

**Intitulé :** Etablissement de pangénome graphe d'agents phytopathogènes à but de suivi de population.

Type de poste: Stage M2/M1

Type de contrat: Stage

Durée : 6 mois

Labo : BIOGER

Ville : Palaiseau (Agrocampus)

Contact : Nicolas Lapalu

Email : [nicolas.lapalu@inrae.fr](mailto:nicolas.lapalu@inrae.fr)

Mots-clés : pangénomique, bioinformatique, phytopathogène, GWAS, développement

### **Contexte :**

La possibilité d'obtenir des génomes complets télomère-à-télomères (T2T) grâce aux technologies de séquençage de 3ème génération (long reads) a ouvert la voie à l'analyse de la diversité génétique des individus en termes de contenu de séquence ou de réarrangement à l'échelle du génome entier. Ainsi, là où de nombreuses approches de génétique des populations et de génétique quantitative utilisaient majoritairement les différences observées entre un génome de référence et les individus/populations étudiés, il est maintenant possible de compléter ces analyses en prenant en compte les variants structuraux (SV) et les régions spécifiques absentes du génome de référence. Ces dernières années ont vu apparaître le développement d'outils permettant d'intégrer l'ensemble de l'information de séquence nucléotidique dans une structure de graphe à partir d'alignements de génomes [1]. La variabilité génétique est alors représentée par un branchement dans le graphe lorsque le contenu en séquence diverge entre individus. La notion de graphe de pangénome (PGG) est ainsi apparue ouvrant la voie à l'utilisation d'une structure plus informative que la simple séquence d'un génome de référence. Le développement d'outils et les formats de fichiers permettant de manipuler ces PGG sont actuellement un front de science dont on peut constater l'importance par les publications récentes autour du pangénome humain [2]–[4].

L'unité d'accueil, BIOGER, est une unité de recherche dont les activités sont centrées sur l'étude et le suivi d'agents phytopathogènes de grandes cultures tels que *Zymoseptoria tritici*, agent de la septoriose du blé, *Leptosphaeria maculans* responsable de la verse du colza, ou encore *Botrytis cinerea* responsable de la pourriture grise. L'étudiant sera accueilli au sein du plateau bioinformatique, (bioinfobioger) qui gère les données génomiques de l'unité, mais réalise aussi de nombreuses analyses (assemblage, annotations, transcriptomiques, ...). Le plateau assure entre autres l'analyse du polymorphisme de séquence des populations collectées par les équipes afin de réaliser des suivis d'allèles d'intérêt (surveillance des populations) et de conduire des analyses génétiques d'association génotype-phénotype ou génotype-environnement (GWAS et GEA). Le protocole d'analyse actuel se base sur l'utilisation d'un ou plusieurs génomes de référence identifiés par la communauté, mais qui ne représente pas la complétude de la diversité génétique des espèces. Le but du stage est d'initier la transition de l'utilisation de génome de référence vers l'utilisation de PGG, en éprouvant les méthodologies et en comparant les résultats aux études déjà réalisées. L'utilisation de PGG permettra de centraliser l'ensemble des informations de diversité génétique connues pour les agents phytopathogènes faisant l'objet d'un suivi de population et donc d'accéder à l'exhaustivité de la diversité génétique de l'espèce. Le stage se fera en interaction avec une autre équipe de l'unité (2

chercheurs, 1 post-doc et stagiaires) travaillant sur *Z.tritici*, ainsi qu'une équipe partenaire à Montpellier, travaillant sur le modèle *Pyricularia oryzae*, et qui seront les premiers utilisateurs des résultats produits pendant le stage.

### Objectifs du stage :

- Etablissement de PGG sur les modèle *Z. tritici* et *P. oryzae*, en utilisant des génomes complets publiés [5][6] ou disponibles en interne.
- Validation de l'évolutivité du graphe en incorporant progressivement de la variabilité par ajout de nouveaux génomes ou régions spécifiques.
- Utilisation de population d'isolats prélevés au champs et séquencés en short-read dans le but de réaliser l'identification des variants à partir des graphes de pangénomes. Validation des résultats en comparaison avec des résultats de GWAS disponibles dans l'unité [7].
- Mise en place d'un protocole de suivi d'allèles ou de régions génomiques à partir du PGG [8], [9].
- L'ensemble des analyses seront tracées dans un gestionnaire de projet et rejouables sous forme de notebook jupyter ou autre. En fonction de l'avancée du projet, l'étudiant pourra initier/améliorer des workflow d'analyses à partir des méthodologies mises en place pendant le stage.

L'ensemble des données du projet sont déjà disponibles, ce qui permettra de commencer immédiatement. Ce stage s'inscrit dans un projet avec plusieurs partenaires (biologistes/bioinformaticien). Une capacité forte à expliciter et présenter son travail est attendu.

**Compétences recherchées :** Connaissances générales bio-informatique, génomique et programmation (Python, shell). Utilisation de l'environnement Linux/Unix, installation d'outil. Envie de communiquer et d'échanger. Connaissances pratiques, manager de workflow (snakemake) et d'outil de développement (git, gitlab, environnement de tests) seraient un plus, mais seront acquises pendant les stages si manquants.

### Références :

- [1] J. M. Eizenga *et al.*, "Pangenome Graphs," *Annu. Rev. Genomics Hum. Genet.*, vol. 21, pp. 139–162, Aug. 2020, doi: 10.1146/annurev-genom-120219-080406.
- [2] Y. Gao *et al.*, "A pangenome reference of 36 Chinese populations.," *Nature*, vol. 619, no. 7968, pp. 112–121, Jul. 2023, doi: 10.1038/s41586-023-06173-7.
- [3] K. H. Miga and T. Wang, "The Need for a Human Pangenome Reference Sequence," *Annu. Rev. Genomics Hum. Genet.*, vol. 22, pp. 81–102, 2021, doi: 10.1146/ANNUREV-GENOM-120120-081921.
- [4] T. Wang *et al.*, "The Human Pangenome Project: a global resource to map genomic diversity," *Nature*, vol. 604, no. 7906, pp. 437–446, Apr. 2022, doi: 10.1038/s41586-022-04601-8.
- [5] T. Badet, U. Oggenfuss, L. Abraham, B. A. McDonald, and D. Croll, "A 19-isolate reference-quality global pangenome for the fungal wheat pathogen *Zymoseptoria tritici*," *BMC Biol.*, vol. 18, no. 1, Feb. 2020, doi: 10.1186/s12915-020-0744-3.
- [6] Z. Peng *et al.*, "Effector gene reshuffling involves dispensable mini-chromosomes in the wheat

blast fungus," *PLoS Genet.*, vol. 15, no. 9, p. e1008272, 2019, doi: 10.1371/JOURNAL.PGEN.1008272.

- [7] R. Amezrou *et al.*, "Whole-genome sequencing reveals diverse mechanisms underlying quantitative pathogenicity and host adaptation in a fungal plant pathogen," *bioRxiv*, p. 2022.12.23.521735, Jan. 2022, doi: 10.1101/2022.12.23.521735.
- [8] S. Omrane *et al.*, "Plasticity of the MFS1 Promoter Leads to Multidrug Resistance in the Wheat Pathogen *Zymoseptoria tritici*," *mSphere*, vol. 2, no. 5, Oct. 2017, doi: 10.1128/MSPHERE.00393-17.
- [9] H. Bellah, G. Gazeau, S. Géglise, R. Amezrou, T. C. Marcel, and D. Croll, "A highly multiplexed assay to monitor pathogenicity, fungicide resistance and gene flow in the fungal wheat pathogen *Zymoseptoria tritici*," *PLoS One*, vol. 18, no. 2, Feb. 2023, doi: 10.1371/JOURNAL.PONE.0281181.

Lieu : INRAE UR BIOGER, Univ. Paris-Saclay, AgroCampus 22 Place de l'Agronomie, 91120 Palaiseau – France

Contact : Nicolas Lapalu (encadrant), [nicolas.lapalu@inrae.fr](mailto:nicolas.lapalu@inrae.fr)