NGS approach for investigating evolutionary transition form oviparity to viviparity in squamate reptiles.

Luca Cornetti[1,2], Andrea Benazzo[2], Alex Panziera[2], Mike W. Bruford[3], Giorgio Bertorelle[2], Cristiano Vernesi[1]

[1] Dept. of Biodiversity and Molecular Ecology, Research and Innovation Center, Edmund Mach Foundation, Trento, Italy; [2] Department of Evolutionary Biology, University of Ferrara, Italy; [3] School of Biosciences, Cardiff University, Cardiff, UK

Introduction

The lizard *Zootoca vivipara* (Figure 1) is one on the few example in Nature which shows, within the same species, populations with different reproductive modalities. Oviparous populations live in the southern part of its distributional range (the newly discovered *Z.v.carniolica* in Eastern-Italian Alps and *Z.v.louislantzi* in the Pyrenees), while viviparous subspecies (e.g. *Z.v.vivipara* and *Z.v.sachalinensis*) are widely distributed from British Isles and central France to Scandinavia and north-eastern Asia (Figure 2) [1]. This species is, therefore, particularly well suited for studying the evolutionary shift in reproductive mode.

Results and Discussion

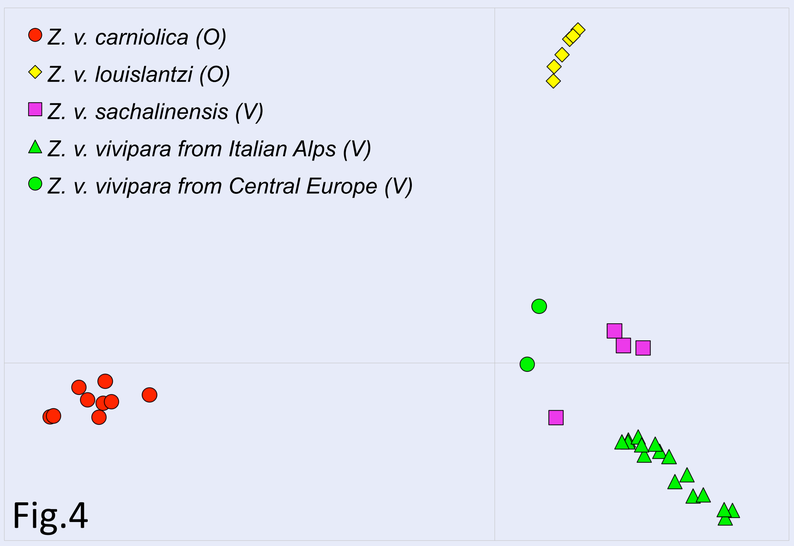


Fig.4 MultiDimensional Scaling based on genetic distances between individuals according to 87k SNPs.

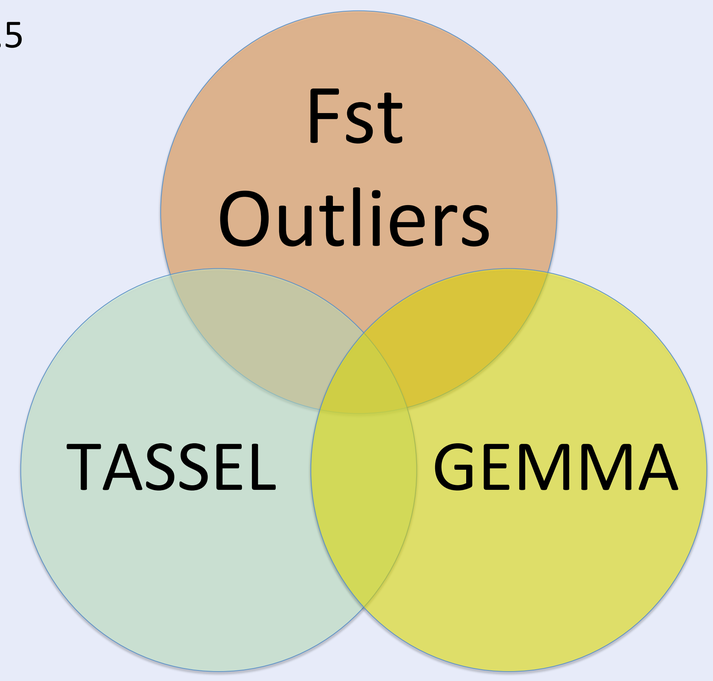
Fig.5 GWAS approach. Putative SNPs (217) were chosen among 4908 markers that showed significant signal of selection in at least two methods.

Bioinformatic analysis performed using the pipeline software Stacks produced about 100.000 local alignments and about 260.000 Single Nucleotide Polymorphisms (SNPs). 87.385 SNPs, selected from the 75bp single-end reads showing no more than 5 SNPs, were used to describe the overall genetic variation between subspecies. The result of MDS (Figure 2) seems to confirm (according to [1]) the existence of two oviparous clades (*Z.v.carniolica* and *Z.v.louislantzi*) and one viviparous clade (composed by *Z.v.vivipara* and *Z.v.sachalinensis*).
We additionally analyzed a restricted marker-set (4908 SNP) in order to have at least 50% of present data in each clade, for trying to investigate the genetic basis of evolution of viviparity dividing our dataset according to reproductive mode.
We adopted three different approaches (Figure 5):

(1) Fst Outlier (intersection of markers with paiwise Fst ≥ 0.5 between *vivipara* (V) and *carniolica* (O) and *vivipara* (V) and *lousilantzi* (O) and Fst ≤ 0.05 within *vivipara* (V));
(2) Tassel (genotype-phenotype association using Mixed Linear Model) [3];
(3) Gemma (software implementing the Genome-wide Effcient Mixed Model Association algorithm)[4].

We selected 217 SNPs that were identified in at least two methods. Genomic sequences (200-500 bp long, achieved with Illumina Paired-end protocol) physically linked to these markers were then BLASTed against the *Anolis Carolinensis* genome [5] (the only reptile genome available at the moment; *A. carolinensis* and *Z. vivipara* share the most recent common ancestor about 180 mya).

Fig.5



Genes (predicted proteins)	Function
Immunoglobulin light chain	Immune system
Suppressor of cytokine signaling	
V-set and immunoglobulin domain	
Immunoglobulin superfamily member	
Interferon induced protein	
Vasopressin V1A	Vascularization
Angiopoietin-related protein	
Progesterone binding factor	Hormone receptor

Tab. 1 Examples of identified genes possibly related to reproductive mode

Conclusions

Looking for genes and markers showing signals of selection is becoming relatively straightforward with the advent of NGS; RADtag, together with Paired-end sequencing, is a useful method to assemble millions of genomic reads into contigs which can be compared to known sequences in existing databases. However, not all genes can be identified, since non-model species may contain unknown genes or the closest reference genome may not be so close.
We analyzed contigs physically linked to 217 SNPs which showed signal of selection and we found sequence similarities in about 60% of them (max E-value 0.5, against *A. carolinensis* genome). Among the predicted proteins observed, eight genes could be possibly related to reproductive mode according to previous studies that focused on different physiological pathways that differ between oviparous and viviparous reproductive mode [8].
Evolution of viviparity poses a major immunological hurdle for mother and foetus. For instance, cytokines and other proteins involved in immunological response seem to play a similar role in mammals and viviparous squamates. Hormone receptors have an essential part in evolution of viviparity, in fact they may regulate follicular development and oocyte maturation. Also proteins involved in placental development and vascularization are peculiar elements that characterize mammals and viviparous squamates. Examples of these gene categories were identified in this study and listed in Table 1.
So far, only gene-by-gene or protein-by-protein approaches have been taken; this study is the first attempt to analyze the oviparity/viviparity transition at genomic level, with the consciousness that this shift is a very complex physiological process, probably mediated by hundreds of genes.

Materials and Methods

A Next Generation Sequencing technique was used to analyse 40 samples of *Zootoca vivipara* to cover the overall genetic variation of the species. RADtag sequencing (Figure 3) uses Illumina HiSeq technology to simultaneously discover and analyse thousands of SNPs at genome level [6]. Bioinformatic analyses have been conducted using the pipeline software Stacks v1.02 [7].

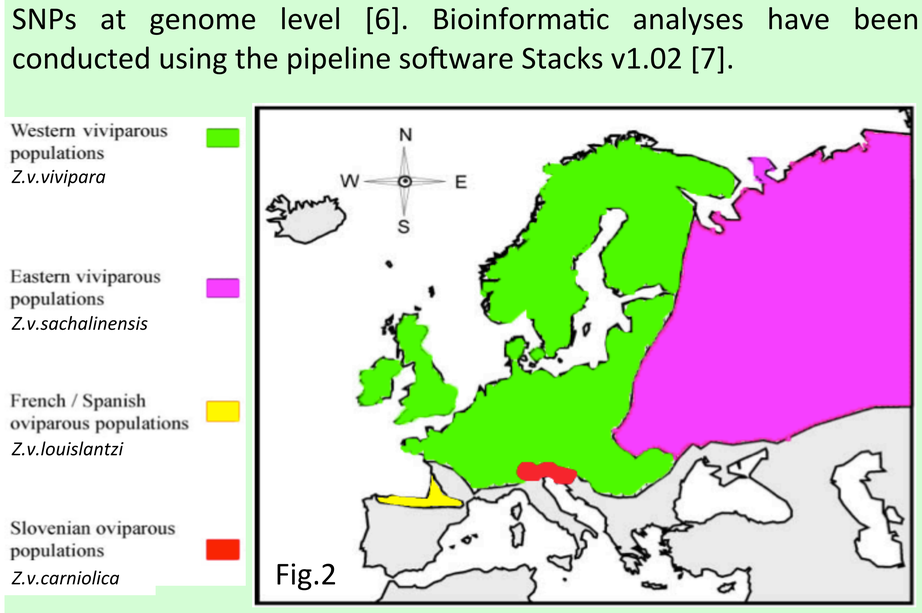


Fig.2 *Zootoca vivipara* European distribution and subspecific pattern (modified from [2])

Fig.1 Adult individual of *Zootoca vivipara* (photo by K. Tabarelli de Fatis)

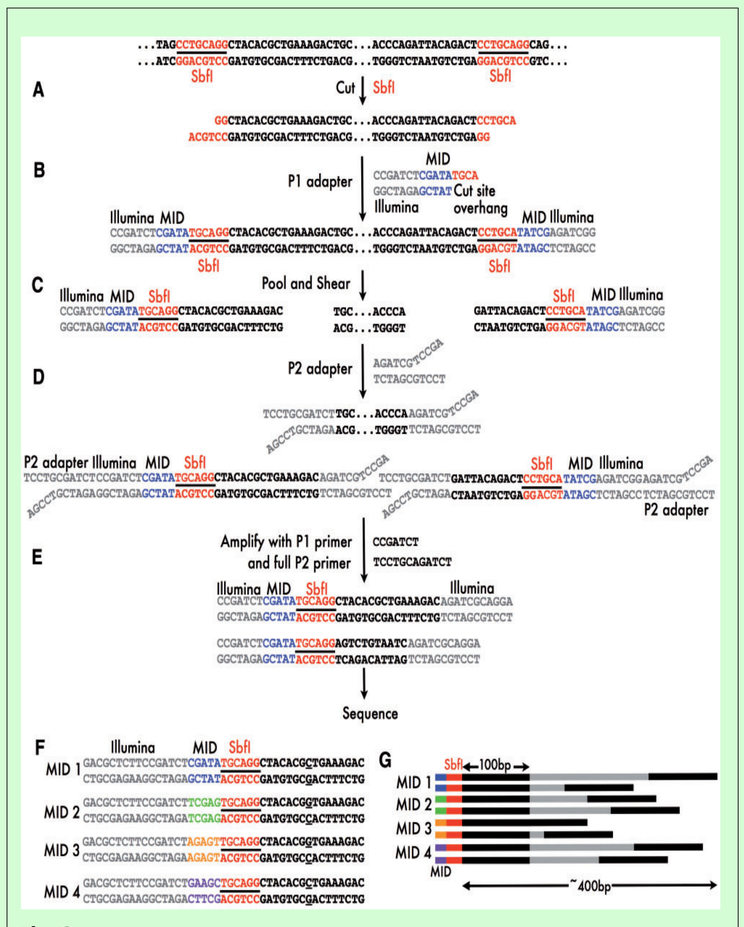


Fig.3 The process of RADSeq. (**A**) Genomic DNA is sheared with a restriction enzyme of choice (SbfI in this example). (**B**) PI adapter is ligated to SbfI-cut fragments. The PI adapter is adapted from the Illumina sequencing adapter (full sequence not shown here), with a molecular identifier (MID; CGATA in this example) and a cut site overhang at the end (TGCA in this example). (**C**) Samples from multiple individuals are pooled together and all fragments are randomly sheared. Only a subset of the resulting fragments contains restriction sites and PI adapters. (**D**) P2 adapter is ligated to all fragments. The P2 adapter has a divergent end. (**E**) PCR amplification with PI and P2 primers. The P2 adapter will be completed only in the fragments ligated with PI adapter, and so only these fragments will be fully amplified. (**F**) Pooled samples with different MIDs are separated bioinformatically and SNPs called (C/G SNP underlined). (**G**) As fragments are sheared randomly, paired end sequences from each sequenced fragment will cover a 300–400 bp region downstream of the restriction site.

Acknowledgments

We would like to thank Autonomous Province of Trento for founding the ACE-SAP project

References

[1] Surget-Groba Y et al., (2001). Intraspecific Phylogeography of *Lacerta vivipara* and the Evolution of Viviparity. *Mol. Phylogenet. Evol.*18, 449-459.
[2] Surget-Groba Y et al., (2002). Phylogeography and conservation of the populations of *Zootoca vivipara carniolica*. *Biol. Cons*. 106, 365-372.
[3] Bradbury PJ et al. (2007) TASSEL: software for association mapping of complex traits in diverse samples. *Bioinformatics* 23, 2633-2635.
[4] Zhou X & Stephens M (2012). Genome-wide effcient mixed-model analysis for association studies. Nature Genetics. 44, 821-824.
[5] Alfoldi J et al., (2011). The genome of the green anole lizard and a comparative analysis with birds and mammals. *Nature*. 477, 587–591
[6] Davey JW, Blaxter ML (2011) RADSeq: next generation population genetics. *Brief Funct Genomics*. 9, 416-423
[7] Catchen J et al., Stacks: an analysis tool set for population genomics (2013). *Mol. Ecol*. 22, 3124-3140
[8] Murphy BF & Thompson MB.(2011) A review of the evolution of viviparity in squamate reptiles: the past, present and future role of molecular biology and genomics. *J Comp Physiol B Biochem Syst Environ Physiol*. 181, 575-594.