

Un scan de sélection pour identifier des gènes sous sélection équilibrante chez le tournesol
(english version below)

Université Claude Bernard Lyon 1

Laboratoire de Biométrie et Biologie Evolutive UMR CNRS 5558 Villeurbanne

Encadrante : Hélène Badouin (helene.badouin@univ-lyon1.fr), <https://lbbe.univ-lyon1.fr/-Equipe-Sexe-et-Evolution-.html?lang=fr>,

<https://scholar.google.fr/citations?user=QkgCKZsAAAAJ&hl=fr>

Contexte: Comprendre comment et pourquoi les espèces évitent l'autofécondation est une question importante en biologie. Chez de nombreuses plantes à fleurs, l'autofécondation est empêchée par des mécanismes d'auto-incompatibilité génétique (AI), où l'interaction entre les protéines spécifiques du pistil et du pollen entraîne le rejet de l'autopollen¹. L'AI est généralement contrôlé par un locus unique, multi-allélique (locus S) qui évolue dans le cadre d'une forte sélection équilibrante, c'est-à-dire une sélection pour le maintien de la diversité génétique². L'acquisition et la perte multiples d'AI ont eu lieu au cours de l'évolution des Angiospermes, mais les gènes sous-jacents (gènes S) sont inconnus dans la plupart des familles, limitant grandement notre compréhension des forces évolutives et des mécanismes proximaux qui régissent ces transitions³. Le tournesol commun *Helianthus annuus* est actuellement l'un des très rares modèles de plantes AI dans lequel les ressources et les outils nécessaires pour une étude intégrative de la découverte des gènes S sont disponibles. Un locus de caractères quantitatifs pour l'AI a été identifié⁴, et un génome de référence de haute qualité a été récemment obtenu (Badouin et al., 2017, Nature⁵). Le projet GENESIS (Exploration génomique des gènes d'auto-incompatibilité chez le tournesol) utilise une combinaison d'approches génomiques et génétiques pour identifier et valider les gènes causant l'auto-incompatibilité chez le tournesol.

Objectifs: Le but de cette formation est d'effectuer un scan du génome⁶ pour détecter la sélection équilibrante dans un ensemble de données de génomique des populations de tournesols sauvages. Nous avons reséquencé les génomes complets de 36 individus sauvages et d'un groupe externe. L'alignement des lectures contre un génome de référence de haute qualité (mapping) et la détection des variants (SNP calling) ont été réalisés. L'étudiant ou étudiante utilisera l'ensemble de ces données pour : 1) Identifier les gènes montrant une couverture suffisante pour être analysé ; 2) Calculer des statistiques descriptives de génomique de population; 3) Effectuer des tests de sélection équilibrante (tests Tajima D, Fay et Wu H). Cela sera fait avec les bibliothèques scientifiques existantes pour la génomique des populations (bio ++, libsequence ou egglib⁷).

Compétences: l'étudiant ou étudiante acquerra ou améliorera des compétences en bioinformatique (utilisation d'un environnement Unix, d'un cluster informatique partagé, langage script Python ou Perl, R), analyse de données NGS et génomique des populations.

Références:

1. Nasrallah, J. B. Recognition and Rejection of Self in Plant Reproduction. *Science* **296**, 305–308 (2002).
2. Charlesworth, D., Vekemans, X., Castric, V. & Glémin, S. Plant self-incompatibility systems: a molecular evolutionary perspective. *New Phytol.* **168**, 61–69 (2005).
3. Allen, A. M. & Hiscock, S. J. Evolution and Phylogeny of Self-Incompatibility Systems in Angiosperms. in *Self-Incompatibility in Flowering Plants* 73–101 (Springer Berlin Heidelberg, 2008). doi:10.1007/978-3-540-68486-2_4
4. Gandhi, S. D. *et al.* The self-incompatibility locus (S) and quantitative trait loci for self-pollination and seed dormancy in sunflower. *Theor. Appl. Genet.* **111**, 619–629 (2005).
5. **Badouin**, H. *et al.* The sunflower genome provides insights into oil metabolism, flowering and Asterid evolution. *Nature* **546**, 148–152 (2017).
6. **Badouin**, H. *et al.* Widespread selective sweeps throughout the genome of model plant pathogenic fungi and identification of effector candidates. *Mol. Ecol.* **26**, 2041–2062 (2017).
7. Mita, S. D. & Siol, M. EggLib: processing, analysis and simulation tools for population genetics and genomics. *BMC Genet.* **13**, 27 (2012).

A genome scan for balancing selection for identifying self-incompatibility genes in sunflower

Université Claude Bernard Lyon 1

Laboratoire de Biométrie et Biologie Evolutive UMR CNRS 5558 Villeurbanne

Encadrante : Hélène Badouin (helene.badouin@univ-lyon1.fr), <https://lbbe.univ-lyon1.fr/-Equipe-Sexe-et-Evolution-.html?lang=fr>,

<https://scholar.google.fr/citations?user=QkgCKZsAAAAJ&hl=fr>

Context: Understanding how and why species avoid self-fertilization is an important question in biology. In many flowering plants, self-fertilization is prevented by **genetic self-incompatibility (SI)**, where interaction between specific pistil and pollen proteins leads to the rejection of self-pollen¹. SI is generally controlled by a single, multi-allelic locus (S locus) that evolves under strong balancing selection, i.e. selection for maintenance of genetic diversity². **Multiple acquisition and loss of SI have occurred during Angiosperm evolution, but the underlying genes (S genes) are unknown in most families, greatly limiting our understanding of the evolutionary forces and proximal mechanisms that govern these transitions³.** The common sunflower *Helianthus annuus* is currently **one of the very few SI plant models in which the necessary resources and tools for an integrative study for S genes discovery are available.** A quantitative trait locus for SI has been identified⁴, and a high-quality reference genome was recently obtained (Badouin *et al.* 2017, Nature⁵). GENESIS project (Genomic Exploration of Self-Incompatibility Genes In Sunflower) uses a **combination of genomics and genetics approaches to identify and validate the genes causing self-incompatibility in sunflower.**

Objectives: The goal of this training is to perform a **genome scan⁶ for balancing selection** in a dataset of **population genomics** in wild sunflowers. We have resequenced the complete

genomes of 36 wild individuals and an outgroup using paired-end Illumina. Mapping of short-reads on a high-quality reference genome and SNP calling have been completed. The student will use these data to 1) Identify genes that show sufficient coverage to be analyzed, 2) compute descriptive statistics of population genomics, 3) perform tests of balancing selection (Tajima's D, Fay and Wu's H, HKA tests). This will be done with existing scientific libraries for population genomics (bio++, libsequence or egglib⁷). Depending on interest and background, the student could also take part in the analysis of a long-reads transcriptomics pooled dataset for which a *ad hoc* bioinformatics pipeline will have to be developed.

Competences: the student will acquire or improve skills in **bioinformatics** (use of a unix environment, of a shared computing cluster, scripting language python or perl, R, C++ if bases in this language), analysis of **NGS data** and in **population genomics** (theoretical and practical knowledge).