

**Sujet: Améliorations des performances et de l'utilisabilité d'outils d'analyse de variants génomiques.**

- Laboratoire d'accueil: Laboratoire Reproduction et Développement des plantes ([RDP](#)), ENS Lyon, Lyon
- Équipe d'accueil: Signalisation hormonale et développement ([page web](#)).
- Encadrant: Fabrice Besnard (chercheur, CR INRAe)

## Présentation générale (contexte et objectif)

Comprendre les bases génétiques des processus et propriétés biologiques observables (les phénotypes) reste une question centrale de la biologie. Pour aborder ce problème de manière expérimentale, il est indispensable de disposer d'organismes qui porte des variations génétiques (différents génotypes), afin de corréliser variations des génotypes et variations des phénotypes. Les techniques modernes de séquençage des génomes révolutionnent les pratiques de la génétique, car il est désormais possible de connaître rapidement, de manière précise et presque exhaustive la séquence du génome d'individus ou d'échantillons. Que ce soit au sein de populations naturelles naturellement porteuses de variations génétiques, au travers de banques de mutants disponibles pour quelques organismes modèles ou suite à de nouvelles expériences de mutagenèse, l'accès à des ressources génétiques bien caractérisée n'est en théorie plus un problème. En pratique, les données brutes fournies par les procédés actuels de séquençage massif en parallèle (aussi appelées "Next Generation Sequencing") sont inutilisables telles quelles et nécessitent des analyses bio-informatiques poussées pour identifier de manière fiable des différences de séquences entre plusieurs échantillons. Pour profiter au mieux de la puissance du séquençage moderne, il est donc nécessaire de se doter d'outils d'analyses performants.

Au laboratoire, nous étudions la génétique du développement de plantes modèles (notamment *Arabidopsis thaliana* et *Physcomitrium patens*) pour lesquelles des ressources génétiques naturelles et artificielles (mutagenèse) existent. Toutefois peu de personnes maîtrisent les outils d'analyses pour que leur utilisation se généralise dans le laboratoire et que ce type d'approche soit utilisée de manière plus routinière par les généticiens expérimentaux.

L'enjeu de ce stage est donc d'améliorer les outils existant d'analyse de variants génomiques pour les rendre plus fiables, plus performants et plus accessibles à des scientifiques sans compétences de bio-informatiques.

## Description des axes de travail

Nous avons découpé le travail en trois axes complémentaires:

1. Améliorer les outils d'analyses de variants, en particulier ceux liés à l'analyse des variants structuraux ;
2. Mise en place d'un système de gestion de chaîne de traitement ("pipeline") ;
3. Analyse de données concernant des expériences en cours au laboratoire ;

Ces trois axes pourront être approfondis plus ou moins séparément en fonction des préférences et des aptitudes du stagiaire ainsi que des opportunités particulières rencontrées lors de leur développement.

### Axe 1: Amélioration des outils d'analyses

1.1. Méthode actuelle: amélioration des scripts d'analyses comparée des sorties des programmes [Pindel](#) et [Breakdancer](#) (montée version python, paramétrisation et automatisation pour l'utilisateur, architecture et dépendance des scripts).

1.2. Étude bibliographique des logiciels de détection de variants structuraux utilisés dans le domaine, implémentation et tests (ex: [MetaSV](#), approches mixtes comme dans [Konrad A. et al., 2019](#)).

1.3. Mise à niveau (version) des outils d'analyse des petits variants (principalement montée de version GATK3.6 → [GATK latest](#) - 4.1+)

1.3. Amélioration du format final des sorties (export facile vers des tableurs)

*Axe 2: Système de gestion de chaîne de traitement*

2.1. Automatisation de la soumission d'analyse en lots parallèles

2.2. Analyse des solutions existantes (par exemple: [Luigi](#) ou [autres systèmes](#))

2.3. Mise en place et tests d'une ou deux solutions

*Axe3: Analyse de données réelles*

3.1. Analyse d'échantillons issues d'une mutagenèse en cours sur la mousse modèle *Physcomitrium patens*.

3.2. Caractérisation de mutants par insertion T-DNA chez *Arabidopsis thaliana*

## Références

*Publications de membre de l'équipe utilisant les techniques de génomiques:*

**Besnard, F.**, Koutsovoulos, G., Dieudonné, S., Blaxter, M., and Félix, M.-A. (2017). Toward Universal Forward Genetics: Using a Draft Genome Sequence of the Nematode *Oscheius tipulae* To Identify Mutations Affecting Vulva Development. *Genetics* 206, 1747–1761.

Vargas-Velazquez, A.M., **Besnard, F.**, and Félix, M.-A. (2019). Necessity and Contingency in Developmental Genetic Screens: EGF, Wnt and Semaphorin Pathways in Vulval Induction of the Nematode *Oscheius tipulae*. *Genetics* genetics.301970.2019.

**Besnard, F.**, Picao-Osorio, J., Dubois, C., and Félix, M.-A. (2020). A broad mutational target explains a fast rate of phenotypic evolution. *ELife* 9, e54928.

*Code public existant:* [Andalusian Mapping](#)