INDEPENDENT AMPHIOXUS EXPANSION AND TWO-GENE ANCESTRAL STATE IN VERTEBRATES FOR CHORDATE MYOD-RELATED MRFS

M. E. Aase-Remedios [1,*], C. Coll-Lladó [1], D. E. K. Ferrier [1]

[1]Scottish Oceans Institute, Gatty Marine Laboratory, University of St Andrews, St Andrews, United Kingdom

Poster Submission: MyoD (Myogenic differentiation) has long been recognized as a master developmental control gene and a key element of the bilaterian developmental toolkit. The MyoD family of bHLH transcription factors (Myogenic regulatory factors, MRFs) drives myogenesis across the bilaterians, though these comparisons are complicated by multiple instances of gene duplication and loss in several lineages. Following duplications, for instance the two rounds of whole genome duplication (2R WGD) at the origin of the vertebrates, duplicate regulatory genes like the MRFs often subfunctionalise, whereby the function of the ancestral gene is partitioned amongst the daughter genes, a process which is frequently focused on the complex regulatory regions characteristic of developmental transcription factors. Subfunctionalisation has been well-documented for MRFs in the vertebrates, where *MyoD* and *Myf5* act early in myogenic determination while *Myog* and *Myf6* are expressed later, in differentiating myoblasts. Comparing chordate MRFs, we find an independent expansion of MRFs in the invertebrate chordate amphioxus, with evidence for a parallel instance of subfunctionalisation relative to that of vertebrates. Conserved synteny amongst chordate MRF loci supports the 2R WGD events as a major force in shaping the evolution of vertebrate MRFs. We also resolve vertebrate MRF complements and organization and infer an ancestral two-gene state in the vertebrates which corresponds to the creation of early- and late-acting types of MRFs. This necessitates a revision of previous conclusions about the simple one-to-four origin of vertebrate MRFs.

The molecular basis of major transitions in evolution
SMBE-LBA-010
GENOMIC SIGNATURES OF PHOTOSYNTHETIC TRANSITIONS WITHIN A GRASS SPECIES
M. Bianconi [1,*], E. Curran [1], J. Olofsson [1], L. Dunning [1], P.-A. Christin [1]
[1]Animal and Plant Sciences, University of Sheffield, Sheffield, United Kingdom

Poster Submission: The genomic signatures of adaptive physiological transitions are often blurred by evolutionary processes that take place over millions of years. This limits our understanding of the order of events and the underlying evolutionary processes that led to the evolution of a particular physiological trait. The grass *Alloteropsis semialata* provides an outstanding system to address such questions, as it encompasses populations that recently evolved the derived C_4 photosynthetic metabolism as well as non-C_4 populations. Here we use whole genome resequencing data of individuals sampled across the geographical and photosynthetic range of *A. semialata*, and the recently published chromosome-level genome assembly of this species to investigate the evolutionary processes that led to the photosynthetic divergence in the group. We estimate the distribution of fitness effects of mutations in several genomic regions, and test whether loci known to be involved in C_4 photosynthesis have evolved under selection. Our study sheds new light onto the adaptive genetic changes underlying the evolution of a complex physiological trait in plants.

The molecular basis of major transitions in evolution

SMBE-LBA-031

THE BACTERIAL ORIGINS OF CHEMICAL SIGNALLING IN HUMANS

P. Taylor [1,*], J. Miles [1], C. Ali [1], F. Ly [1], F. Cullinane [1], D. Wang [1], C. Perry [2], A. Millard [3], D. Scanlan [4]

[1]Astbury Centre for Structural Molecular Biology, University of Leeds, Leeds, [2]Department of Chemistry, University of Warwick, Coventry, [3]Department of Genetics & Genome Biology, University of Leicester, Leicester, [4]School of Life Sciences, University of Warwick, Coventry, United Kingdom

Poster Submission: We have previously postulated that retinoic acid is not of animal provenance, as previously thought, but that this key component of Vitamin A may be a product of bacterial origin and present in humans through a Lateral (or Horizontal) Gene Transfer (LGT) event of aldehyde dehydrogenase and CYP120 from cyanobacteria to animals. Lateral Gene Transfers between cyanobacteria and animals, in particular of cyclooxygenases**,** calpains and WD40 domains have been suggested by other authors, indicating that these events are likely to have shaped processes across the animal kingdom.

We are now commencing more systematic identifcation of LGT from ancestral cyanobacteria, and other complex bacteria, to ancestral animals, using a BLASTp protocol to compare the human proteome with bacterial proteomes, followed by phylogenetic analysis of candidate LGT genes / proteins. To validate this approach, we applied the protocol to a set of proteins that Marchadier et al. defined as the minimum "toolkit" for calcium signalling in eukaryotes.

Pleasingly, the initial BLASTp screen identified calpains as a "hit", consistent with the existing phylogenetic analysis by Rawlings mentioned above, as well as other proteins involved in calcium regulated apoptosis, caspases, calmodulin dependent protein kinases (CAMK) and the master calcium regulator calmodulin. A further important calmodulin related enzyme, nitric oxide synthase, also emerged from the BLASTp screen.

Separate phylogenetic analyses of calmodulin, calmodulin dependent kinases, caspases and nitric oxide synthases showed that orthologous genes / proteins from cyanobacteria and other complex bacteria had more recent common ancestors with animals than would be expected from the "standard" species tree of life. This is consistent with the aforementioned literature results from ourselves and others on genes / proteins controlling other aspects of chemical signalling. Importantly, key amino acid residues or motifs seem to be particularly well conserved, for example the ligands for calcium in calmodulin, suggesting that biochemical mechanisms may be similar in the diverse species.

Our new results add to an emerging picture of surprisingly close links between chemical signalling pathways in animals, including humans, and complex bacteria, in particular cyanobacteria. Consistent with the literature, we propose that such incongruencies with "standard" species trees are most likely explained by Lateral Gene Transfer. Proposed LGT events from ancestral cyanobacteria to ancestral animals add new richness to our understanding of eukaryotic evolution. Furthermore, identification of putative human-like chemical signalling proteins in complex bacteria may provide novel opportunities to study biochemical pathways of significant biomedical importance.

The molecular basis of major transitions in evolution
SMBE-LBA-046
ORIGIN AND EVOLUTION OF STEROIDOGENESIS
Y. Hoshino[*], E. Gaucher[1]
[1]Biology, Georgia State University, Atlanta, United States

Poster Submission: Isoprenoids (or terpenoids) represent one of the largest groups of organic compounds in nature and are distributed in the three domains of life. Among isoprenoids, steroids are universally found in eukaryotes and are integral components of eukaryotic membranes used to regulate cell rigidity and fluidity. Even though a common origin of all steroid biosynthesis pathways (steroidogenesis) has been inferred, its evolutionary history is still enigmatic. Our current study provides insight into the early evolution of steroid biosynthesis using the technique of Ancestral Sequence Reconstruction (ASR). ASR enables us to investigate the physical and chemical properties of ancestral proteins that could have existed at the early stage of steroidogenesis evolution. We characterized several ancestral proteins in terms of their enzymatic products and discuss the potential evolutionary trajectory from bacterial non-steroidal homologs (hopanoids) towards eukaryotic steroids. Since steroids are inferred to be present in the common ancestor of eukaryotes, the evolution of eukaryotes is closely linked to that of steroids. Our current study provides a molecular basis for a major transition that occurred in the evolution of life – emergence of eukaryotes.

Understanding the genomics of climate change response

SMBE-LBA-055

THE GENOMIC SIGNATURES OF ADAPTATION ALONG A THERMAL GRADIENT IN A MARINE FISH

A. Jacobs[*], N. O. Therkildsen[1]

[1]Department of Natural Resources, Cornell University, Ithaca, United States

Poster Submission: Marine ectotherms show a higher vulnerability to climate change related warming than terrestrial ones. Yet, we still do not fully understand how marine species can adapt to different environmental conditions, such as differences in temperature, without strong barriers to gene flow. Understanding how marine species adapt to diverse environmental conditions will allow us to more accurately predict species and population responses to climate change.

Atlantic silversides (*Menidia menidia*) are an ecological textbook example of a species exhibiting clear local adaptation despite extensive connectivity across one of the world's steepest marine temperature gradients along the North American Atlantic Coast. Atlantic silversides display strong countergradient variation in a range of independent adaptive traits, including growth rate, with northern populations showing faster growth compared to southern populations. However, the genomic processes underlying this countergradient variation and the adaptation to different thermal regimes are not known to date.

Here, we use low-coverage whole-genome re-sequencing data from over 550 individuals sampled across eleven populations along this gradient in combination with a de novo reference genome and linkage maps to identify signatures of adaptive genomic divergence and uncover how selection, recombination, and demography shape the landscape of genomic variation in this ecological model species. We identify strong genomic signatures of local adaptation across the genome that are partially explained by differences in sea surface temperature variation across the year (i.e. seasonality). In particular, large-scale inversions seem to play a predominant role in local adaptation. The landscape of differentiation is furthermore shaped by strong differences in gene flow and selection. Overall, this study provides detailed insights into the processes and mechanisms underlying local adaptation of an ecological model species along a thermal gradient, enhancing our understanding of how marine species can adapt to different environments in the absence of barriers to gene flow.

Using Ancient DNA to Study Natural Selction: New Models and New Data

SMBE-LBA-054

GENEALOGICAL INFERENCE FROM THOUSANDS OF ANCIENT AND MODERN SAMPLES

A. W. Wohns [1,*], Y. Wong [1], G. McVean [1]

[1]Big Data Institute, Li Ka Shing Centre for Health Information and Discovery, University of Oxford, Oxford, United Kingdom

Poster Submission: The thousands of ancient human genomes published over the last decade have provided important insights into ancestral demography and patterns of selection. Using these genomes together with the millions of publicly available modern genomes in a genealogical inference framework, such as tsinfer, could indicate how ancient and modern samples are related. This would give a more complete understanding of the forces shaping human genetic diversity. The tsinfer algorithm uses relative age estimates for ancestral haplotypes to create genealogical topologies of contemporaneous samples; however, non-contemporaneous (ancient) genomes may be used if they are inserted as potential ancestral haplotypes at the correct relative age. Integration of ancient samples into tsinfer is thus dependent on estimating the age of nodes in an inferred genealogy.

In this work, we first develop, implement, and test an importance sampling-based method for estimating node ages conditional on tree sequence topologies. Second, we use these dated genealogies, combined with ancient samples, to improve estimates of mutation age. Finally, we infer genealogies where ancient haplotypes can serve as ancestors of modern samples.

Using simulations, we demonstrate that the accuracy of both node and mutation age estimates is improved with both increasing sample sizes, from n=100 to 10,000, and population-scaled mutation rate, as well as with the inclusion of greater numbers of ancient samples. We demonstrate that the resulting tree sequences contain rich information on genomic descent from ancient samples, with preliminary results from a time series dataset of ancient genomes.

Using Ancient DNA to Study Natural Selction: New Models and New Data
SMBE-LBA-052
FROM BONES TO BATCH EFFECTS: CIRCUMVENTING BIAS IN ADNA DATA
L. M. Cassidy [1,*]
[1]Smurfit Institute of Genetics, Trinity College Dublin, Dublin, Ireland

Poster Submission: Batch effects present one of the largest challenges to genomic research, where rapid technological improvement of both wet and dry lab methodologies prevents convergence on uniform processing pipelines. This is a particular concern for palaeogenomics, given that prehistoric genomes are a finite resource and currently published data may not be amenable to resampling and sequencing in the coming decades.

Despite this added impetus for establishing consistent protocols in ancient genomics, most crucially in the wetlab, the degraded nature of aDNA has produced the opposite effect - a diversity of work-around methods designed to increase endogenous yields and minimise the impact of post-mortem damage. Post-DNA extraction, these include a move from shotgun sequencing to capture approaches (most prominently SNP capture), full and partial UDG-treatments and selection for smaller fragment sizes.

Much variation also exists in bioinformatic pipelines, although from a conservation standpoint this is inconsequential as long as raw sequence data is retained and made publically available. However, different programs, parameters and filters chosen during read processing and genotype calling can work to exacerbate biases introduced in wet lab procedures. Conversely, when chosen correctly, bioinformatic pipelines can provide the opportunity to minimise these. We explore this potential here, with particular focus on various genotype inference methods. We use data produced following different wet lab methodologies but originating from the same ancient individuals to explore bias and recommend pipelines to allow better co-analysis of data from different research groups.

Using Ancient DNA to Study Natural Selction: New Models and New Data

SMBE-LBA-053

'DEVOLVING' HUMANS BY ENGINEERING ANCIENT GENES INTO THE MODERN GENOME USING THE CRISPR/CAS9 SYSTEM.

L. Balico [1,*], L. Tran [1], C. R. Gaucher [1], E. Slack [1], E. A. Gaucher [1]

[1]Georgia State University, Atlanta, United States

Poster Submission: Uric acid is an insoluble metabolic product of purine metabolism and high concentration of uric acid is associated with human diseases, like gout. Typically, uric acid is metabolized to 5-hydroxyisourate by the enzyme uricase. Most mammals have a functional uricase enzyme, but in the Lesser and Great Apes (including humans), uricase is a pseudogene, and pseudogenization occurred during Mid Miocene. Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR) associated protein 9 (CRISPR/Cas9) system has recently emerged as an efficient methodology for genome editing. We are exploited the CRISPR/Cas9 system to disrupt the native human uricase pseudogene in HepG2 cells and simultaneously replace it with the ancestral uricase gene most-recently functional in the human/mammalian lineage as previously demonstrated in our group. This work allows to investigate the hypothesis that deactivating the uricase gene may have provided a selective advantage for our ancestors.