Genomic Analysis of European *Drosophila melanogaster* Populations Reveals Longitudinal Structure, Continent-Wide Selection, and Previously Unknown DNA Viruses

Martin Kapun,*†¶§[1,2,3] Maite G. Barrón,†[1,4] Fabian Staubach,‡[1,5] Darren J. Obbard,[1,6] R. Axel W. Wiberg,[1,7,8] Jorge Vieira,[1,9,10] Clément Goubert,[1,11,12] Omar Rota-Stabelli,[1,13] Maaria Kankare,‡[1,14] María Bogaerts-Márquez,[1,4] Annabelle Haudry,[1,11] Lena Waidele,[1,5] Iryna Kozeretska,[1,15,16] Elena G. Pasyukova,[1,17] Volker Loeschcke,[1,18] Marta Pascual,[1,19,20] Cristina P. Vieira,[1,9,10] Svitlana Serga,[1,15] Catherine Montchamp-Moreau,[1,21] Jessica Abbott,[1,22] Patricia Gibert,[1,11] Damiano Porcelli,[1,23] Nico Posnien,[1,24] Alejandro Sánchez-Gracia,[1,19,20] Sonja Grath,[1,25] Élio Sucena,[1,26,27] Alan O. Bergland,‡[1,28] Maria Pilar Garcia Guerreiro,[1,29] Banu Sebnem Onder,[1,30] Eliza Argyridou,[1,25] Lain Guio,[1,4] Mads Fristrup Schou,[1,18,22] Bart Deplancke,[1,31] Cristina Vieira,[1,11] Michael G. Ritchie,[1,7] Bas J. Zwaan,[1,32] Eran Tauber,[1,33,34] Dorcas J. Orengo,[1,19,20] Eva Puerma,[1,19,20] Montserrat Aguadé,[1,19,20] Paul Schmidt,‡[1,35] John Parsch,[1,25] Andrea J. Betancourt,[1,36] Thomas Flatt,*†‡[1,2,3] and Josefa González*†‡[1,4]

[1]The European Drosophila Population Genomics Consortium (DrosEU)
[2]Department of Ecology and Evolution, University of Lausanne, Lausanne, Switzerland
[3]Department of Biology, University of Fribourg, Fribourg, Switzerland
[4]Institute of Evolutionary Biology, CSIC-Universitat Pompeu Fabra, Barcelona, Spain
[5]Department of Evolutionary Biology and Ecology, University of Freiburg, Freiburg, Germany
[6]Institute of Evolutionary Biology, University of Edinburgh, Edinburgh, United Kingdom
[7]Centre for Biological Diversity, School of Biology, University of St. Andrews, St Andrews, Scotland
[8]Department of Environmental Sciences, Zoological Institute, University of Basel, Basel, Switzerland
[9]Instituto de Biologia Molecular e Celular (IBMC), University of Porto, Porto, Portugal
[10]Instituto de Investigação e Inovação em Saúde (I3S), University of Porto, Porto, Portugal
[11]Laboratoire de Biométrie et Biologie Evolutive UMR 5558, CNRS, Université Lyon 1, Université de Lyon, Villeurbanne, France
[12]Department of Molecular Biology and Genetics, Cornell University, Ithaca, NY
[13]Research and Innovation Centre, Fondazione Edmund Mach, San Michele all' Adige, Italy
[14]Department of Biological and Environmental Science, University of Jyväskylä, Jyväskylä, Finland
[15]General and Medical Genetics Department, Taras Shevchenko National University of Kyiv, Kyiv, Ukraine
[16]State Institution National Antarctic Scientific Center of Ministry of Education and Science of Ukraine, Kyiv, Ukraine
[17]Laboratory of Genome Variation, Institute of Molecular Genetics of RAS, Moscow, Russia
[18]Department of Bioscience—Genetics, Ecology and Evolution, Aarhus University, Aarhus C, Denmark
[19]Departament de Genètica, Microbiologia i Estadística, Facultat de Biologia, Universitat de Barcelona, Barcelona, Spain
[20]Institut de Recerca de la Biodiversitat (IRBio), Universitat de Barcelona, Barcelona, Spain
[21]Université Paris-Saclay, CNRS, IRD, UMR Évolution, Génomes, Comportement et Écologie, 91198, Gif-sur-Yvette, France
[22]Section for Evolutionary Ecology, Department of Biology, Lund University, Lund, Sweden
[23]Department of Animal and Plant Sciences, Sheffield, United Kingdom
[24]Johann-Friedrich-Blumenbach-Institut für Zoologie und Anthropologie, Universität Göttingen, Göttingen, Germany
[25]Division of Evolutionary Biology, Faculty of Biology, Ludwig-Maximilians-Universität München, Planegg, Germany
[26]Instituto Gulbenkian de Ciência, Oeiras, Portugal
[27]Departamento de Biologia Animal, Faculdade de Ciências da Universidade de Lisboa, Lisboa, Portugal
[28]Department of Biology, University of Virginia, Charlottesville, VA
[29]Departament de Genètica i Microbiologia, Universitat Autònoma de Barcelona, Barcelona, Spain
[30]Department of Biology, Faculty of Science, Hacettepe University, Ankara, Turkey
[31]Institute of Bio-engineering, School of Life Sciences, EPFL, Lausanne, Switzerland

Mol. Biol. Evol. 37(9):2661–2678 doi:10.1093/molbev/msaa120 Advance Access publication May 15, 2020

[32]Laboratory of Genetics, Department of Plant Sciences, Wageningen University, Wageningen, Netherlands
[33]Department of Evolutionary and Environmental Biology, University of Haifa, Haifa, Israel
[34]Institute of Evolution, University of Haifa, Haifa, Israel
[35]Department of Biology, University of Pennsylvania, Philadelphia, PA
[36]Department of Evolution, Ecology, and Behaviour, University of Liverpool, Liverpool, United Kingdom
[†]These authors contributed equally to this work.
[‡]*Drosophila* Real Time Evolution (Dros-RTEC) Consortium.
[¶]Present address: Department of Evolutionary Biology and Environmental Sciences, University of Zürich, Zürich, Switzerland
[§]Present address: Division of Cell and Developmental Biology, Medical University of Vienna, Vienna, Austria
***Corresponding authors:** E-mails: martin.kapun@uzh.ch; thomas.flatt@unifr.ch; josefa.gonzalez@ibe.upf-csic.es.
Associate editor: Daniel Falush

Abstract

Genetic variation is the fuel of evolution, with standing genetic variation especially important for short-term evolution and local adaptation. To date, studies of spatiotemporal patterns of genetic variation in natural populations have been challenging, as comprehensive sampling is logistically difficult, and sequencing of entire populations costly. Here, we address these issues using a collaborative approach, sequencing 48 pooled population samples from 32 locations, and perform the first continent-wide genomic analysis of genetic variation in European *Drosophila melanogaster*. Our analyses uncover longitudinal population structure, provide evidence for continent-wide selective sweeps, identify candidate genes for local climate adaptation, and document clines in chromosomal inversion and transposable element frequencies. We also characterize variation among populations in the composition of the fly microbiome, and identify five new DNA viruses in our samples.

Key words: **population genomics, adaptation, demography, selection, clines, SNPs, structural variants.**

Introduction

Understanding processes that influence genetic variation in natural populations is fundamental to understanding the process of evolution (Dobzhansky 1970; Lewontin 1974; Kreitman 1983; Kimura 1984; Hudson et al. 1987; McDonald and Kreitman 1991; Adrian and Comeron 2013). Until recently, technological constraints have limited studies of natural genetic variation to small regions of the genome and small numbers of individuals. With the development of population genomics, we can now analyze patterns of genome-wide genetic variation for large numbers of individuals, with samples structured across space and time. As a result, we have new insight into the evolutionary dynamics of genetic variation in natural populations (Begun et al. 2007; Hohenlohe et al. 2010; Cheng et al. 2012; Pool et al. 2012; Harpur et al. 2014; Zanini et al. 2015). But, despite this technological progress, extensive large-scale sampling and genome sequencing of populations remains prohibitively expensive and too labor-intensive for most individual research groups.

Here, we present the first comprehensive, continent-wide genomic analysis of genetic variation of European *Drosophila melanogaster*, based on 48 pool-sequencing samples from 32 localities collected in 2014 (fig. 1) by the European *Drosophila* Population Genomics Consortium (*DrosEU*; https://droseu.net). *Drosophila melanogaster* offers several advantages for genomic studies of evolution in space and time. It boasts a relatively small genome, a broad geographic range, a multivoltine life history which allows sampling across generations on short timescales, simple standard techniques for collecting wild samples, and a well-developed context for population genomic analysis (Powell 1997; Keller 2007; Hales et al. 2015). Importantly, this species is studied by an extensive international research community, with a long history of developing shared resources (Hales et al. 2015; Bilder and Irvine 2017; Haudry et al. 2020).

Our study complements and extends previous studies of genetic variation in *D. melanogaster*, both from its native range in Southeastern Africa and from its world-wide expansion as a human commensal. The expansion into Europe is thought to have occurred ~4,100–19,000 years ago and into North America and Australia in the last few centuries (David and Capy 1988; Lachaise et al. 1988; Li and Stephan 2006; Keller 2007; Kapopoulou, Kapun, et al. 2018; Arguello et al. 2019; Sprengelmeyer et al. 2020). The colonization of novel habitats and climate zones on multiple continents makes *D. melanogaster* especially useful for studying parallel local adaptation, with previous studies finding pervasive latitudinal clines in allele frequencies (Schmidt and Paaby 2008; Turner et al. 2008; Kolaczkowski et al. 2011; Fabian et al. 2012; Bergland et al. 2014; Kapun, Fabian, et al. 2016; Machado et al. 2016), structural variants such as chromosomal inversions (reviewed in Kapun and Flatt 2019), transposable elements (TEs) (Boussy et al. 1998; González et al. 2008, 2010), and complex phenotypes (de Jong and Bochdanovits 2003; Schmidt and Paaby 2008; Schmidt et al. 2008; Kapun, Schmidt, et al. 2016; Behrman et al. 2018), especially along the North American and Australian east coasts. In addition to parallel local adaptation, these latitudinal clines are, however,

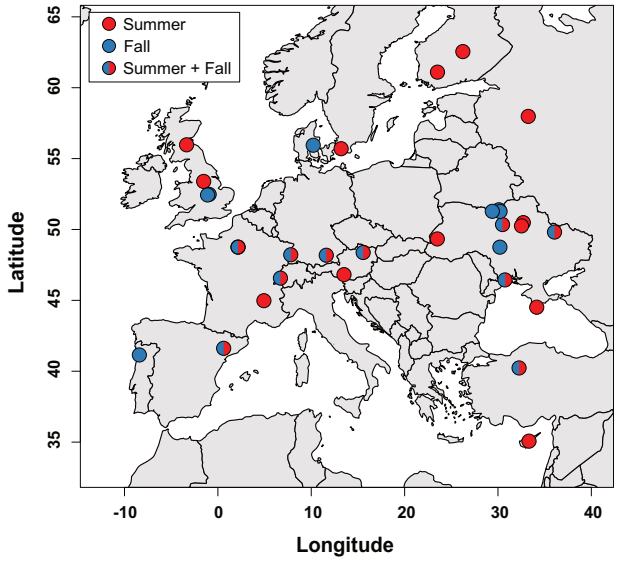

FIG. 1. The geographic distribution of population samples. Locations of all samples in the 2014 *DrosEU* data set. The color of the circles indicates the sampling season for each location: ten of the 32 locations were sampled at least twice, once in summer and once in fall (see table 1 and supplementary table S1, Supplementary Material online). Note that some of the 12 Ukrainian locations overlap in the map.

also affected by admixture with flies from Africa and Europe (Caracristi and Schlötterer 2003; Yukilevich and True 2008a, 2008b; Duchen et al. 2013; Kao et al. 2015; Bergland et al. 2016).

In contrast, the population genomics of *D. melanogaster* on the European continent remains largely unstudied (Božičević et al. 2016; Pool et al. 2016; Mateo et al. 2018). Because Eurasia was the first continent colonized by *D. melanogaster* as they migrated out of Africa, we sought to understand how this species has adapted to new habitats and climate zones in Europe, where it has been established the longest (David and Capy 1988; Lachaise et al. 1988). We analyze our data at three levels: 1) variation at single-nucleotide polymorphisms (SNPs) in nuclear and mitochondrial (mtDNA) genomes ($\sim 5.5 \times 10^6$ SNPs in total); 2) structural variation, including TE insertions and chromosomal inversion polymorphisms; and 3) variation in the microbiota associated with flies, including bacteria, fungi, protists, and viruses.

Results and Discussion

As part of the *DrosEU* consortium, we collected 48 population samples of *D. melanogaster* from 32 geographical locations across Europe in 2014 (table 1 and fig. 1). We performed pooled sequencing (Pool-Seq) of all 48 samples, with an average autosomal coverage $\geq 50\times$ (supplementary table S1, Supplementary Material online). Of the 32 locations, ten were sampled at least once in summer and once in fall (fig. 1), allowing a preliminary analysis of seasonal change in allele frequencies on a genome-wide scale.

A description of the basic patterns of genetic variation of these European *D. melanogaster* population samples, based on SNPs, is provided in Supplementary Material online (see supplementary results and table S1, Supplementary Material online). For each sample, we estimated genome-wide levels of π, Watterson's θ, and Tajima's D (corrected for pooling; Futschik and Schlötterer 2010; Kofler et al. 2011). In brief, patterns of genetic variability and Tajima's D were largely consistent with what has been previously observed on other continents (Fabian et al. 2012; Langley et al. 2012; Lack et al. 2015, 2016), and genetic diversity across the genome varies mainly with recombination rate (Langley et al. 2012). We also found little spatiotemporal variation among European populations in overall levels of sequence variability (table 2).

Below we focus on the identification of selective sweeps, previously unknown longitudinal population structure across the European continent, patterns of local adaptation and clines, and microbiota.

Several Genomic Regions Show Signatures of Continent-Wide Selective Sweeps

To identify genomic regions that have likely undergone selective sweeps in European populations of *D. melanogaster*, we used *Pool-hmm* (Boitard et al. 2013; see supplementary table S2A, Supplementary Material online), which identifies candidate sweep regions via distortions in the allele frequency spectrum. We ran *Pool-hmm* independently for each sample and identified several genomic regions that coincide with previously identified, well-supported sweeps in the proximity of *Hen1* (Kolaczkowski et al. 2011), *Cyp6g1* (Daborn et al. 2002), *wapl* (Beisswanger et al. 2006), and around the chimeric gene *CR18217* (Rogers and Hartl 2012), among others (supplementary table S2B, Supplementary Material online). These regions also showed local reductions in π and Tajima's D, consistent with selective sweeps (fig. 2 and supplementary figs. S1 and S2, Supplementary Material online). The putative sweep regions that we identified in the European populations included 145 of the 232 genes previously identified using *Pool-hmm* in an Austrian population (Boitard et al. 2012; supplementary table S2C, Supplementary Material online). We also identified other regions which have not previously been described as targets of selective sweeps (supplementary table S2A, Supplementary Material online). Of the regions analyzed, 64 showed signatures of selection across all European populations (supplementary table S2D, Supplementary Material online). Of these, 52 were located in the 10% of regions with the lowest values of Tajima's D (SuperExactTest; $P < 0.001$). These may represent continent-wide sweeps that predate the colonization of Europe (Beisswanger et al. 2006) or which have recently swept across the majority of European populations (supplementary table S2D, Supplementary Material online). Indeed, 43 of the 64 genes (67%) that showed signatures of selection across all European populations were located in regions with reduced Tajima's D (lowest 10%) in African populations, suggesting that selective sweeps in these genes might predate the out-of-Africa expansion (supplementary table S2D, Supplementary Material online).

Table 1. Sample Information for All Populations in the *DrosEU* Data Set.

ID	Country	Location	Coll. Date	Number ID	Lat (°)	Lon (°)	Alt (m)	Season	*n*	Coll. Name
AT_Mau_14_01	Austria	Mauternbach	2014-07-20	1	48.38	15.56	572	S	80	Andrea J. Betancourt
AT_Mau_14_02	Austria	Mauternbach	2014-10-19	2	48.38	15.56	572	F	80	Andrea J. Betancourt
TR_Yes_14_03	Turkey	Yesiloz	2014-08-31	3	40.23	32.26	680	S	80	Banu Sebnem Onder
TR_Yes_14_04	Turkey	Yesiloz	2014-10-23	4	40.23	32.26	680	F	80	Banu Sebnem Onder
FR_Vil_14_05	France	Viltain	2014-08-18	5	48.75	2.16	153	S	80	Catherine Montchamp-Moreau
FR_Vil_14_07	France	Viltain	2014-10-27	7	48.75	2.16	153	F	80	Catherine Montchamp-Moreau
FR_Got_14_08	France	Gotheron	2014-07-08	8	44.98	4.93	181	S	80	Cristina Vieira
UK_She_14_09	United Kingdom	Sheffield	2014-08-25	9	53.39	−1.52	100	S	80	Damiano Porcelli
UK_Sou_14_10	United Kingdom	South Queensferry	2014-07-14	10	55.97	−3.35	19	S	80	Darren Obbard
CY_Nic_14_11	Cyprus	Nicosia	2014-08-10	11	35.07	33.32	263	S	80	Eliza Argyridou
UK_Mar_14_12	United Kingdom	Market Harborough	2014-10-20	12	52.48	−0.92	80	F	80	Eran Tauber
UK_Lut_14_13	United Kingdom	Lutterworth	2014-10-20	13	52.43	−1.10	126	F	80	Eran Tauber
DE_Bro_14_14	Germany	Broggingen	2014-06-26	14	48.22	7.82	173	S	80	Fabian Staubach
DE_Bro_14_15	Germany	Broggingen	2014-10-15	15	48.22	7.82	173	F	80	Fabian Staubach
UA_Yal_14_16	Ukraine	Yalta	2014-06-20	16	44.50	34.17	72	S	80	Iryna Kozeretska
UA_Yal_14_18	Ukraine	Yalta	2014-08-27	18	44.50	34.17	72	S	80	Iryna Kozeretska
UA_Ode_14_19	Ukraine	Odesa	2014-07-03	19	46.44	30.77	54	S	80	Iryna Kozeretska
UA_Ode_14_20	Ukraine	Odesa	2014-07-22	20	46.44	30.77	54	S	80	Iryna Kozeretska
UA_Ode_14_21	Ukraine	Odesa	2014-08-29	21	46.44	30.77	54	S	80	Iryna Kozeretska
UA_Ode_14_22	Ukraine	Odesa	2014-10-10	22	46.44	30.77	54	F	80	Iryna Kozeretska
UA_Kyi_14_23	Ukraine	Kyiv	2014-08-09	23	50.34	30.49	179	S	80	Iryna Kozeretska
UA_Kyi_14_24	Ukraine	Kyiv	2014-09-08	24	50.34	30.49	179	F	80	Iryna Kozeretska
UA_Var_14_25	Ukraine	Varva	2014-08-18	25	50.48	32.71	125	S	80	Oleksandra Protsenko
UA_Pyr_14_26	Ukraine	Pyriatyn	2014-08-20	26	50.25	32.52	114	S	80	Oleksandra Protsenko
UA_Dro_14_27	Ukraine	Drogobych	2014-08-24	27	49.33	23.50	275	S	80	Iryna Kozeretska
UA_Cho_14_28	Ukraine	Chornobyl	2014-09-13	28	51.37	30.14	121	F	80	Iryna Kozeretska
UA_Cho_14_29	Ukraine	Chornobyl Yaniv	2014-09-13	29	51.39	30.07	121	F	80	Iryna Kozeretska
SE_Lun_14_30	Sweden	Lund	2014-07-31	30	55.69	13.20	51	S	80	Jessica Abbott
DE_Mun_14_31	Germany	Munich	2014-06-19	31	48.18	11.61	520	S	80	John Parsch
DE_Mun_14_32	Germany	Munich	2014-09-03	32	48.18	11.61	520	F	80	John Parsch
PT_Rec_14_33	Portugal	Recarei	2014-09-26	33	41.15	−8.41	175	F	80	Jorge Vieira
ES_Gim_14_34	Spain	Gimenells (Lleida)	2014-10-20	34	41.62	0.62	173	F	80	Lain Guio
ES_Gim_14_35	Spain	Gimenells (Lleida)	2014-08-13	35	41.62	0.62	173	S	80	Lain Guio
FI_Aka_14_36	Finland	Akaa	2014-07-25	36	61.10	23.52	88	S	80	Maaria Kankare
FI_Aka_14_37	Finland	Akaa	2014-08-27	37	61.10	23.52	88	S	80	Maaria Kankare
FI_Ves_14_38	Finland	Vesanto	2014-07-26	38	62.55	26.24	121	S	66	Maaria Kankare
DK_Kar_14_39	Denmark	Karensminde	2014-09-01	39	55.95	10.21	15	F	80	Mads Fristrup Schou
DK_Kar_14_41	Denmark	Karensminde	2014-11-25	41	55.95	10.21	15	F	80	Mads Fristrup Schou
CH_Cha_14_42	Switzerland	Chalet à Gobet	2014-07-24	42	46.57	6.70	872	S	80	Martin Kapun
CH_Cha_14_43	Switzerland	Chalet à Gobet	2014-10-05	43	46.57	6.70	872	F	80	Martin Kapun
AT_See_14_44	Austria	Seeboden	2014-08-17	44	46.81	13.51	591	S	80	Martin Kapun
UA_Kha_14_45	Ukraine	Kharkiv	2014-07-26	45	49.82	36.05	141	S	80	Svitlana Serga
UA_Kha_14_46	Ukraine	Kharkiv	2014-09-14	46	49.82	36.05	141	F	80	Svitlana Serga
UA_Cho_14_47	Ukraine	Chornobyl Applegarden	2014-09-13	47	51.27	30.22	121	F	80	Svitlana Serga
UA_Cho_14_48	Ukraine	Chornobyl Polisske	2014-09-13	48	51.28	29.39	121	F	70	Svitlana Serga
UA_Kyi_14_49	Ukraine	Kyiv	2014-10-11	49	50.34	30.49	179	F	80	Svitlana Serga
UA_Uma_14_50	Ukraine	Uman	2014-10-01	50	48.75	30.21	214	F	80	Svitlana Serga
RU_Val_14_51	Russia	Valday	2014-08-17	51	57.98	33.24	217	S	80	Elena Pasyukova

NOTE.—Origin, collection date, season, and sample size (number of chromosomes: *n*) of the 48 samples in the *DrosEU* 2014 data set. Additional information can be found in supplementary table S1, Supplementary Material online.

We then asked if there was any indication of selective sweeps particular to a certain habitat. To this end, we classified the populations according to the Köppen–Geiger climate classification (Peel et al. 2007) and identified several putative sweeps exclusive to arid, temperate, and cold regions (supplementary table S2*A*, Supplementary Material online). To shed light on potential phenotypes affected by selective sweeps, we performed a gene ontology (GO) analysis. For temperate climates, this analysis showed enrichment for functions such as "response to stimulus," "transport," and "nervous system development." For cold climates, it showed enrichment for "vitamin and cofactor metabolic processes" (supplementary table S2*E*, Supplementary Material online). There was no enrichment of any GO category for sweeps associated with arid regions.

Thus, we identified several new candidate selective sweeps in European populations of *D. melanogaster*, many of which occur in the majority of European populations and which

Table 2. Clinality of Genetic Variation and Population Structure.

Factor	Latitude	Longitude	Altitude	Season	Moran's *I*
$\pi_{(X)}$	4.11*	1.62	*15.23***	1.65	0.86
$\pi_{(Aut)}$	0.91	2.54	*27.18***	0.16	−0.86
$\theta_{(X)}$	2.65	1.31	*15.54***	2.22	0.24
$\theta_{(Aut)}$	0.48	1.44	*13.66***	0.37	−1.13
$D_{(X)}$	0.02	0.38	5.93*	3.26	−2.08
$D_{(Aut)}$	0.09	0.76	5.33*	0.71	−1.45
PC1	0.63	*118.08***(***)	3.64	0.75	*4.2***
PC2	*4.69**	*7.15**	*11.77***	1.68	−0.32
PC3	0.39	0.23	*19.91***	0.28	1.38

NOTE.—Effects of geographic variables and/or seasonality on genome-wide average levels of diversity (π, θ, and Tajima's *D*; top rows) and on the first three axes of a PCA based on allele frequencies at neutrally evolving sites (bottom rows). The values represent *F* ratios from general linear models. Italic type indicates *F* ratios that are significant after Bonferroni correction (adjusted $\alpha' = 0.0055$). Asterisks in parentheses indicate significance when accounting for spatial autocorrelation by spatial error models. These models were only calculated when Moran's *I* test, as shown in the last column, was significant.
*$P < 0.05$; **$P < 0.01$; ***$P < 0.001$.

merit future study, using sequencing of individual flies and functional genetic experiments.

European Populations Are Structured along an East–West Gradient

We next investigated whether patterns of genetic differentiation might be due to demographic substructuring. Overall, pairwise differentiation as measured by F_{ST} was relatively low, particularly for the autosomes (autosomal F_{ST} 0.013–0.059; *X*-chromosome F_{ST}: 0.043–0.076; Mann–Whitney *U* test; $P < 0.001$; supplementary table S1, Supplementary Material online). The *X* chromosome is expected to have higher F_{ST} than the autosomes, given its relatively smaller effective population size (Mann–Whitney *U* test; $P < 0.001$; Hutter et al. 2007). One population, from Sheffield (UK), was unusually differentiated from the others (average pairwise $F_{ST} = 0.027$; SE= 0.00043 vs. $F_{ST} = 0.04$; SE= 0.00055 for comparisons without this population and with this population only; supplementary table S1, Supplementary Material online). Including this sample in the analysis could potentially lead to exaggerated patterns of geographic differentiation, as it is both highly differentiated and the furthest west. We therefore excluded it from the following analyses of geographic differentiation, as this approach is conservative. (For details, see Supplementary Material online; including or excluding this population did not qualitatively change our results and their interpretation.)

Despite low overall levels of among-population differentiation, we found that European populations exhibit clear evidence of geographic substructuring. For this analysis, we focused on SNPs located within short introns, with a length ≤60 bp and which most likely reflect neutral population structure (Haddrill et al. 2005; Singh et al. 2009; Parsch et al. 2010; Clemente and Vogl 2012; Lawrie et al. 2013). We restricted our analyses to polymorphisms in regions of high recombination ($r > 3$ cM/Mb; Comeron et al. 2012) and to SNPs at least 1 Mb away from the breakpoints of common inversions (and excluding the inversion bodies themselves), resulting in 4,034 SNPs used for demographic analysis.

We found two signatures of geographic differentiation using these putatively neutral SNPs. First, we identified a weak but significant correlation between pairwise F_{ST} and geographic distance, consistent with isolation by distance (IBD; Mantel test; $P < 0.001$; $R^2 = 0.12$, max. $F_{ST} \sim 0.045$; fig. 3*A*). Second, a principal components analysis (PCA) on allele frequencies showed that the three most important PC axes explain >25% of the total variance (PC1: 16.71%, PC2: 5.83%, PC3: 4.6%, eigenvalues = 159.8, 55.7, and 44, respectively; fig. 3*B*). The first axis, PC1, was strongly correlated with longitude ($F_{1,42} = 118.08$, $P < 0.001$; table 2). Again, this pattern is consistent with IBD, as the European continent extends further in longitude than latitude. We repeated the above PCA using SNPs in 4-fold degenerate sites, as these are also assumed to be relatively unaffected by selection (Akashi 1995; Halligan and Keightley 2006; supplementary fig. S3, Supplementary Material online), and found highly consistent results.

Because there was a significant spatial autocorrelation between samples (as indicated by Moran's test on residuals from linear regressions with PC1; $P < 0.001$; table 2), we repeated the analysis with an explicit spatial error model; the association between PC1 and longitude remained significant. To a lesser extent PC2 was likewise correlated with longitude ($F_{1,42} = 7.15$, $P < 0.05$), but also with altitude ($F_{1,42} = 11.77$, $P < 0.01$) and latitude ($F_{1,42} = 4.69$, $P < 0.05$; table 2). Similar to PC2, PC3 was strongly correlated with altitude ($F_{1,42} = 19.91$, $P < 0.001$; table 2). We also examined these data for signatures of genetic differentiation between samples collected at different times of the year. For the data set as a whole, no major PC axes were correlated with season, indicating that there were no strong differences in allele frequencies shared between all our summer and fall samples ($P > 0.05$ for all analyses; table 2). For the ten locations sampled in both summer and fall, we performed separate PC analyses for summer and fall. Summer and fall values of PC1 (adjusted R^2: 0.98; $P < 0.001$), PC2 (R^2: 0.74; $P < 0.001$) and PC3 (R^2: 0.81; $P < 0.001$) were strongly correlated across seasons. This indicates a high degree of seasonal stability in local genetic variation.

Next, we attempted to determine if populations could be statistically classified into clusters of similar populations. Using hierarchical model fitting based on the first four PC axes from the PCA mentioned above, we found two distinct clusters (fig. 3*B*) separated along PC1, supporting the notion

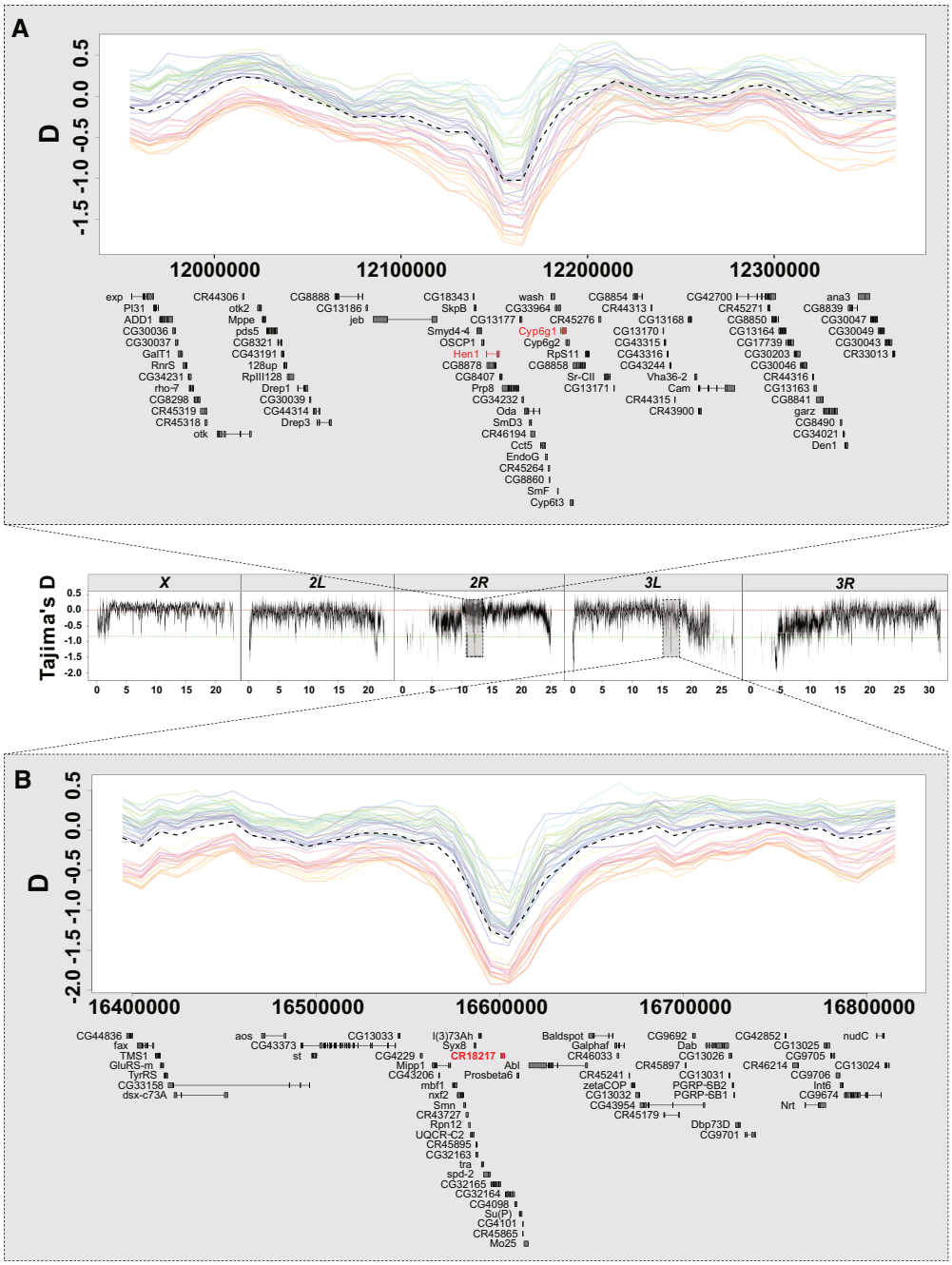

FIG. 2. Candidate signals of selective sweeps in European populations. The central panel shows the distribution of Tajima's *D* in 50-kb sliding windows with 40-kb overlap, with red and green dashed lines indicating Tajima's $D = 0$ and -1, respectively. The top panel shows a detail of a genomic region on chromosomal arm *2R* in the vicinity of *Cyp6g1* and *Hen1* (highlighted in red), genes reportedly involved in pesticide resistance. This strong sweep signal is characterized by an excess of low-frequency SNP variants and overall negative Tajima's *D* in all samples. Colored solid lines depict Tajima's *D* for each sample (see supplementary fig. S2, Supplementary Material online, for color codes, Supplementary Material online); the black dashed line shows Tajima's *D* averaged across all samples. The bottom panel shows a region on *3L* previously identified as a potential target of selection, which shows a similar strong sweep signature. Notably, both regions show strongly reduced genetic variation (supplementary fig. S1, Supplementary Material online).

of strong longitudinal differentiation among European populations. Similarly, model-based spatial clustering also showed that populations were separated mainly by longitude (fig. 3C; using ConStruct, with $K = 3$ spatial layers chosen based on model selection procedure via cross-validation). We also inferred levels of admixture among populations from this analysis, based on the relationship between F_{ST} and migration rate (Wright 1949) and using recent estimates of N_e in European populations ($N_e \sim 3.1 \times 10^6$; Duchen et al. 2013; for pairwise migration rates see supplementary table S3, Supplementary Material online). Within the Western European cluster and between the clusters, $4N_em$ was similar ($4N_em$-WE = 43.76, $4N_em$-between = 45.97); in Eastern Europe, estimates of $4N_em$ indicate significantly higher levels of admixture, despite

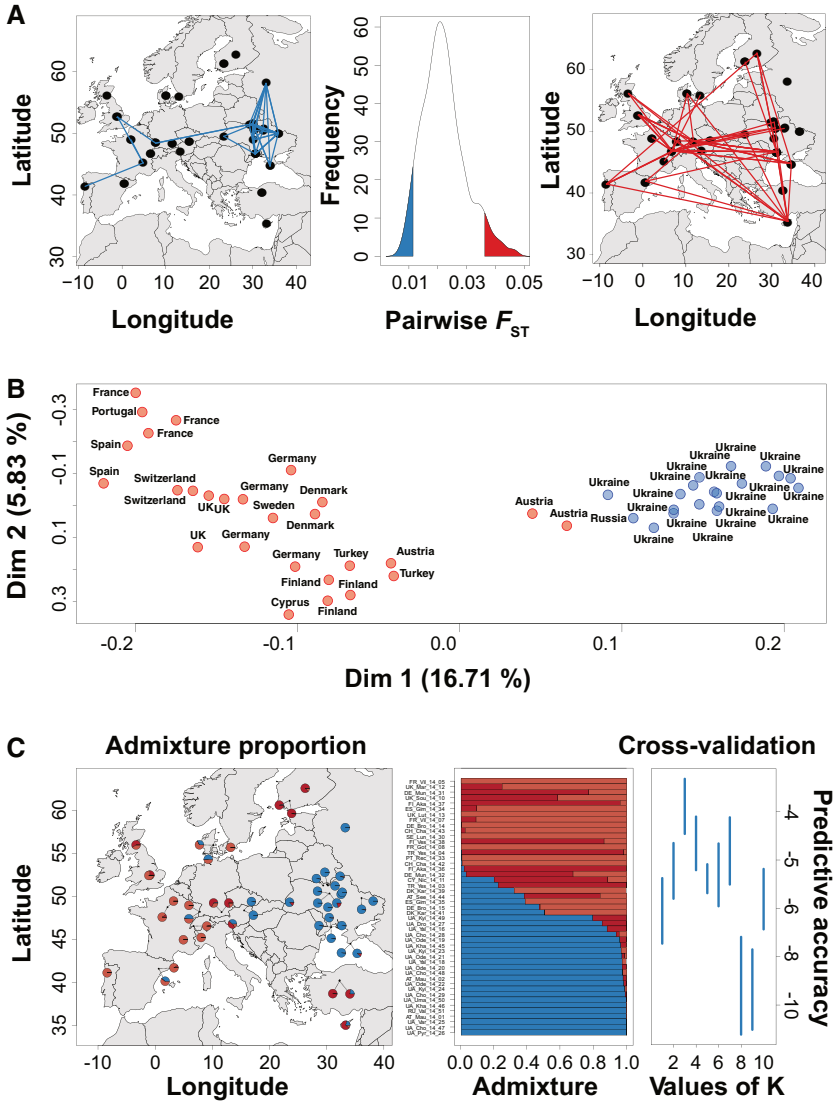

Fig. 3. Genetic differentiation among European populations. (*A*) Average F_{ST} among populations at putatively neutral sites. The center plot shows the distribution of F_{ST} values for all 1,128 pairwise population comparisons, with the F_{ST} values for each comparison obtained from the mean across all 4,034 SNPs used in the analysis. Plots on the left and the right show population pairs in the lower (blue) and upper (red) 5% tails of the F_{ST} distribution. (*B*) PCA analysis of allele frequencies at the same SNPs reveals population substructuring in Europe. Hierarchical model fitting using the first four PCs showed that the populations fell into two clusters (indicated by red and blue), with cluster assignment of each population subsequently estimated by *k*-means clustering. (*C*) Admixture proportions for each population inferred by model-based clustering with *ConStruct* are highlighted as pie charts (left plot) or structure plots (center). The optimal number of three spatial layers (*K*) was inferred by cross-validation (right plot).

the larger geographic range covered by these samples ($4N_em = 74.17$; Mann–Whitney *U* test; $P < 0.001$). This result suggests that longitudinal differentiation in Europe might be partly driven by high levels of genetic exchange in Eastern Europe, perhaps due to migration and recolonization after harsh winters in that region. However, these estimates of gene flow must be interpreted with caution, as unknown demographic events can confound estimates of migration rates from F_{ST} (Whitlock and McCauley 1999).

In addition to restricted gene flow between geographic areas, local adaptation may explain population substructure, even at neutral sites, if nearby and closely related populations are responding to similar selective pressures. We investigated whether any of 19 climatic variables, obtained from the WorldClim database (Hijmans et al. 2005), were associated with the genetic structure in our samples. These climatic variables represent interpolated averages across 30 years of observation at the geographic coordinates corresponding to our sampling locations. As many of these variables are highly intercorrelated, we analyzed their joint effects on genetic variation, by using PCA to summarize the information they capture. The first three climatic PC axes capture >77% of the variance in the 19 climatic variables (supplementary table S4, Supplementary Material online). PC1 explained 36% of the variance and was strongly correlated ($r > 0.75$ or $r < -0.75$) with climatic variables differentiating "hot and dry" from "cold and wet" climates (e.g., maximum temperature of the warmest month, $r = 0.84$; mean temperature of warmest

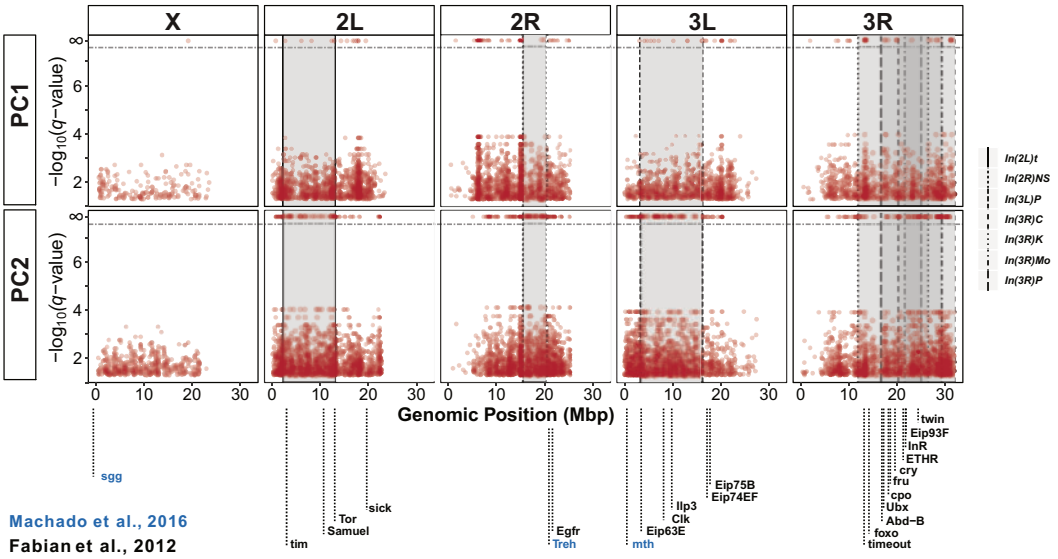

FIG. 4. Manhattan plots of SNPs with q values <0.05 in BayeScEnv association tests with PC1 or PC2 of bioclimatic variables. Vertical lines denote the breakpoints of common inversions. The gene names highlight some candidate genes found in our study and which have previously been identified as varying clinally by Fabian et al. (2012) and Machado et al. (2016) along the North American east coast. Note that q values of 0 (which are infinite on a log-scale) are plotted at the top of each figure, above the gray dash-dotted horizontal lines in order to separate them from the other candidates with q values >0. These zero values are unlikely to be spurious as the densities of these infinite values tend to line up with peaks of $\log_{10}(q)$ below the dashed line, suggesting that they represent highly significant continuations of these peaks.

quarter, $r = 0.86$; annual mean temperature, $r = 0.85$; precipitation during the warmest quarter, $r = -0.87$). Conversely, PC2 (27.3% of variance explained) distinguished climates with low and high differences between seasons (e.g., isothermality, $r = 0.83$; temperature seasonality, $r = 0.88$; temperature annual range, $r = -0.78$; precipitation in coldest quarter, $r = 0.79$). PC1 was strongly correlated with latitude (linear regression: $R^2 = 0.48$, $P < 0.001$), whereas PC2 was strongly associated with longitude ($R^2 = 0.58$, $P < 0.001$). PC2 was also correlated with latitude ($R^2 = 0.11$, $P < 0.05$) and altitude ($R^2 = 0.12$, $P < 0.01$).

We next asked whether any of these climate PCs explained any of the genetic structure uncovered above. Pairwise linear regressions of the first three PC axes based on allele frequencies of intronic SNPs against the first three climatic PCs revealed only one significant correlation after Bonferroni correction: between climatic PC2 ("seasonality") versus genetic PC1 (longitude; adjusted $\alpha = 0.017$; $R^2 = 0.49$, $P < 0.001$). This suggests that longitudinal differentiation along the European continent might be partly driven by the transition from oceanic to continental climate, possibly leading to local adaptation to gradual changes in temperature seasonality and the severity of winter conditions.

Interestingly, the central European division into an eastern and a western clade of *D. melanogaster* closely resembles known hybrid zones of organisms which form closely related pairs of sister taxa. These biogeographic patterns have been associated with long-term reductions of gene flow between eastern and western populations during the last glacial maximum, followed by postglacial recolonization of the continent from southern refugia (Hewitt 1999). However, in contrast to many of these taxa, which often exhibit pronounced pre- and postzygotic isolation (Szymura and Barton 1986; Haas and Brodin 2005; Macholán et al. 2008; Knief et al. 2019), we found low genome-wide differentiation among eastern and western populations (average max. $F_{ST} \sim 0.045$), perhaps indicating that the longitudinal division of European *D. melanogaster* is not the result of postglacial secondary contact.

Climatic Predictors Identify Genomic Signatures of Local Climate Adaptation

To further explore climatic patterns, and to identify signatures of local adaptation caused by climatic differences among populations independent of neutral demographic effects, we tested for associations of SNP alleles with climatic PC1 and PC2 using BayeScEnv (de Villemereuil and Gaggiotti 2015). The total number of SNPs tested and the number of "top SNPs" (q value<0.05) are given in supplementary table S5A, Supplementary Material online. A large proportion of the top SNPs were intergenic (PC1: 33.5%; PC2: 32.2%) or intronic variants (PC1: 50.1%; PC2: 50.5%). Manhattan plots of q values for all SNPs are shown in figure 4. These figures show some distinct "peaks" of highly differentiated SNPs along with some broader regions of moderately differentiated SNPs (fig. 4). For example, the circadian rhythm gene *timeout* and the ecdysone signaling genes *Eip74EF* and *Eip75B* all lie near peaks associated with climatic PC1 ("hot/dry" vs. "cold/wet"; fig. 4, top panels). We note that the corresponding genes have been identified in previous studies of clinal (latitudinal) differentiation in North American *D. melanogaster* (Fabian et al. 2012; Machado et al. 2016). We found a significant overlap between genes associated with PC1 and PC2 (both of which are correlated with latitude) in our study and candidate gene sets from these previous studies of latitudinal clines (SuperExactTest; $P < 0.001$; Fabian et al. 2012; Machado et al. 2016). For example, out of 1,974 latitudinally varying loci

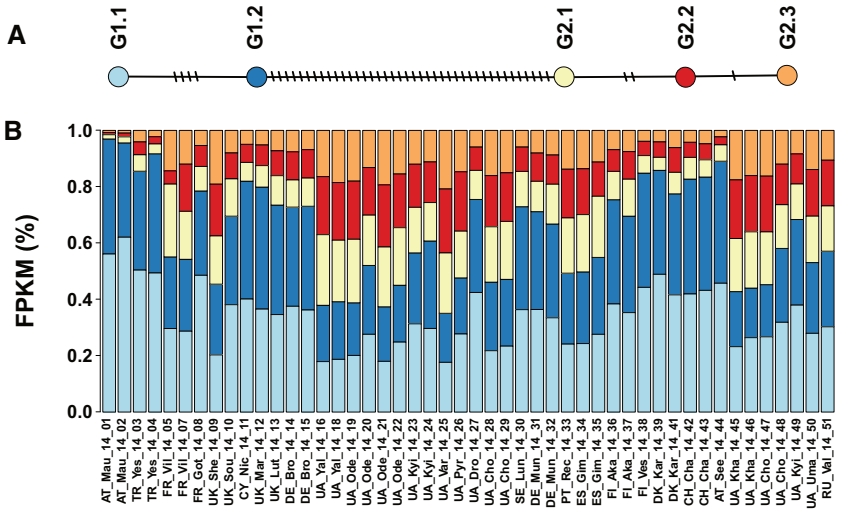

Fig. 5. Mitochondrial haplotypes. (*A*) Network showing the relationship of five common mitochondrial haplotypes. (*B*) Estimated frequency of each mitochondrial haplotype in 48 European samples.

along the North American east coast identified by Fabian et al. (2012), we found 403 (20%) and 505 (26%) of them to also be associated with PC1 and PC2 in European populations, respectively (supplementary table S5*B* and *C*, Supplementary Material online). Moreover, the BayeScEnv analysis and *Pool-hmm* analysis together identify four regions with both climatic associations and evidence for continent-wide selective sweeps (supplementary table S5*B* and *C*, Supplementary Material online). Finally, four other BayeScEnv candidate genes were previously identified as targets of selection in African and North American populations based on significant McDonald–Kreitman tests (Langley et al. 2012; see supplementary table S5*B* and *C*, Supplementary Material online).

We next asked whether any insights into the targets of local selection could be gleaned from examining the functions of genes near the BayeScEnv peaks. We examined annotated features within 2 kb of significantly associated SNPs (PC1: 3,545 SNPs near 2,078 annotated features; PC2: 5,572 SNPs near 2,717 annotated features; supplementary table S5*B* and *C*, Supplementary Material online). First, we performed a GO term analysis with GOwinda (Kofler and Schlötterer 2012) to ask whether SNPs associated with climatic PCs are enriched for any gene functions. For PC1, we found no GO term enrichment. For PC2, we found enrichment for "cuticle development" and "UDP-glucosyltransferase activity." Next, we performed functional annotation clustering with DAVID (v6.8; Huang et al. 2009), and identified 37 and 47 clusters with an enrichment score $>$ 1.3 for PC1 and PC2, respectively (supplementary table S5*D* and *E*, Supplementary Material online, Huang et al. 2009). PC1 was enriched for categories such as "sex differentiation" and "response to nicotine," whereas PC2 was enriched for functional categories such as "response to nicotine," "integral component of membrane," and "sensory perception of chemical stimulus" (supplementary table S5*D* and *E*, Supplementary Material online).

We also asked whether the SNPs identified by BayeScEnv show consistent signatures of local adaptation. Many associated genes (1,205) were also shared between PC1 and PC2. Some genes have indeed been previously implicated in climatic and clinal adaptation, such as the circadian rhythm genes *timeless*, *timeout*, and *clock*, the sexual differentiation gene *fruitless*, and the *couch potato* locus which underlies the latitudinal cline in reproductive dormancy in North America (Tauber et al. 2007; Schmidt et al. 2008; Fabian et al. 2012). Notably, these also include the major insulin signaling genes *insulin-like receptor* (*InR*) and *forkhead box subgroup O* (*foxo*), which have strong genomic and experimental evidence implicating these loci in clinal, climatic adaptation along the North America east coast (Paaby et al. 2010, 2014; Fabian et al. 2012; Durmaz et al. 2019). Thus, European populations share multiple potential candidate targets of selection with North American populations (cf. Fabian et al. 2012; Machado et al. 2016; also see Božičević et al. 2016). We next turned to examining polymorphisms other than SNPs, that is, mitochondrial haplotypes as well as inversion and TE polymorphisms.

Mitochondrial Haplotypes Also Exhibit Longitudinal Population Structure

Mitochondrial haplotypes also showed evidence of longitudinal demographic structure in European population. We identified two main alternative mitochondrial haplotypes in Europe, G1 and G2, each with several subhaplotypes (G1.1 and G1.2 and G2.1, G2.2, and G2.3). The two subtypes, G1.2 and G2.1, are separated by 41 mutations (fig. 5*A*). The frequencies of the alternative G1 and G2 haplotypes varied among populations between 35.1% and 95.6% and between 4.4% and 64.9%, respectively (fig. 5*B*). Qualitatively, three types of European populations could be distinguished based on these haplotypes: 1) central European populations, with a high frequency ($>$60%) of G1 haplotypes, 2) Eastern European populations in summer, with a low frequency ($<$40%) of G1 haplotypes, and 3) Iberian and Eastern European populations in fall, with a frequency of G1 haplotypes between 40% and 60% (supplementary fig. S4,

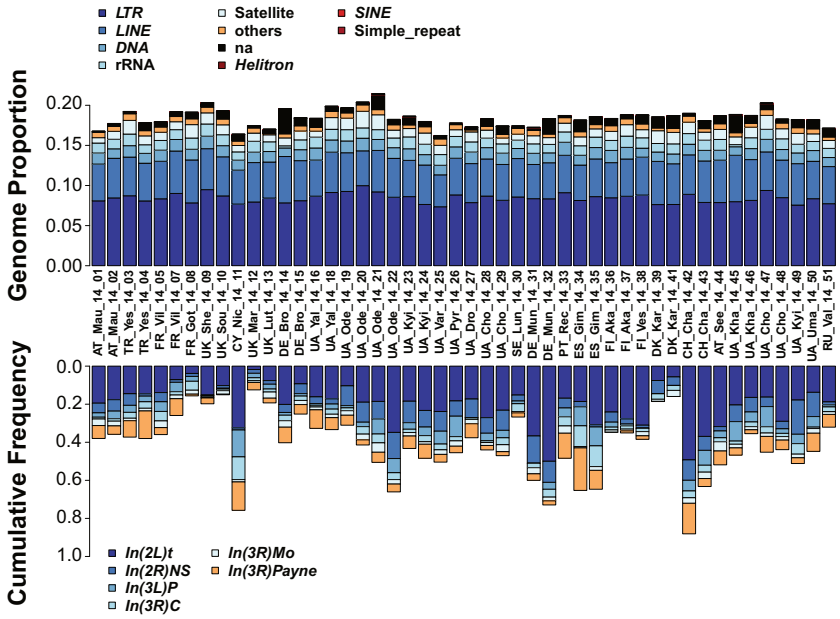

FIG. 6. Geographic patterns of structural variants. The upper panel shows stacked bar plots with the relative abundances of transposable elements (TEs) in all 48 population samples. The proportion of each repeat class was estimated from sampled reads with dnaPipeTE (two samples per run, 0.1× coverage per sample). The lower panel shows stacked bar plots depicting absolute frequencies of six cosmopolitan inversions in all 48 population samples.

Supplementary Material online). Analyses of mitochondrial haplotypes from a North American population (Cooper et al. 2015) as well as from world-wide samples (Wolff et al. 2016) also revealed high levels of haplotype diversity.

Although there was no correlation between the frequency of G1 haplotypes and latitude, G1 haplotypes and longitude were weakly but significantly correlated ($r^2 = 0.10$; $P < 0.05$). We thus divided the data set into an eastern and a western subset along the 20° meridian, corresponding to the division of two major climatic zones, temperate (oceanic) versus cold (continental) (Peel et al. 2007). This split revealed a clear correlation (r^2=0.5; $P < 0.001$) between longitude and the frequency of G1 haplotypes, explaining as much as 50% of the variation in the western group (supplementary fig. S4*B*, Supplementary Material online). Similarly, in eastern populations, longitude and the frequency of G1 haplotypes were correlated ($r^2 = 0.2$; $P < 0.001$), explaining ~20% of the variance (supplementary fig. S4*B*, Supplementary Material online). Thus, these mitochondrial haplotypes appear to follow a similar east–west population structure as observed for the nuclear SNPs described above.

The Frequency of Polymorphic TEs Varies with Longitude and Altitude

To examine the population genomics of structural variants, we first focused on TEs. Similar to previous findings, the repetitive content of the 48 samples ranged from 16% to 21% of the nuclear genome size (Quesneville et al. 2005; fig. 6). The vast majority of detected repeats were TEs, mostly long-terminal repeat elements (LTRs; range 7.55–10.15%) and long interspersed nuclear elements (LINEs; range 4.18–5.52%), along with a few DNA elements (range 1.16–1.65%) (supplementary table S6, Supplementary Material online). LTRs have been previously described as being the most abundant TEs in the *D. melanogaster* genome (Kaminker et al. 2002; Bergman et al. 2006). Correspondingly, variation in the proportion of LTRs best explained variation in total TE content (LINE+LTR+DNA) (Pearson's $r = 0.87$, $P < 0.01$, vs. DNA $r = 0.58$, $P = 0.0117$, and LINE $r = 0.36$, $P < 0.01$ and supplementary fig. S5*A*, Supplementary Material online).

For each of the 1,630 euchromatic TE insertion sites annotated in the *D. melanogaster* reference genome *v.6.04*, we estimated the frequency at which a copy of the TE was present at that site using *T-lex2* (Fiston-Lavier et al. 2015; see supplementary table S7, Supplementary Material online). On an average, 56% were fixed in all samples. The remaining polymorphic TEs mostly segregated at low frequency in all samples (supplementary fig. S5*B*, Supplementary Material online), potentially due to purifying selection (González et al. 2008; Petrov et al. 2011; Kofler et al. 2012; Cridland et al. 2013; Blumenstiel et al. 2014). However, 246 were present at intermediate frequencies (>10% and <95%) and located in regions of nonzero recombination (Fiston-Lavier et al. 2010; Comeron et al. 2012; see supplementary table S7, Supplementary Material online). Although some of these insertions might be segregating neutrally at transposition–selection balance (Charlesworth et al. 1994; see supplementary fig. S5*B*, Supplementary Material online), they are likely enriched for candidate adaptive mutations (Rech et al. 2019).

In each of the 48 samples, TE frequency and recombination rate were negatively correlated genome-wide (Spearman rank sum test; $P < 0.01$), as has also been previously reported for *D. melanogaster* (Bartolomé et al. 2002; Petrov et al. 2011; Kofler et al. 2012). This remains true when fixed (population

Table 3. Clinality and/or Seasonality of Chromosomal Inversions.

Factor	Latitude	Longitude	Altitude	Season	Moran's *I*
In(2L)t	2.2	*10.09***	*43.94****	0.89	−0.92
In(2R)NS	0.25	*14.43****	2.88	2.43	1.25
In(3L)P	*21.78****	2.82	0.62	3.6	−1.61
In(3R)C	*18.5*** (***)	0.75	1.42	0.04	2.79**
In(3R)Mo	0.3	0.09	0.35	0.03	−0.9
In(3R)Payne	*43.47****	0.66	1.69	1.55	−0.89

NOTE.—The values represent *F* ratios from binomial generalized linear models to account for frequency data. Underlined italic type indicates deviance values that were significant after Bonferroni correction (adjusted $\alpha' = 0.0071$). Asterisks in parentheses indicate significance when accounting for spatial autocorrelation by spatial error models. These models were only calculated when Moran's *I* test, as shown in the last column, was significant.
**$P < 0.01$; ***$P < 0.001$.

frequency ≥95%) TE insertions were excluded from the analysis, although it was not statistically significant for some chromosomes and populations (supplementary table S8, Supplementary Material online). In both cases, the correlation was stronger when broad-scale (Fiston-Lavier et al. 2010) rather than fine-scale (Comeron et al. 2012) recombination rate estimates were used, indicating that the former may best capture long-term population recombination patterns (see supplementary materials and methods and table S8, Supplementary Material online).

We next tested whether variation in TE frequencies among samples was associated with spatially or temporally varying factors. We focused on 111 TE insertions that segregated at intermediate frequencies, were located in nonzero recombination regions, and that showed an interquartile range (IQR) >10 (see supplementary materials and methods, Supplementary Material online). Of these insertions, 57 were significantly associated with an at least one variable of interest after multiple testing correction (supplementary table S9*A*, Supplementary Material online): 13 were significantly associated with longitude, 13 with altitude, five with latitude, three with season, and 23 insertions with more than one of these variables (supplementary table S9*A*, Supplementary Material online). These 57 TEs were mainly located inside genes (42 out of 57; Fisher's exact test, $P > 0.05$; supplementary table S9*A*, Supplementary Material online).

The 57 TEs significantly associated with these environmental variables were enriched for two TE families: the LTR *297* family with 11 copies, and the DNA *pogo* family with five copies (χ^2 values after Yate's correction < 0.05; supplementary table S9*B*, Supplementary Material online). Interestingly, 17 of the 57 TEs coincided with previously identified adaptive candidate TEs, suggesting that our data set might be enriched for adaptive insertions (SuperExactTest, $P < 0.001$), several of which exhibit spatial frequency clines that deviate from neutral expectation (SuperExactTest, $P < 0.001$, supplementary table S9*A*, Supplementary Material online; cf.; Rech et al. 2019). Moreover, 18 of the 57 TEs also show significant correlations with either geographical or temporal variables in North American populations (SuperExactTest, $P < 0.001$, supplementary table S9*A*, Supplementary Material online; cf. Lerat et al. 2019).

Inversions Exhibit Latitudinal and Longitudinal Clines in Europe

Polymorphic chromosomal inversions, another class of structural variants besides TEs, are well-known to exhibit pronounced spatial (clinal) patterns in North American, Australian, and other populations, possibly due to spatially varying selection (reviewed in Kapun and Flatt 2019; also see Mettler et al. 1977; Knibb et al. 1981; Lemeunier and Aulard 1992; Hoffmann and Weeks 2007; Fabian et al. 2012; Kapun et al. 2014; Adrion et al. 2015; Rane et al. 2015; Kapun, Fabian, et al. 2016). However, in contrast to North America and Australia, inversion clines in Europe remain very poorly characterized (Lemeunier and Aulard 1992; Kapun and Flatt 2019). We therefore sought to examine the presence and frequency of six cosmopolitan inversions (*In(2L)t*, *In(2R)NS*, *In(3L)P*, *In(3R)C*, *In(3R)Mo*, *In(3R)Payne*) in our European samples, using a panel of highly diagnostic inversion-specific marker SNPs, identified through sequencing of cytologically determined karyotypes by Kapun et al. (2014) (also see Kapun, Fabian, et al. 2016). All 48 samples were polymorphic for one or more inversions (fig. 6). However, only *In(2L)t* segregated at substantial frequencies in most populations (average frequency = 20.2%); all other inversions were either absent or rare (average frequencies: *In(2R)NS* = 6.2%, *In(3L)P* = 4%, *In(3R)C* = 3.1%, *In(3R)Mo* =2.2%, *In(3R)Payne* = 5.7%) (cf. Kapun, Fabian, et al. 2016; Kapun and Flatt 2019).

Despite their overall low frequencies, several inversions showed pronounced clinality, in qualitative agreement with findings from other continents (Lemeunier and Aulard 1992; Kapun and Flatt 2019). For the analyses below, we tested for potentially confounding effects of significant residual spatial autocorrelation among samples; all of these tests were negative, except for *In(3R)C* (Moran's $I \approx 0$, $P > 0.05$ for all tests; table 3). We observed significant latitudinal clines in Europe for *In(3L)P*, *In(3R)C*, and *In(3R)Payne* (binomial generalized linear model: Inversion frequency ~ Latitude + Longitude + Altitude + Season; effect of Latitude: $P < 0.001$ for all; see table 3). Clines for *In(3L)P* and *In(3R)Payne* were similar between Europe and North America (with frequencies for both decreasing with latitude, $P < 0.05$; see supplementary table S10, Supplementary Material online). However, all inversions differed in their frequency at the same latitude between North America and Europe ($P < 0.001$ for the Latitude × Continent interaction; supplementary table S10, Supplementary Material online).

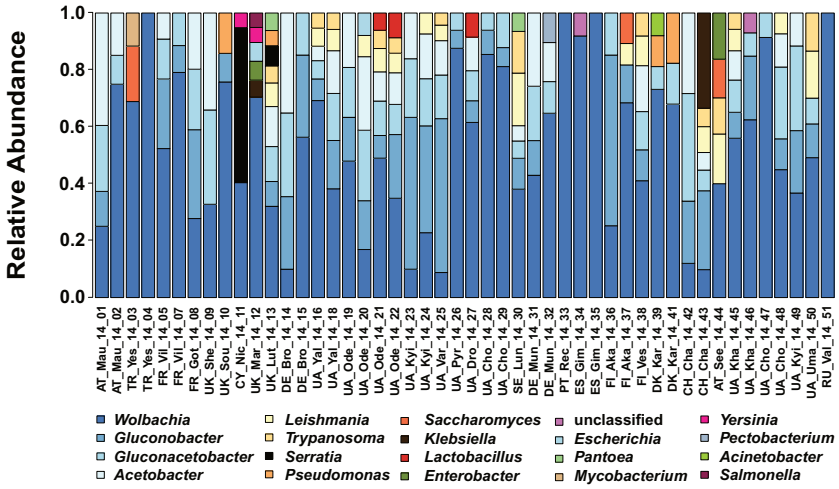

FIG. 7. Microbiota associated with *Drosophila*. Relative abundance of *Drosophila*-associated microbes as assessed by MGRAST classified shotgun sequences. Microbes had to reach at least 3% relative abundance in one of the samples to be represented.

Latitudinal inversion clines previously observed along the North American and Australian east coasts (supplementary fig. S6 and table S10, Supplementary Material online; Kapun, Fabian, et al. 2016) have been attributed to spatially varying selection, especially in the case of *In(3R)Payne* (Anderson et al. 2005; Umina et al. 2005; Kennington et al. 2006; Rako et al. 2006; Kapun, Fabian, et al. 2016; Kapun, Schmidt, et al. 2016; Durmaz et al. 2018; Kapun and Flatt 2019). Similar to patterns in North America (Kapun, Fabian, et al. 2016), we observed that clinality of the three inversion polymorphisms was markedly stronger than for putatively neutral SNPs in short introns (see supplementary table S11, Supplementary Material online), suggesting that these polymorphisms are maintained nonneutrally. Together, these findings suggest that latitudinal inversion clines in Europe are shaped by spatially varying selection, as they are in North America (Kapun, Fabian, et al. 2016; Kapun and Flatt 2019).

We also detected longitudinal clines for *In(2L)t* and *In(2R)NS*, with both polymorphisms decreasing in frequency from east to west (see table 3; $P < 0.01$; also cf. Kapun and Flatt 2019). Longitudinal clines for these two inversions have also been found in North America (cf. Kapun and Flatt 2019). One of these inversions, *In(2L)t*, also changed in frequency with altitude (table 3; $P < 0.001$). These longitudinal and altitudinal inversion clines did, however, not deviate from neutral expectation (supplementary table S11, Supplementary Material online).

European *Drosophila* Microbiomes Contain Entomophthora, Trypanosomatids, and Previously Unknown DNA Viruses

The microbiota can affect life-history traits, immunity, hormonal physiology, and metabolic homeostasis of their fly hosts (Martino et al. 2017; Trinder et al. 2017) and might thus reveal interesting patterns of local adaptation. We therefore examined the bacterial, fungal, protist, and viral microbiota sequence content of our samples. To do this, we characterized the taxonomic origin of the non-*Drosophila* reads in our data set using MGRAST, which identifies and counts short protein motifs (features) within reads (Meyer et al. 2008). We examined 262 million reads in total. Of these, most were assigned to *Wolbachia* (mean 53.7%; fig. 7 and supplementary table S1, Supplementary Material online), a well-known endosymbiont of *Drosophila* (Werren et al. 2008). The abundance of *Wolbachia* protein features relative to other microbial protein features (relative abundance) varied strongly between samples, ranging from 8.8% in a sample from Ukraine to almost 100% in samples from Spain, Portugal, Turkey, and Russia (supplementary table S12, Supplementary Material online). Similarly, *Wolbachia* loads varied 100-fold between samples, as estimated from the ratio of *Wolbachia* protein features to *Drosophila* protein features (supplementary table S12, Supplementary Material online). In contrast to a previous study (Kriesner et al. 2016), there was no evidence for clinality of *Wolbachia* loads ($P = 0.13$, longitude; $P = 0.41$, latitude; Kendall's rank correlation). However, these authors measured infection frequencies whereas we measured *Wolbachia* loads in pooled samples. Because the frequency of infection does not necessarily correlate with microbial loads measured in pooled samples, we might not have been able to detect such a signal in our data.

Acetic acid bacteria of the genera *Gluconobacter*, *Gluconacetobacter*, and *Acetobacter* were the second largest group, with an average relative abundance of 34.4% among microbial protein features. Furthermore, we found evidence for the presence of several genera of Enterobacteria (*Serratia*, *Yersinia*, *Klebsiella*, *Pantoea*, *Escherichia*, *Enterobacter*, *Salmonella*, and *Pectobacterium*). *Serratia* occurs only at low frequencies or is absent from most of our samples, but reaches a very high relative abundance among microbial protein features in the Nicosia (Cyprus) summer collection (54.5%). This high relative abundance was accompanied by an 80× increase in *Serratia* bacterial load.

We also detected several eukaryotic microorganisms, although they were less abundant than the bacteria. We found trypanosomatids, previously reported to be associated with *Drosophila* in other studies (Wilfert et al. 2011; Chandler and

James 2013; Hamilton et al. 2015), in 16 of our samples, on an average representing 15% of all microbial protein features identified in these samples.

Fungal protein features make up <3% of all but three samples (from Finland, Austria, and Turkey; supplementary table S12, Supplementary Material online). This is somewhat surprising because yeasts are commonly found on rotting fruit, the main food substrate of *D. melanogaster*, and co-occur with flies (Barata et al. 2012; Chandler et al. 2012). This result suggests that, although yeasts can attract flies and play a role in food choice (Becher et al. 2012; Buser et al. 2014), they might not be highly prevalent in or on *D. melanogaster* bodies. One reason might be that they are actively digested and thus not part of the microbiome. We also found the fungal pathogen *Entomophthora muscae* in 14 samples, making up 0.18% of the reads (Elya et al. 2018).

Our data also allowed us to identify DNA viruses. Only one DNA virus has been previously described for *D. melanogaster* (Kallithea virus; Webster et al. 2015; Palmer et al. 2018) and only two additional ones from other Drosophilid species (*D. innubila* Nudivirus, Unckless 2011; Invertebrate Iridovirus 31 in *D. obscura* and *D. immigrans*, Webster et al. 2016). In our data set, ~2 million reads came from *Kallithea* nudivirus (Webster et al. 2015), allowing us to assemble the first complete *Kallithea* genome (>300-fold coverage in the Ukrainian sample UA_Kha_14_46; GenBank accession number KX130344).

We also found reads from five additional DNA viruses that were previously unknown (supplementary table S13, Supplementary Material online). First, around 1,000 reads come from a novel nudivirus closely related to both *Kallithea* virus and to *D. innubila* nudivirus (Unckless 2011) in sample DK_Kar_14_41 from Karensminde, Denmark (supplementary table S13, Supplementary Material online). As the reads from this virus were insufficient to assemble the genome, we identified a publicly available data set (SRR3939042: 27 male *D. melanogaster* from Esparto, California; Machado et al. 2016) with sufficient reads to complete the genome (provisionally named "*Esparto* Virus"; KY608910). Second, we also identified two novel Densoviruses (*Parvoviridae*). The first is a relative of *Culex pipiens* densovirus, provisionally named "*Viltain* virus," found at 94-fold coverage in sample FR_Vil_14_07 (Viltain; KX648535). The second is "*Linvill Road* virus," a relative of *Dendrolimus punctatus* densovirus, represented by only 300 reads here, but with high coverage in data set SRR2396966 from a North American sample of *D. simulans*, permitting assembly (KX648536; Machado et al. 2016). Third, we detected a novel member of the *Bidnaviridae* family, "*Vesanto* virus," a bidensovirus related to *Bombyx mori* densovirus 3 with ~900-fold coverage in sample FI_Ves_14_38 (Vesanto; KX648533 and KX648534). Finally, in one sample (UA_Yal_14_16), we detected a substantial number of reads from an Entomopox-like virus, which we were unable to fully assemble (supplementary table S13, Supplementary Material online).

Using a detection threshold of >0.1% of the *Drosophila* genome copy number, the most commonly detected viruses were *Kallithea* virus (30/48 of the pools) and *Vesanto* virus (25/48), followed by *Linvill Road* virus (7/48) and *Viltain* virus (5/48), with *Esparto* virus and the entomopox-like virus being the rarest (2/48 and 1/48, respectively). Because *Wolbachia* can protect *Drosophila* from viruses (Teixeira et al. 2008), we hypothesized that *Wolbachia* loads might correlate negatively with viral loads, but found no evidence of such a correlation ($P = 0.83$ Kallithea virus; $P = 0.76$ Esparto virus; $P = 0.52$ Viltain virus; $P = 0.96$ Vesanto 1 virus; $P = 0.93$ Vesanto 2 virus; $P = 0.5$ Linvill Road virus; Kendall's rank correlation). Perhaps this is because the *Kallithea* virus, the most prevalent virus in our data set, is not expected to be affected by *Wolbachia* (Palmer et al., 2018). Similarly, Shi et al. (2018) found no link between *Wolbachia* and the prevalence or abundance of RNA viruses in data from individual flies.

The variation in bacterial microbiomes across space and time reported here is analyzed in more detail in Wang et al. (2020); this study suggests that some of this variation is structured geographically (cf. Walters et al. 2020). Thus, microbiome composition might contribute to phenotypic differences and local adaptation among populations (Haselkorn et al. 2009; Richardson et al. 2012; Staubach et al. 2013; Kriesner et al. 2016; Wang and Staubach 2018).

Conclusions

Here, we have comprehensively sampled and sequenced European populations of *D. melanogaster* for the first time (fig. 1). We find that European *D. melanogaster* populations are longitudinally differentiated for putatively neutral SNPs, mitochondrial haplotypes as well as for inversion and TE insertion polymorphisms. Potentially adaptive polymorphisms also show this pattern, possibly driven by the transition from oceanic to continental climate along the longitudinal axis of Europe. We note that this longitudinal differentiation qualitatively resembles the one observed for human populations in Europe (Cavalli-Sforza 1966; Xiao et al. 2004; Francalacci and Sanna 2008; Novembre et al. 2008). Given that *D. melanogaster* is a human commensal (Keller 2007; Arguello et al. 2019), it is thus tempting to speculate that the demographic history of European populations might have been influenced by past human migration. Outside Europe, east–west structure has been previously found in sub-Saharan Africa populations of *D. melanogaster*, with the split between eastern and western African populations having occurred ~70 ka (Michalakis and Veuille 1996; Aulard et al. 2002; Kapopoulou, Pfeifer, et al. 2018), a period that coincides with a wave of human migration from eastern into western Africa (Nielsen et al. 2017). However, in contrast to the pronounced pattern observed in Europe, African east–west structure is relatively weak, explaining only ~2.7% of variation, and is primarily due to an inversion whose frequency varies longitudinally. In contrast, our demographic analyses are based on SNPs located in >1 Mb distance from the breakpoints of the most common inversions and excluding the inversion bodies, making it unlikely that the longitudinal pattern we observe is driven by inversions.

Our extensive sampling was feasible only due to synergistic collaboration among many research groups. Our efforts in

Europe are paralleled in North America by the *Dros-RTEC* consortium (Machado et al. 2019), with whom we are collaborating to compare population genomic data across continents. Together, we have sampled both continents annually since 2014; we aim to continue to sample and sequence European and North American *Drosophila* populations with increasing spatiotemporal resolution in future years. With these efforts, we hope to provide a rich community resource for biologists interested in molecular population genetics and adaptation genomics.

Materials and Methods

A detailed description of the Materials and Methods is provided in supplementary materials and methods, Supplementary Material online; here, we give a brief overview of the data set and the basic methods used. The 2014 *DrosEU* data set represents the most comprehensive spatiotemporal sampling of European *D. melanogaster* populations to date (fig. 1 and supplementary table S1, Supplementary Material online). It comprises 48 samples of *D. melanogaster* collected from 32 geographical locations across Europe at different time points in 2014 through a joint effort of 18 research groups. Collections were mostly performed with baited traps using a standardized protocol (see supplementary materials and methods, Supplementary Material online). From each collection, we pooled 33–40 wild-caught males. We used males as they are more easily distinguishable morphologically from similar species than females. Despite our precautions, we identified a low level of *D. simulans* contamination in our sequences; we computationally filtered these sequences from the data prior to further analysis (see Supplementary Material online). To sequence these samples, we extracted DNA and barcoded each sample, and sequenced the ~40 flies per sample as a pool (Pool-Seq; Schlötterer et al. 2014), as paired-end fragments on a *Illumina NextSeq 500* sequencer at the Genomics Core Facility of Pompeu Fabra University. Samples were multiplexed in five batches of ten samples, except for one batch of eight samples (supplementary table S1, Supplementary Material online). Each multiplexed batch was sequenced on four lanes at ~50× raw coverage per sample. The read length was 151 bp, with a median insert size of 348 bp (range 209–454 bp). Our genomic data set is available under NCBI Bioproject accession PRJNA388788. Sequences were processed and mapped to the *D. melanogaster* reference genome (v.6.12) and reference sequences from common commensals and pathogens. Our bioinformatic pipeline is available at https://github.com/capoony/DrosEU_pipeline (last accessed May 22, 2020). To call SNPs, we developed custom software (*PoolSNP*; see supplementary materials and methods, Supplementary Material online; https://github.com/capoony/PoolSNP, last accessed May 22, 2020), using stringent heuristic parameters. In addition, we obtained genome sequences from African flies from the *Drosophila* Genome Nexus (DGN; http://www.johnpool.net/genomes.html, last accessed May 22, 2020; see supplementary table S14, Supplementary Material online, for SRA accession numbers). We used data from 14 individuals from Rwanda and 40 from Siavonga (Zambia). We mapped these data to the *D. melanogaster* reference genome using the same pipeline as for our own data above, and built consensus sequences for each haploid sample by only considering alleles with >0.9 allele frequencies. We converted consensus sequences to *VCF* and used *VCFtools* (Danecek et al. 2011) for downstream analyses. Additional steps in the mapping and variant calling pipeline and further downstream analyses of the data are detailed in supplementary materials and methods, Supplementary Material online.

Supplementary Material

Supplementary data are available at *Molecular Biology and Evolution* online.

Acknowledgments

We acknowledge support of the publication fee by the CSIC Open Access Publication Support Initiative through its Unit of Information Resources for Research (URICI). We thank three anonymous reviewers and the editors for their helpful comments on a previous version of our article. We are grateful to the members of the *DrosEU* and Dros-RTEC consortia and to Dmitri Petrov (Stanford University) for support and discussion. *DrosEU* is funded by a Special Topic Networks (STN) grant from the European Society for Evolutionary Biology (ESEB). Computational analyses were partially executed at the Vital-IT bioinformatics facility of the University of Lausanne (Switzerland), the computing facilities of the CC LBBE/PRABI in Lyon (France), the bwUniCluster of the state of Baden–Württemberg (bwHPC), and the University of St Andrews Bioinformatics Unit which is funded by a Wellcome Trust ISSF award (Grant No. 105621/Z/14/Z). We are also grateful to Simon Boitard and Oscar Gaggiotti for their helpful technical advice on *Pool-hmm* and BayeScEnv analyses, respectively. This study was supported by University of Freiburg Research Innovation Fund 2014, Deutsche Forschungsgemeinschaft (DFG) (Grant No. STA1154/4-1Project 408908608 to F.S.); Academy of Finland (Grant Nos. 268241 and 272927 to M.K.); Russian Foundation of Basic Research (Grant No. 15-54-46009 CT_a to E.G.P.); Danish Natural Science Research Council (Grant No. 4002-00113 to V.L.); Ministerio de Economia y Competitividad (Grant Nos. CTM2017-88080 [AEI/FEDER, UE] and CGL2013-42432-P to M.P. and M.P.G.G., respectively); Centre National de la Recherche Scientifique (Grant No. UMR 9191 to C.M.-M.); Vetenskapsrådet (Grant Nos. 2011-05679 and 2015-04680 to J.A.); Emmy Noether Programme of the DFG (Grant No. PO 1648/3-1 to N.P.); National Institute of Health (Grant Nos. R35GM119686 and R01GM100366 to A.O.B. and P.S.S., respectively); Scientific and Technological Research Council of Turkey (TUBITAK) (Grant No. 214Z238 to B.S.O.); Agence Nationale de la Recherche Exhyb (Grant No. 14-CE19-0016 to C.V.); FP6, Network of Excellence LifeSpan (Grant No. FP6 036894 to B.J.Z.); FP7, IDEAL (Grant No. FP7/2007-2011/259679 to B.J.Z.); Israel Science Foundation (Grant No. 1737/17 to E.T.); Deutsche Forschungsgemeinschaft (Grant No. PA 903/

8-1 to J.P.); Austrian Science Fund (FWF) (Grant Nos. P32275 and P27048 to M.K. and A.J.B., respectively); Biotechnology and Biological Sciences Research Council (BBSRC) (Grant No. BB/P00685X/1 to A.J.B.); Swiss National Science Foundation (SNSF) (Grant Nos. PP00P3_133641, PP00P3_165836, and 31003A_182262 to T.F.); European Commission (Grant No. H2020-ERC-2014CoG-647900 to J.G.); Secretaria d'Universitats i Recerca, Department Economia i Coneixement, Generalitat de Catalunya (Grant No. GRC 2017 SGR 880 to J.G.); Ministerio de Economia y Competitividad/FEDER (Grant No. BFU2014-57779-P to J.G.); and Ministerio de Ciencia e Innovación/AEI/FEDER, EU (Grant No. BFU2017-82937-P to J.G.)

References

Adrian AB, Comeron JM. 2013. The *Drosophila* early ovarian transcriptome provides insight to the molecular causes of recombination rate variation across genomes. *BMC Genomics* 14(1):794.

Adrion JR, Hahn MW, Cooper BS. 2015. Revisiting classic clines in *Drosophila melanogaster* in the age of genomics. *Trends Genet.* 31(8):434–444.

Akashi H. 1995. Inferring weak selection from patterns of polymorphism and divergence at "silent" sites in *Drosophila* DNA. *Genetics* 139:1067–1076.

Anderson AR, Hoffmann AA, McKechnie SW, Umina PA, Weeks AR. 2005. The latitudinal cline in the *In(3R)Payne* inversion polymorphism has shifted in the last 20 years in Australian *Drosophila melanogaster* populations. *Mol Ecol.* 14(3):851–858.

Arguello JR, Laurent S, Clark AG. 2019. Demographic history of the human commensal *Drosophila melanogaster*. *Genome Biol Evol.* 11(3):844–854.

Aulard S, David JR, Lemeunier F. 2002. Chromosomal inversion polymorphism in Afrotropical populations of *Drosophila melanogaster*. *Genet Res.* 79(1):49–63.

Barata A, Santos SC, Malfeito-Ferreira M, Loureiro V. 2012. New insights into the ecological interaction between grape berry microorganisms and *Drosophila* flies during the development of sour rot. *Microb Ecol.* 64(2):416–430.

Bartolomé C, Maside X, Charlesworth B. 2002. On the abundance and distribution of transposable elements in the genome of *Drosophila melanogaster*. *Mol Biol Evol.* 19(6):926–937.

Becher PG, Flick G, Rozpędowska E, Schmidt A, Hagman A, Lebreton S, Larsson MC, Hansson BS, Piškur J, Witzgall P, et al. 2012. Yeast, not fruit volatiles mediate *Drosophila melanogaster* attraction, oviposition and development. *Funct Ecol.* 26(4):822–828.

Begun DJ, Holloway AK, Stevens K, Hillier LW, Poh Y-P, Hahn MW, Nista PM, Jones CD, Kern AD, Dewey CN, et al. 2007. Population genomics: whole-genome analysis of polymorphism and divergence in *Drosophila simulans*. *PLoS Biol.* 5(11):e310.

Behrman EL, Howick VM, Kapun M, Staubach F, Bergland AO, Petrov DA, Lazzaro BP, Schmidt PS. 2018. Rapid seasonal evolution in innate immunity of wild *Drosophila melanogaster*. *Proc R Soc B.* 285(1870):20172599.

Beisswanger S, Stephan W, De Lorenzo D. 2006. Evidence for a selective sweep in the *wapl* region of *Drosophila melanogaster*. *Genetics* 172(1):265–274.

Bergland AO, Behrman EL, O'Brien KR, Schmidt PS, Petrov DA. 2014. Genomic evidence of rapid and stable adaptive oscillations over seasonal time scales in *Drosophila*. *PLoS Genet.* 10(11):e1004775.

Bergland AO, Tobler R, González J, Schmidt P, Petrov D. 2016. Secondary contact and local adaptation contribute to genome-wide patterns of clinal variation in *Drosophila melanogaster*. *Mol Ecol.* 25(5):1157–1174.

Bergman CM, Quesneville H, Anxolabehere D, Ashburner M. 2006. Recurrent insertion and duplication generate networks of transposable element sequences in the *Drosophila melanogaster* genome. *Genome Biol.* 7(11):R112.

Bilder D, Irvine KD. 2017. Taking Stock of the Drosophila Research Ecosystem. Genetics 206(3):1227–1236.

Blumenstiel JP, Chen X, He M, Bergman CM. 2014. An age-of-allele test of neutrality for transposable element insertions. *Genetics* 196(2):523–538.

Boitard S, Kofler R, Françoise P, Robelin D, Schlötterer C, Futschik A. 2013. Pool-hmm: a Python program for estimating the allele frequency spectrum and detecting selective sweeps from next generation sequencing of pooled samples. *Mol Ecol Resour.* 13(2):337–340.

Boitard S, Schlötterer C, Nolte V, Pandey RV, Futschik A. 2012. Detecting selective sweeps from pooled next-generation sequencing samples. *Mol Biol Evol.* 29(9):2177–2186.

Boussy IA, Itoh M, Itoh M, Rand D, Woodruff RC. 1998. Origin and decay of the P element-associated latitudinal cline in Australian *Drosophila melanogaster*. *Genetica* 104(1):45–57.

Božičević V, Hutter S, Stephan W, Wollstein A. 2016. Population genetic evidence for cold adaptation in European *Drosophila melanogaster* populations. *Mol Ecol.* 25(5):1175–1191.

Buser CC, Newcomb RD, Gaskett AC, Goddard MR. 2014. Niche construction initiates the evolution of mutualistic interactions. *Ecol Lett.* 17(10):1257–1264.

Caracristi G, Schlötterer C. 2003. Genetic differentiation between American and European *Drosophila melanogaster* populations could be attributed to admixture of African alleles. *Mol Biol Evol.* 20(5):792–799.

Cavalli-Sforza LL. 1966. Population structure and human evolution. *Proc R Soc Lond B Biol Sci.* 164(995):362–379.

Chandler JA, Eisen JA, Kopp A. 2012. Yeast communities of diverse *Drosophila* species: comparison of two symbiont groups in the same hosts. *Appl Environ Microbiol.* 78(20):7327–7336.

Chandler JA, James PM. 2013. Discovery of trypanosomatid parasites in globally distributed *Drosophila* species. *PLoS One* 8(4):e61937.

Charlesworth B, Sniegowski P, Stephan W. 1994. The evolutionary dynamics of repetitive DNA in eukaryotes. *Nature* 371(6494):215–220.

Cheng C, White BJ, Kamdem C, Mockaitis K, Costantini C, Hahn MW, Besansky NJ. 2012. Ecological genomics of *Anopheles gambiae* along a latitudinal cline: a population-resequencing approach. *Genetics* 190(4):1417–1432.

Clemente F, Vogl C. 2012. Unconstrained evolution in short introns? – An analysis of genome-wide polymorphism and divergence data from *Drosophila*. *J Evol Biol.* 25(10):1975–1990.

Comeron JM, Ratnappan R, Bailin S. 2012. The many landscapes of recombination in Drosophila melanogaster. *PLoS Genet.* 8(10):e1002905.

Cooper BS, Burrus CR, Ji C, Hahn MW, Montooth KL. 2015. Similar efficacies of selection shape mitochondrial and nuclear genes in both *Drosophila melanogaster* and *Homo sapiens*. *G3 (Bethesda)* 5:2165–2176.

Cridland JM, Macdonald SJ, Long AD, Thornton KR. 2013. Abundance and distribution of transposable elements in two *Drosophila* QTL mapping resources. *Mol Biol Evol.* 30(10):2311–2327.

Daborn PJ, Yen JL, Bogwitz MR. 2002. A single p450 allele associated with insecticide resistance in *Drosophila*. *Science* 297(5590):2253–2256.

Danecek P, Auton A, Abecasis G, Albers CA, Banks E, DePristo MA, Handsaker RE, Lunter G, Marth GT, Sherry ST, et al.. 2011. The variant call format and VCFtools. *Bioinformatics* 27(15):2156–2158.

David JR, Capy P. 1988. Genetic variation of *Drosophila melanogaster* natural populations. *Trends Genet.* 4(4):106–111.

de Jong G, Bochdanovits Z. 2003. Latitudinal clines in *Drosophila melanogaster*: body size, allozyme frequencies, inversion frequencies, and the insulin-signalling pathway. *J Genet.* 82(3):207–223.

Dobzhansky T. 1970. Genetics of the evolutionary process. New York: Columbia University Press.

Duchen P, Zivkovic D, Hutter S, Stephan W, Laurent S. 2013. Demographic inference reveals African and European admixture in the North American *Drosophila melanogaster* population. *Genetics* 193(1):291–301.

Durmaz E, Benson C, Kapun M, Schmidt P, Flatt T. 2018. An inversion supergene in *Drosophila* underpins latitudinal clines in survival traits. *J Evol Biol*. 31(9):1354–1364.

Durmaz E, Rajpurohit S, Betancourt N, Fabian DK, Kapun M, Schmidt P, Flatt T. 2019. A clinal polymorphism in the insulin signaling transcription factor *foxo* contributes to life-history adaptation in *Drosophila*. *Evolution* 73(9):1774–1792.

Elya C, Lok TC, Spencer QE, McCausland H, Martinez CC, Eisen MB. 2018. Robust manipulation of the behavior of *Drosophila melanogaster* by a fungal pathogen in the laboratory. *eLife* 7:e34414.

Fabian DK, Kapun M, Nolte V, Kofler R, Schmidt PS, Schlötterer C, Flatt T. 2012. Genome-wide patterns of latitudinal differentiation among populations of *Drosophila melanogaster* from North America. *Mol Ecol*. 21(19):4748–4769.

Fiston-Lavier A-S, Barrón MG, Petrov DA, González J. 2015. T-lex2: genotyping, frequency estimation and re-annotation of transposable elements using single or pooled next-generation sequencing data. *Nucleic Acids Res*. 43(4):e22.

Fiston-Lavier A-S, Singh ND, Lipatov M, Petrov DA. 2010. *Drosophila melanogaster* recombination rate calculator. *Gene* 463(1–2):18–20.

Francalacci P, Sanna D. 2008. History and geography of human Y-chromosome in Europe: a SNP perspective. *J Anthropol Sci*. 86:59–89.

Futschik A, Schlötterer C. 2010. The next generation of molecular markers from massively parallel sequencing of pooled DNA samples. *Genetics* 186(1):207–218.

González J, Karasov TL, Messer PW, Petrov DA. 2010. Genome-wide patterns of adaptation to temperate environments associated with transposable elements in *Drosophila*. *PLoS Genet*. 6(4):e1000905.

González J, Lenkov K, Lipatov M, Macpherson JM, Petrov DA. 2008. High rate of recent transposable element–induced adaptation in *Drosophila melanogaster*. *PLoS Biol*. 6(10):e251.

Haas F, Brodin A. 2005. The Crow *Corvus corone* hybrid zone in southern Denmark and northern Germany. *Ibis*. 147(4):649–656.

Haddrill PR, Charlesworth B, Halligan DL, Andolfatto P. 2005. Patterns of intron sequence evolution in *Drosophila* are dependent upon length and GC content. *Genome Biol*. 6(8):R67.

Hales KG, Korey CA, Larracuente AM, Roberts DM. 2015. Genetics on the fly: a primer on the *Drosophila* model system. *Genetics* 201(3):815–842.

Halligan DL, Keightley PD. 2006. Ubiquitous selective constraints in the *Drosophila* genome revealed by a genome-wide interspecies comparison. *Genome Res*. 16(7):875–884.

Hamilton PT, Votýpka J, Dostálová A, Yurchenko V, Bird NH, Lukeš J, Lemaitre B, Perlman SJ. 2015. Infection Dynamics and Immune Response in a Newly Described Drosophila-*Trypanosomatid* Association. *mBio*. 6(5):e01356–01315.

Harpur BA, Kent CF, Molodtsova D, Lebon JMD, Alqarni AS, Owayss AA, Zayed A. 2014. Population genomics of the honey bee reveals strong signatures of positive selection on worker traits. *Proc Natl Acad Sci U S A*. 111(7):2614–2619.

Haselkorn TS, Markow TA, Moran NA. 2009. Multiple introductions of the *Spiroplasma* bacterial endosymbiont into *Drosophila*. *Mol Ecol*. 18(6):1294–1305.

Haudry A, Laurent S, Kapun M. 2020. Population genomics on the fly: recent advances in Drosophila. In: Dutheil JY, editor. Statistical population genomics. Vol. 2090. New York: Springer. p. 357–396.

Hewitt GM. 1999. Post-glacial re-colonization of European biota. *Biol J Linn Soc*. 68(1–2):87–112.

Hijmans RJ, Cameron SE, Parra JL, Jones PG, Jarvis A. 2005. Very high resolution interpolated climate surfaces for global land areas. *Int J Climatol*. 25(15):1965–1978.

Hoffmann AA, Weeks AR. 2007. Climatic selection on genes and traits after a 100 year-old invasion: a critical look at the temperate-tropical clines in *Drosophila melanogaster* from eastern Australia. *Genetica* 129(2):133–147.

Hohenlohe PA, Bassham S, Etter PD, Stiffler N, Johnson EA, Cresko WA. 2010. Population genomics of parallel adaptation in threespine stickleback using sequenced RAD tags. *PLoS Genet*. 6(2):e1000862.

Huang DW, Sherman BT, Lempicki RA. 2009. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat Protoc*. 4(1):44–57.

Hudson RR, Kreitman M, Aguadé M. 1987. A test of neutral molecular evolution based on nucleotide data. *Genetics* 116(1):153–159.

Hutter S, Li H, Beisswanger S, De Lorenzo D, Stephan W. 2007. Distinctly different sex ratios in African and European populations of *Drosophila melanogaster* inferred from chromosomewide single nucleotide polymorphism data. *Genetics* 177(1):469–480.

Kaminker JS, Bergman CM, Kronmiller B, Carlson J, Svirskas R, Patel S, Frise E, Wheeler DA, Lewis SE, Rubin GM, et al. 2002. The transposable elements of the *Drosophila melanogaster* euchromatin: a genomics perspective. *Genome Biol*. 3(12):research0084.

Kao JY, Zubair A, Salomon MP, Nuzhdin SV, Campo D. 2015. Population genomic analysis uncovers African and European admixture in *Drosophila melanogaster* populations from the southeastern United States and Caribbean Islands. *Mol Ecol*. 24(7):1499–1509.

Kapopoulou A, Kapun M, Pavlidis P, et al. 2018. Early split between African and European populations of *Drosophila melanogaster*. *bioRxiv*. doi: https://doi.org/10.1101/340422.

Kapopoulou A, Pfeifer S, Jensen J, Laurent S. 2018. The demographic history of African *Drosophila melanogaster*. *Genome Biol Evol*. 10(9):2338–2342.

Kapun M, Fabian DK, Goudet J, Flatt T. 2016. Genomic evidence for adaptive inversion clines in *Drosophila melanogaster*. *Mol Biol Evol*. 33(5):1317–1336.

Kapun M, Flatt T. 2019. The adaptive significance of chromosomal inversion polymorphisms in *Drosophila melanogaster*. *Mol Ecol*. 28(6):1263–1282.

Kapun M, Schmidt C, Durmaz E, Schmidt PS, Flatt T. 2016. Parallel effects of the inversion *In(3R)Payne* on body size across the North American and Australian clines in *Drosophila melanogaster*. *J Evol Biol*. 29(5):1059–1072.

Kapun M, van Schalkwyk H, McAllister B, Flatt T, Schlötterer C. 2014. Inference of chromosomal inversion dynamics from Pool-Seq data in natural and laboratory populations of *Drosophila melanogaster*. *Mol Ecol*. 23(7):1813–1827.

Keller A. 2007. *Drosophila melanogaster*'s history as a human commensal. *Curr Biol*. 17(3):R77–R81.

Kennington JW, Partridge L, Hoffmann AA. 2006. Patterns of diversity and linkage disequilibrium within the cosmopolitan inversion *In(3R)Payne* in *Drosophila melanogaster* are indicative of coadaptation. *Genetics* 172(3):1655–1663.

Kimura M. 1984. The neutral theory of molecular evolution. Cambridge, United Kingdom: Cambridge University Press.

Knibb WR, Oakeshott JG, Gibson JB. 1981. Chromosome inversion polymorphisms in *Drosophila melanogaster*. I. Latitudinal clines and associations between inversions in Australasian populations. *Genetics* 98(4):833–847.

Knief U, Bossu CM, Saino N, Hansson B, Poelstra J, Vijay N, Weissensteiner M, Wolf J. 2019. Epistatic mutations under divergent selection govern phenotypic variation in the crow hybrid zone. *Nat Ecol Evol*. 3(4):570–576.

Kofler R, Betancourt AJ, Schlötterer C. 2012. Sequencing of pooled DNA samples (Pool-Seq) uncovers complex dynamics of transposable element insertions in *Drosophila melanogaster*. *PLoS Genet*. 8(1):e1002487.

Kofler R, Orozco-terWengel P, De Maio N, Pandey RV, Nolte V, Futschik A, Kosiol C, Schlötterer C. 2011. PoPoolation: a toolbox for population genetic analysis of next generation sequencing data from pooled individuals. *PLoS One* 6(1):e15925.

Kofler R, Schlötterer C. 2012. GOwinda: unbiased analysis of gene set enrichment for genome-wide association studies. *Bioinformatics* 28(15):2084–2085.

Kolaczkowski B, Kern AD, Holloway AK, Begun DJ. 2011. Genomic differentiation between temperate and tropical Australian populations of *Drosophila melanogaster*. *Genetics* 187(1):245–260.

Kreitman M. 1983. Nucleotide polymorphism at the *alcohol dehydrogenase* locus of *Drosophila melanogaster*. *Nature* 304(5925):412–417.

Kriesner P, Conner WR, Weeks AR, Turelli M, Hoffmann AA. 2016. Persistence of a *Wolbachia* infection frequency cline in *Drosophila melanogaster* and the possible role of reproductive dormancy. *Evolution* 70(5):979–997.

Lachaise D, Cariou M-L, David JR, et al. 1988. Historical biogeography of the *Drosophila melanogaster* species subgroup. In: Hecht MK, Wallace B, Prance GT, editors. Evolutionary biology. Boston: Springer. p. 159–225.

Lack JB, Cardeno CM, Crepeau MW, Taylor W, Corbett-Detig RB, Stevens KA, Langley CH, Pool JE. 2015. The *Drosophila* Genome Nexus: a population genomic resource of 623 *Drosophila melanogaster* genomes, including 197 from a single ancestral range population. *Genetics* 199(4):1229–1241.

Lack JB, Lange JD, Tang AD, Corbett-Detig RB, Pool JE. 2016. A thousand fly genomes: an expanded *Drosophila* Genome Nexus. *Mol Biol Evol*. 33(12):3308–3313.

Langley CH, Stevens K, Cardeno C, Lee YCG, Schrider DR, Pool JE, Langley SA, Suarez C, Corbett-Detig RB, Kolaczkowski B, et al. 2012. Genomic variation in natural populations of *Drosophila melanogaster*. *Genetics* 192(2):533–598.

Lawrie DS, Messer PW, Hershberg R, Petrov DA. 2013. Strong purifying selection at synonymous sites in *D. melanogaster*. *PLoS Genet*. 9(5):e1003527.

Lemeunier F, Aulard S. 1992. Inversion polymorphism in *Drosophila melanogaster*. In: Krimbas CB, Powell JR, editors. Drosophila inversion polymorphism. New York: CRC Press. p. 339–405.

Lerat E, Goubert C, Guirao-Rico S, Merenciano M, Dufour A-B, Vieira C, González J. 2019. Population-specific dynamics and selection patterns of transposable element insertions in European natural populations. *Mol Ecol*. 28(6):1506–1522.

Lewontin RC. 1974. The genetic basis of evolutionary change. New York: Columbia University Press.

Li H, Stephan W. 2006. Inferring the demographic history and rate of adaptive substitution in *Drosophila*. *PLoS Genet*. 2(10):e166.

Machado H, Bergland AO, Taylor R. 2019. Broad geographic sampling reveals predictable, pervasive, and strong seasonal adaptation in *Drosophila*. *bioRxiv*. doi: https://doi.org/10.1101/337543.

Machado HE, Bergland AO, O'Brien KR, Behrman EL, Schmidt PS, Petrov DA. 2016. Comparative population genomics of latitudinal variation in *Drosophila simulans* and *Drosophila melanogaster*. *Mol Ecol*. 25(3):723–740.

Macholán M, Baird SJ, Munclinger P, Dufková P, Bímová B, Piálek J. 2008. Genetic conflict outweighs heterogametic incompatibility in the mouse hybrid zone? *BMC Evol Biol*. 8(1):271.

Martino ME, Ma D, Leulier F. 2017. Microbial influence on *Drosophila* biology. *Curr Opin Microbiol*. 38:165–170.

Mateo L, Rech GE, González J. 2018. Genome-wide patterns of local adaptation in Western European *Drosophila melanogaster* natural populations. *Sci Rep*. 2018;8:16143.

McDonald JH, Kreitman M. 1991. Adaptive protein evolution at the *Adh* locus in *Drosophila*. *Nature* 351(6328):652–654.

Mettler LE, Voelker RA, Mukai T. 1977. Inversion clines in populations of *Drosophila melanogaster*. *Genetics* 87(1):169–176.

Meyer F, Paarmann D, D'Souza M, Olson R, Glass EM, Kubal M, Paczian T, Rodriguez A, Stevens R, Wilke A, et al. 2008. The metagenomics RAST server – a public resource for the automatic phylogenetic and functional analysis of metagenomes. *BMC Bioinformatics* 9(1):386.

Michalakis Y, Veuille M. 1996. Length variation of CAG/CAA trinucleotide repeats in natural populations of *Drosophila melanogaster* and its relation to the recombination rate. *Genetics* 143:1713–1725.

Nielsen R, Akey JM, Jakobsson M, Pritchard JK, Tishkoff S, Willerslev E. 2017. Tracing the peopling of the world through genomics. *Nature* 541(7637):302–310.

Novembre J, Johnson T, Bryc K, Kutalik Z, Boyko AR, Auton A, Indap A, King KS, Bergmann S, Nelson MR, et al. 2008. Genes mirror geography within Europe. *Nature* 456(7218):98–101.

Paaby AB, Bergland AO, Behrman EL, Schmidt PS. 2014. A highly pleiotropic amino acid polymorphism in the *Drosophila* insulin receptor contributes to life-history adaptation. *Evolution* 68(12):3395–3409.

Paaby AB, Blacket MJ, Hoffmann AA, Schmidt PS. 2010. Identification of a candidate adaptive polymorphism for *Drosophila* life history by parallel independent clines on two continents. *Mol Ecol*. 19(4):760–774.

Palmer WH, Medd NC, Beard PM, Obbard DJ. 2018. Isolation of a natural DNA virus of Drosophila melanogaster, and characterisation of host resistance and immune responses. *PLoS Pathog*. 14(6):e1007050.

Parsch J, Novozhilov S, Saminadin-Peter SS, Wong KM, Andolfatto P. 2010. On the utility of short intron sequences as a reference for the detection of positive and negative selection in *Drosophila*. *Mol Biol Evol*. 27(6):1226–1234.

Peel MC, Finlayson BL, McMahon TA. 2007. Updated world map of the Köppen-Geiger climate classification. *Hydrol Earth Syst Sci*. 11(5):1633–1644.

Petrov DA, Fiston-Lavier AS, Lipatov M, Lenkov K, González J. 2011. Population genomics of transposable elements in *Drosophila melanogaster*. *Mol Biol Evol*. 28(5):1633–1644.

Pool JE, Braun DT, Lack JB. 2016. Parallel evolution of cold tolerance within *Drosophila melanogaster*. *Mol Biol Evol*. 34:349–360.

Pool JE, Corbett-Detig RB, Sugino RP, Stevens KA, Cardeno CM, Crepeau MW, Duchen P, Emerson JJ, Saelao P, Begun DJ, et al. 2012. Population genomics of sub-Saharan *Drosophila melanogaster*: African diversity and non-African admixture. *PLoS Genet*. 8(12):e1003080.

Powell JR. 1997. Progress and prospects in evolutionary biology: the Drosophila model. New York and Oxford: Oxford University Press.

Quesneville H, Bergman CM, Andrieu O, Autard D, Nouaud D, Ashburner M, Anxolabehere D. 2005. Combined evidence annotation of transposable elements in genome sequences. *PLoS Comp Biol*. 1(2):e22.

Rako L, Anderson AR, Sgrò CM, Stocker AJ, Hoffmann AA. 2006. The association between inversion *In(3R)Payne* and clinally varying traits in *Drosophila melanogaster*. *Genetica* 128(1–3):373–384.

Rane RV, Rako L, Kapun M, Lee SF, Hoffmann AA. 2015. Genomic evidence for role of inversion *3RP* of *Drosophila melanogaster* in facilitating climate change adaptation. *Mol Ecol*. 24(10):2423–2432.

Rech GE, Bogaerts-Márquez M, Barrón MG, Merenciano M, Villanueva-Cañas JL, Horváth V, Fiston-Lavier A-S, Luyten I, Venkataram S, Quesneville H, et al. 2019. Stress response, behavior, and development are shaped by transposable element-induced mutations in *Drosophila*. *PLoS Genet*. 15(2):e1007900.

Richardson MF, Weinert LA, Welch JJ, Linheiro RS, Magwire MM, Jiggins FM, Bergman CM. 2012. Population genomics of the *Wolbachia* endosymbiont in *Drosophila melanogaster*. *PLoS Genet*. 8(12):e1003129.

Rogers RL, Hartl DL. 2012. Chimeric genes as a source of rapid evolution in *Drosophila melanogaster*. *Mol Biol Evol*. 29(2):517–529.

Schlötterer C, Tobler R, Kofler R, Nolte V. 2014. Sequencing pools of individuals – mining genome-wide polymorphism data without big funding. *Nat Rev Genet*. 15(11):749–763.

Schmidt PS, Paaby AB. 2008. Reproductive diapause and life-history clines in north American populations of *Drosophila melanogaster*. *Evolution* 62(5):1204–1215.

Schmidt PS, Zhu C-T, Das J, Batavia M, Yang L, Eanes WF. 2008. An amino acid polymorphism in the *couch potato* gene forms the basis for climatic adaptation in *Drosophila melanogaster*. *Proc Natl Acad Sci U S A*. 105(42):16207–16211.

Shi M, White VL, Schlub T, Eden J-S, Hoffmann AA, Holmes EC. 2018. No detectable effect of Wolbachia w Mel on the prevalence and abundance of the RNA virome of Drosophila melanogaster. *Proc R Soc B*. 285(1883):20181165.

Singh ND, Arndt PF, Clark AG, Aquadro CF. 2009. Strong evidence for lineage and sequence specificity of substitution rates and patterns in *Drosophila*. *Mol Biol Evol*. 26(7):1591–1605.

Sprengelmeyer QD, Mansourian S, Lange JD, Matute DR, Cooper BS, Jirle EV, Stensmyr MC, Pool JE. 2020. Recurrent collection of *Drosophila*

melanogaster from wild African environments and genomic insights into species history. *Mol Biol Evol*. 37(3):627–638.
Staubach F, Baines JF, Künzel S, Bik EM, Petrov DA. 2013. Host species and environmental effects on bacterial communities associated with *Drosophila* in the laboratory and in the natural environment. *PLoS One* 8(8):e70749.
Szymura JM, Barton NH. 1986. Genetic analysis of a hybrid zone between the fire-bellied toads, *Bombina bombina* and *B. variegata*, near Cracow in Southern Poland. *Evolution* 40(6):1141–1159.
Tauber E, Zordan M, Sandrelli F, Pegoraro M, Osterwalder N, Breda C, Daga A, Selmin A, Monger K, Benna C, et al. 2007. Natural selection favors a newly derived *timeless* allele in *Drosophila melanogaster*. *Science* 316(5833):1895–1899.
Teixeira L, Ferreira A, Ashburner M. 2008. The Bacterial Symbiont Wolbachia Induces Resistance to RNA Viral Infections in Drosophila melanogaster. *PLoS Biol.* 6(12):e2.
Trinder M, Daisley BA, Dube JS, Reid G. 2017. *Drosophila melanogaster* as a high-throughput model for host-microbiota interactions. *Front Microbiol*. 8:751.
Turner TL, Levine MT, Eckert ML, Begun DJ. 2008. Genomic analysis of adaptive differentiation in *Drosophila melanogaster*. *Genetics* 179(1):455–473.
Umina PA, Weeks AR, Kearney MR, McKechnie SW, Hoffmann AA. 2005. A rapid shift in a classic clinal pattern in *Drosophila* reflecting climate change. *Science* 308(5722):691–693.
Unckless RL. 2011. A DNA virus of *Drosophila*. *PLoS One* 6(10):e26564.
Villemereuil P, Gaggiotti OE. 2015. A new FST-based method to uncover local adaptation using environmental variables. *Methods Ecol Evol*. 6(11):1248–1258.
Walters AW, Hughes RC, Call TB, Walker CJ, Wilcox H, Peterson SC, Rudman SM, Newell PD, Douglas AE, Schmidt PS, et al. 2020. The microbiota influences the *Drosophila melanogaster* life history strategy. *Mol Ecol*. 29:639–653.
Wang Y, Kapun M, Waidele L, Kuenzel S, Bergland AO, Staubach F. 2020. Common structuring principles of the *Drosophila melanogaster* microbiome on a continental scale and between host and substrate. *Environ Microbiol Rep*. 12(2):220–228.
Wang Y, Staubach F. 2018. Individual variation of natural *D. melanogaster*-associated bacterial communities. *FEMS Microbiol Lett*. 365(6):fny017.
Webster CL, Longdon B, Lewis SH, Obbard DJ. 2016. Twenty-five new viruses associated with the Drosophilidae (Diptera). *Evol Bioinform Online*. 12:13–25.
Webster CL, Waldron FM, Robertson S, Crowson D, Ferrari G, Quintana JF, Brouqui J-M, Bayne EH, Longdon B, Buck AH, et al. 2015. The discovery, distribution, and evolution of viruses associated with *Drosophila melanogaster*. *PLoS Biol*. 13(7):e1002210.
Werren JH, Baldo L, Clark ME. 2008. *Wolbachia*: master manipulators of invertebrate biology. *Nat Rev Microbiol*. 6(10):741–751.
Whitlock MC, McCauley DE. 1999. Indirect measures of gene flow and migration: $F_{ST} \neq 1/(4Nm+1)$. *Heredity* 82(2):117–125.
Wilfert L, Longdon B, Ferreira AGA, Bayer F, Jiggins FM. 2011. Trypanosomatids are common and diverse parasites of *Drosophila*. *Parasitology* 138(7):858–865.
Wolff JN, Camus MF, Clancy DJ, Dowling DK. 2016. Complete mitochondrial genome sequences of thirteen globally sourced strains of fruit fly (*Drosophila melanogaster*) form a powerful model for mitochondrial research. *Mitochondrial DNA A*. 27(6):4672–4674.
Wright S. 1949. The genetical structure of populations. *Ann Eugen*. 15(1):323–354.
Xiao F-X, Yotova V, Zietkiewicz E, Lovell A, Gehl D, Bourgeois S, Moreau C, Spanaki C, Plaitakis A, Moisan J-P, et al. 2004. Human *X*-chromosomal lineages in Europe reveal Middle Eastern and Asiatic contacts. *Eur J Hum Genet*. 12(4):301–311.
Yukilevich R, True JR. 2008a. Incipient sexual isolation among cosmopolitan *Drosophila melanogaster* populations. *Evolution* 62(8):2112–2121.
Yukilevich R, True JR. 2008b. African morphology, behavior and pheromones underlie incipient sexual isolation between us and Caribbean *Drosophila melanogaster*. *Evolution* 62(11):2807–2828.
Zanini F, Brodin J, Thebo L, Lanz C, Bratt G, Albert J, Neher RA. 2015. Population genomics of intrapatient HIV-1 evolution. *eLife* 4:e11282.