

## Using patterns of genomic base composition to infer past transitions in mating systems

### Project description

Based on the distribution of selfing species on phylogenies, selfing is often considered to be of recent origin. However, except for well-documented and specific cases (such as *Capsella rubella* [1,2,3]) it is difficult to date precisely the origin of selfing. Indeed the shift from outcrossing to selfing does not necessarily coincide with the formation of a new species.

Selfing strongly affects population genetic parameters such as homozygosity, effective size and effective recombination rate, which, in turn, affect patterns of polymorphism and genome evolution [reviewed in 4,5]. Contemporary or very recent selfing rates can be estimated using genetic markers, but it is not clear if more ancient selfing rate, and especially shift in selfing rates, could also be inferred from population genetic data. An alternative approach would be to use signature of the effect of selfing that can last on a genome for longer time periods and that could be recorded in patterns of divergence among species.

A strong candidate is GC content that is affected by GC-biased gene conversion (gBGC) in many species [6,7]. gBGC is a recombination-associated process mimicking selection in favour of G and C bases [8]. During recombination mismatches induced by pairing heterozygotes during heteroduplex formation are preferentially corrected towards G or C instead of A or T. gBGC is supposed to be the main driver of the frequently-observed correlation between recombination and GC content and one of the main driver of variation of base composition within and among genomes [6,9]. Because gBGC occurs on heterozygote sites it is thus expected to be absent or nearly absent in highly selfing species [10] and decline in GC content in selfing species has been observed both at the divergence and polymorphism levels [11,12,13,14]. GC-content and gBGC also vary along genes, likely as a function of recombination gradients, first exons and introns being GC-richer than others [15,16,17,18].

The aim of the project is to use the dynamics of GC-content evolution among closely related species with different mating systems to date the shift from outcrossing to selfing by combining information from genomic regions with contrasted GC content (for example first exons versus last ones). The principle of the method is already developed but remains to be tested and applied. The project will include two parts: (i) a simulation part to assess the power and the robustness of the method and (ii) a data-analysis part where the method will be applied to empirical datasets for which genome-wide data are already available. Possible datasets are:

- *Arabidopsis thaliana*: the age of selfing in this model species is likely much more recent than the divergence with its outcrossing relative *A. lyrata* but is still debated [e.g.19].
- *Capsella orientalis*: the *Capsella* genus, on which we work in the lab, is interesting to study the evolution of mating systems. *C. rubella* is a very recent species (20,000/50,000 yrs) and the transition to selfing is concomitant to speciation [1,2,3]. *C. orientalis* is much older (~1 Myrs) but we don't know if transition occurred at the speciation time as in *C. rubella* or much later as in *A. Arabidopsis*.

- *Oryza* genus (AA clade): among the rice species close to the domesticated ones, there is variation in mating systems from outcrossing to different level of selfing and the age and number of transition are unclear. All species of the group are fully sequenced [20].
- *Caenorhabditis* nematodes: selfing evolved three times independently in this genus and, as for *C. orientalis*, the age of transition is still unknown and could differ among the three species [21,22]

A possible extension of the project is to characterized GC gradient along genes in all fully sequenced eukaryotic genomes as an indirect but easy and rapid characterization of recombination gradient, which could serve as a first basis for a global characterization of variation of recombination patterns across eukaryotes.

## Contact

Supervisor: Sylvain Glémin – [sylvain.glemin@ebc.uu.se](mailto:sylvain.glemin@ebc.uu.se)

Co-supervisor: Martin Lascoux – [martin.lascoux@ebc.uu.se](mailto:martin.lascoux@ebc.uu.se)

More info: <http://sylvain.glemin.pagesperso-orange.fr/sylvain-glemin/Home.html>

Group at EBC: <http://www.ieg.uu.se/plant/lascoux-group/>

Department: Plant Ecology and Evolution, IEG, UU

## References

1. Foxe JP, Slotte T, Stahl EA, Neuffer B, Hurka H, et al. (2009) Recent speciation associated with the evolution of selfing in *Capsella*. *Proceedings of the National Academy of Sciences of the United States of America* 106: 5241-5245.
2. Guo YL, Bechsgaard JS, Slotte T, Neuffer B, Lascoux M, et al. (2009) Recent speciation of *Capsella rubella* from *Capsella grandiflora*, associated with loss of self-incompatibility and an extreme bottleneck. *Proceedings of the National Academy of Sciences of the United States of America* 106: 5246-5251.
3. Brandvain Y, Slotte T, Hazzouri KM, Wright SI, Coop G (2013) Genomic identification of founding haplotypes reveal the history of the selfing species *Capsella rubella*. *PLoS Genetics* 9.
4. Burgarella C, Glémin S (2017) Population Genetics and Genome Evolution of Selfing Species. *Encyclopedia of Life Sciences: John Wiley and Sons*.
5. Glémin S, Galtier N (2012) Genome evolution in outcrossing versus selfing versus asexual species. *Methods in Molecular Biology* 855: 311-335.
6. Duret L, Galtier N (2009) Biased gene conversion and the evolution of mammalian genomic landscapes. *Annual Review of Genomics and Human Genetics* 10: 285-311.
7. Pessia E, Popa A, Mousset S, Rezvoy C, Duret L, et al. (2012) Evidence for widespread GC-biased gene conversion in eukaryotes. *Genome Biology and Evolution* 4: 675-682.
8. Marais G (2003) Biased gene conversion: implications for genome and sex evolution. *Trends in Genetics* 19: 330-338.
9. Mugal CF, Weber CC, Ellegren H (2015) GC-biased gene conversion links the recombination landscape and demography to genomic base composition: GC-biased gene conversion drives genomic base composition across a wide range of species. *Bioessays* 37: 1317-1326.
10. Marais G, Charlesworth B, Wright SI (2004) Recombination and base composition: the case of the highly self-fertilizing plant *Arabidopsis thaliana*. *Genome Biology* 5: R45.
11. Muyle A, Serres-Giardi L, Ressayre A, Escobar J, Glémin S (2011) GC-biased gene conversion and selection affect GC content in the *Oryza* genus (rice). *Molecular Biology and Evolution* 28: 2695-2706.

12. Haudry A, Cenci A, Guilhaumon C, Paux E, Poirier S, et al. (2008) Mating system and recombination affect molecular evolution in four Triticeae species. *Genetical Research* 90: 97-109.
13. Burgarella C, Gayral P, Ballenghien M, Bernard A, David P, et al. (2015) Molecular Evolution of Freshwater Snails with Contrasting Mating Systems. *Molecular Biology & Evolution* 32: 2403-2416.
14. Clement Y, Sarah G, Holtz Y, Homa F, Pointet S, et al. (2017) Evolutionary forces affecting synonymous variations in plant genomes. *PLoS Genet* 13: e1006799.
15. Ressayre A, Glémin S, Montalent P, Serre-Giardi L, Dillmann C, et al. (2015) Introns Structure Patterns of Variation in Nucleotide Composition in *Arabidopsis thaliana* and Rice Protein-Coding Genes. *Genome Biol Evol* 7: 2913-2928.
16. Glémin S, Clement Y, David J, Ressayre A (2014) GC content evolution in coding regions of angiosperm genomes: a unifying hypothesis. *Trends in Genetics* 30: 263-270.
17. Stoletzki N (2011) The surprising negative correlation of gene length and optimal codon use--disentangling translational selection from GC-biased gene conversion in yeast. *BMC Evolutionary Biology* 11: 93.
18. Clement Y, Fustier MA, Nabholz B, Glémin S (2015) The bimodal distribution of genic GC content is ancestral to monocot species. *Genome Biology and Evolution* 7: 336-348.
19. Castric V, Billiard S, Vekemans X (2014) Trait transitions in explicit ecological and genomic contexts: plant mating systems as case studies. *Adv Exp Med Biol* 781: 7-36.
20. Jacquemin J, Bhatia D, Singh K, Wing RA (2013) The International Oryza Map Alignment Project: development of a genus-wide comparative genomics platform to help solve the 9 billion-people question. *Curr Opin Plant Biol* 16: 147-156.
21. Thomas CG, Woodruff GC, Haag ES (2012) Causes and consequences of the evolution of reproductive mode in *Caenorhabditis* nematodes. *Trends in Genetics* 28: 213-220.
22. Fierst JL, Willis JH, Thomas CG, Wang W, Reynolds RM, et al. (2015) Reproductive Mode and the Evolution of Genome Size and Structure in *Caenorhabditis* Nematodes. *PLoS Genet* 11: e1005323.