

Using demographic inference methods to reveal invisible speciation events

Coordination: Damien de Vienne, Laure Ségurel, Sylvain Mousset, Carina Mugal

In 2011, a new method (PSMC, [1]) revolutionized the study of population demographic history. It allows estimating past effective population size variations from the genome of a single diploid individual (Figure 1). The method involves segmenting the genome into coalescent age segments (based on the level of heterozygosity), then examining the frequency of the different ages. If many segments in the genome coalesce at the same time, it indicates that the effective population size was low at that time. Since then, several studies have shown that factors unrelated to changes in effective population size, such as population structure or gene flow, can affect this method [2,3], and various improvements have been proposed (e.g., [4]).

Given that population structure can create a signal in demographic inferences, it is legitimate to ask whether speciation events in the past — often associated with structuration of population and a more or less rapid decline in gene flow between new populations — could also produce such a signal, which would be mistakenly

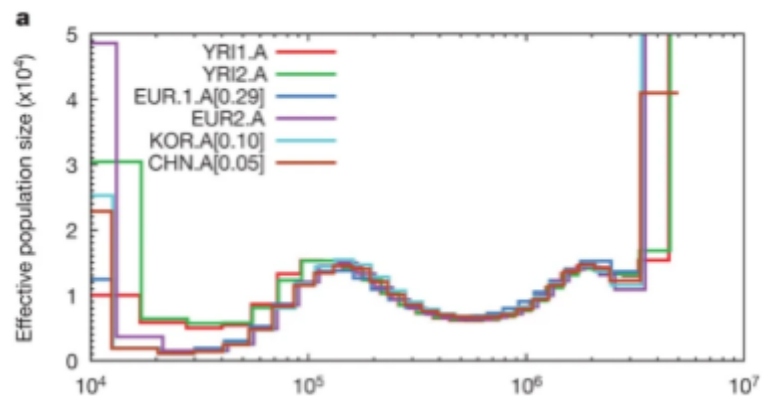


Figure 1. Example of a PSMC plot. The population sizes (y-axis) were inferred for 6 individuals at different times in the past (x-axis). Copied from [1].

interpreted as a change in effective population size. If this was the case, it would not be necessary to observe the two daughter species resulting from a speciation event to detect this event: we could detect "speciation scars" in the phylogenetic trees of the studied species, which would be a significant advancement in phylogenomics and beyond.

This Master project (M1 or M2) aims to explore this question through two parallel approaches: (1) forward-in-time simulations with the tool SLIM [5] to determine the conditions (parameters associated with speciation) under which a speciation event leaves a detectable signal by methods like PSMC; (2) empirical analysis of biological data (six species of *Fidecula*) to see if the inferred

variations in effective population size with PSMC (and other methods) coincide with speciation events visible in the phylogeny.

The consortium: Laure Segurel and Carina Mugal have used (and continue to use) the PSMC method and other demographic history inference methods in several of their projects. Carina Mugal has data and expertise on six bird species of the genus *Fidecula*, which will be used in the studies. Sylvain Mousset has extensive expertise in theoretical population genomics and is proficient with the simulation tool SLIM. Finally, Damien de Vienne is specialized in the study of signals (and associated biases) left by unknown lineages (*ghosts*) in biological data, and the possibility of detecting, quantifying, and placing them in the tree of life through several complementary approaches (gene transfers, metagenomics, genomics, ...).

Bibliography

- [1] Li H, Durbin R (2011) Inference of human population history from individual whole genome sequences. *Nature* 475: 493–496. doi: 10.1038/nature10231
- [2] Mazet O, Rodriguez W, Chikhi L (2015). Demographic inference using genetic data from a single individual: Separating population size variation from population structure. *Theor Popul Biol* 104: 46–58. doi: 10.1016/j.tpb.2015.06.003
- [3] Mazet O, Rodriguez W, Grusea S, Boitard S, Chikhi L (2016) On the importance of being structured: instantaneous coalescence rates and human evolution—lessons for ancestral population size inference. *Heredity* 116: 362-371. doi: 10.1038/hdy.2015.104
- [4] Junfeng Liu, Xianchao Ji & Hua Chen (2022). Beta-PSMC: uncovering more detailed population history using beta distribution. *BMC Genomics*. 23, Article number: 785. <https://doi.org/10.1186/s12864-022-09021-6>
- [5] Haller, B.C., & Messer, P.W. (2023). SLiM 4: Multispecies eco-evolutionary modeling. *The American Naturalist* 201(5), E127–E139. <https://doi.org/10.1086/723601>