Temporal FST genome scans: the case of partially selfing populations

Miguel Navascués, Arnaud Becheler, Laurène Gay, Joelle Ronfort, Renaud Vitalis

Temporal F_{ST} genome scans: the case of partially selfing populations

Miguel Navascués Arnaud Becheler Laurène Gay Joëlle Ronfort Renaud Vitalis

miguel.navascues@inra.fr

Background

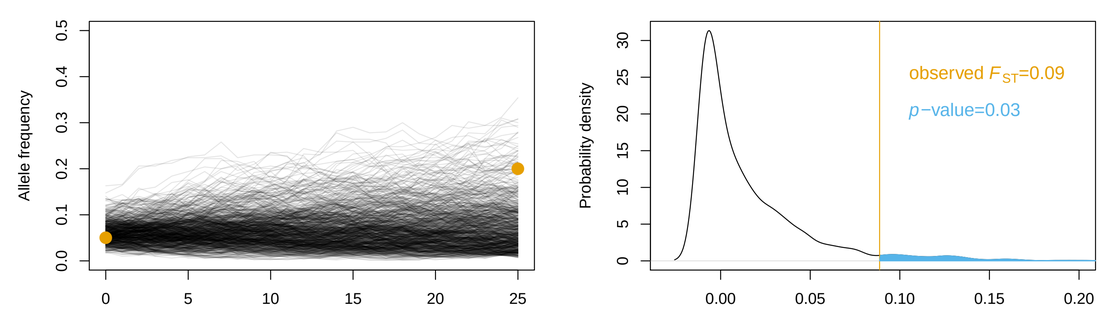
Temporal change in allele frequency, which may be characterized by a measure of genetic differentiation between two or more samples, is generally used to infer the effective size of a population (Waples 1989; *Genetics* **121**:379–91).

Because some loci may be targeted by other evolutionary forces than genetic drift, it has been proposed to test for the homogeneity of locus-specific estimates of temporal differentiation to detect outlier loci that may be targeted by natural selection (Goldringer & Bataillon 2004; doi:10.1534/genetics.103.025908).

Self-fertilization, however, reduces the effective size of populations, and therefore the effective recombination rate between loci, which promotes hitch-hiking over long chromosome stretches, and may jeopardize the detection of loci under selection.

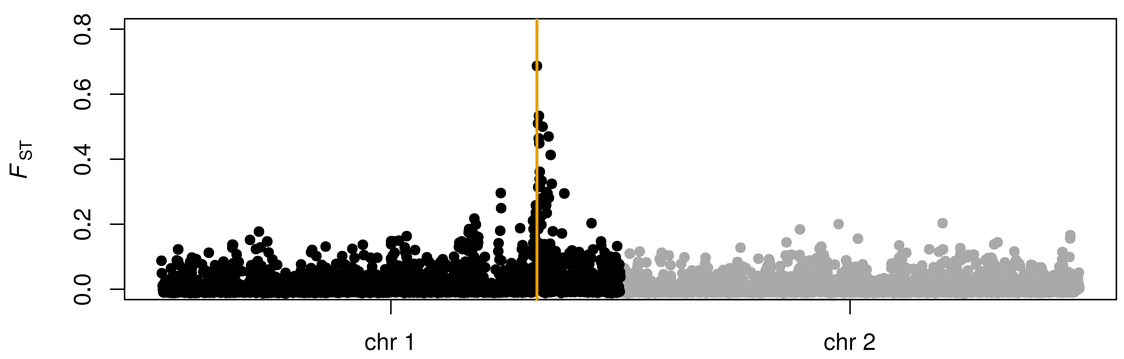
Temporal F_{ST} genome scan

We extended the method proposed by Goldringer & Bataillon (2004). In our method, genomic differentiation between two temporal samples is measured by multilocus F_{ST}, which is then used to infer effective population size ($\hat{N}_e$). We simulate alleles frequencies conditionnally on the initial frequency at a focal locus and $\hat{N}_e$ we then evaluate whether the observed F_{ST} at a focal locus is expected under the null hypothesius of neutral evolution: we compute the p-value of our test as the proportion of simulated F_{ST} values larger than the observed F_{ST}.



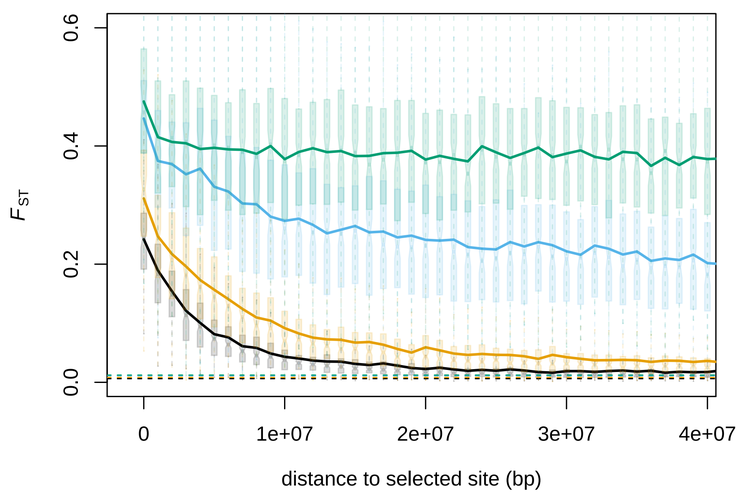
Evaluation of the method: simulations

Simulations were carried out with SLiM (Messer 2013, doi:10.1534/genetics.113.152181). Simulated populations consisted of $N = 500$ diploid individuals with a genome of 2 chromosomes (500Mb total) with $\mu = 10^{-8}$ and $r = 10^{-8}$ per bp. Two samples of 50 individuals genotyped at 10,000 SNPs were obtained 25 generations apart. A new advantageous mutation appears, or existing neutral allele becomes advantageous ($s = 0.5$), at first sampled generation. F_{ST} for each locus was calculated as in the example below (outcrossing population). Position of locus under selection.



Hitch-hiking along the chromosome

Genetic differentiation along the chromosome, as a function of the distance from the selected site, shows the extent of hitch-hiking. Selfing increases the length of the region affected by selection. How does this affect the probability to detect selection using our genome-scan approach? Does it increase the size of the genomic region for which we detect selection or does it reduce power?

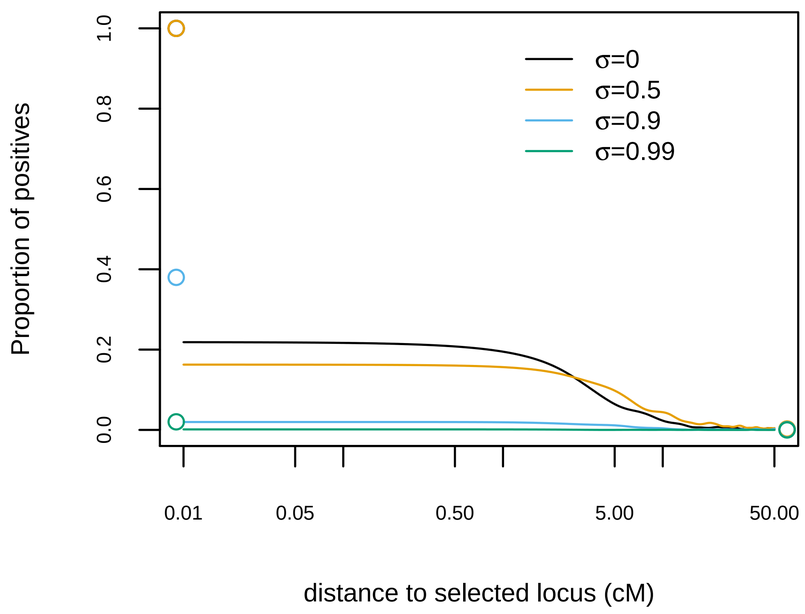


Selfing rate: $\sigma = 0$, $\sigma = 0.5$, $\sigma = 0.9$, $\sigma = 0.99$

——— mean F_{ST} - - - - expected $F_{ST} \approx \frac{\tau}{\tau + 4N_e}$ under neutrality

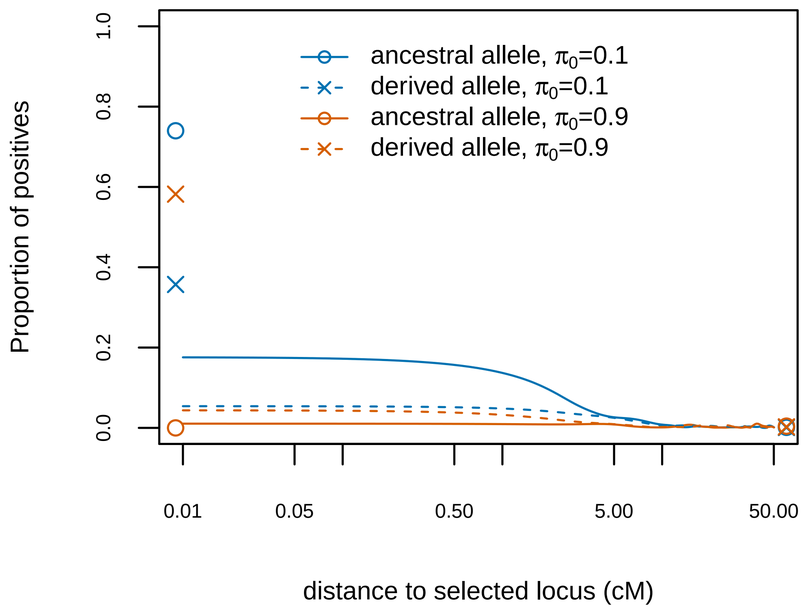
Footprints of selection acting on new mutations

Neutral markers in the vicinity of the site under selection have a higher probability to be called positive in our temporal F_{ST} test, due to hitch-hiking. This extended footprint of selection improves our chances to characterize the genomic region under selection. With larger selfing rates, however, the footprint of selection vanishes, as the whole genome is affected by selection.



Footprints of selection acting on standing variation

In highly selfing populations ($\sigma = 0.95$), recombination events occurring before the selective sweep allow to distinguish between the region around the site under selection and the rest of the genome. Advantageous alleles that start at a low frequency are almost as easy to detect as new advantageous mutations in outcrossing populations. Ancestral advantageous alleles have a higher probability of detection because they had longer time to recombine with the genetic background.



Conclusion

Temporal F_{ST} scans are powerful to detect footprints of selection on regions around loci under selection. In predominant selfing populations, low effective recombination strongly limits the detection of selection, unless it is based on standing variation.

Acknowledgements

This work was funded by INRA's meta-program ACCAF (SELFADAPT project).