

Stage M2 : Reconstruction d'haplotype à partir de données de séquençage en pool ; Application à l'étude de la diversité de population ancienne de maïs

L'évaluation de la diversité des ressources génétiques est une étape clef pour aborder les nouveaux objectifs de sélection face au changement climatique. Les études au niveau des populations nécessitent la caractérisation de nombreux individus, le séquençage génomique (WGS) ou le génotypage d'individus à cette échelle n'est pas envisageable. L'approche en mélange (Pool-Seq) constitue une alternative intéressante à l'étude séparée des individus. Grâce au développement de nombreux logiciels, le Pool-seq est de plus en plus utilisé pour l'étude de la diversité des populations d'organismes modèles et non modèles, aux caractéristiques de génomes variées (e.g. taille de génome et ploïdie). Le Pool-seq repose sur la construction d'une banque unique à partir de plusieurs ADNs individuels et le séquençage NGS de cette banque. Les erreurs de séquençage sont ensuite distinguées des « vrais » SNPs présents à faible fréquence en appliquant un seuil minimum de fréquence allélique.

Si de nombreuses études ont montré l'intérêt du Pool-Seq pour identifier des SNP utiles à l'étude de la diversité génomique des populations. La structure sous-jacente des haplotypes reste cependant encore mal caractérisée alors que les modèles génétiques en considérant les haplotypes devraient être plus résolutifs.

En théorie, les haplotypes peuvent être reconstruit à partir de données pool-seq grâce aux fréquences corrélées des SNP que portent les lectures. Dans ce stage, nous proposons de tester/adapter les différentes approches bioinformatique de reconstruction d'haplotypes à partir de données de séquençage courtes lectures en pool de génome réduits. Les données sont déjà disponibles chez le maïs (le modèle utilisé dans ce stage) et ont été obtenues par des approches de génotypage ciblé de SNPs par séquençage. Le stage s'inscrit pleinement dans le programme de recherche menés chez le maïs, qui vise à décrire la diversité haplotypique au niveau de QTL, des populations anciennes pour une meilleure connaissance et gestion de celles-ci dans les programmes de sélection génomique.

L'étudiant(e) sera basé à l'EPGV, CEA-IbFJ, Evry 2 rue Gaston Crémieux, 91 570 Evry. L'étudiant(e) en bioinformatique ou double compétence bio/bioinfo bénéficiera d'un double encadrement par les chercheurs de l'EPGV et du Moulon équipe GQE

Compétences attendues :

Formation en Bioinformatique ou double compétence Bio/BioInformatique, niveau Master 2.
Langage de programmation (de préférence R) et travail en lignes de commande sous Linux (l'utilisation d'un cluster de calcul serait un plus)

Motivation, autonomie, rigueur et prise d'initiatives.

Compétences pour l'analyse de données NGS et statistiques.

Intérêts pour la génétique / génomique des populations.

Références dans le domaine :

- Guirao-Rico S, González J. Benchmarking the performance of Pool-seq SNP callers using simulated and real sequencing data. *Mol Ecol Resour.* 2021 May;21(4):1216-1229. doi: 10.1111/1755-0998.13343. Epub 2021 Mar 5. PMID: 33534960; PMCID: PMC8251607.
- Mariangela Arca et al., 'Genome-Wide SNP Genotyping of DNA Pools Identifies Untapped Landraces and Genomic Regions That Could Enrich the Maize Breeding Pool', *BioRxiv*, 2 October 2020, 2020.09.30.321018, <https://doi.org/10.1101/2020.09.30.321018>.
- Mariangela Arca et al., 'Deciphering the Genetic Diversity of Landraces With High-Throughput SNP Genotyping of DNA Bulks: Methodology and Application to the Maize 50k Array', *Frontiers in Plant Science* 11 (2021), <https://doi.org/10.3389/fpls.2020.568699>.
- Davide Scaglione et al., 'Single Primer Enrichment Technology as a Tool for Massive Genotyping: A Benchmark on Black Poplar and Maize', *Annals of Botany* 124, no. 4 (29 October 2019): 543–51, <https://doi.org/10.1093/aob/mcz054>.

- Christian Schlötterer et al., 'Sequencing Pools of Individuals — Mining Genome-Wide Polymorphism Data without Big Funding', *Nature Reviews Genetics* 15, no. 11 (November 2014): 749–63, <https://doi.org/10.1038/nrg3803>.

Contexte de travail :

L'unité de service Etude du Polymorphisme des Génomes Végétaux (EPGV), US EPGV, d'INRAE est localisée depuis 2001 à Evry, en accueil au Genoscope (CEA-IbFJ-Evry) Commissariat à l'Energie Atomique et aux énergies alternatives - CEA), plateforme d'envergure et de réputation internationale dans le domaine de la génomique. L'EPGV participe à de nombreux projets d'analyse de génomes des végétaux travaillés à INRAE et Instituts techniques principalement. L'EPGV possède une double expertise biologie moléculaire et bioanalyse. L'unité EPGV transmet son expertise, assure la production de données avec fiabilité et traçabilité.

L'équipe Génétique Quantitative et Méthodologie de la Sélection (GQE) est une équipe majeure de la recherche sur le maïs en France. En partenariat avec les acteurs privés et publics de la filière maïs elle axe ses recherches sur des approches de génétique quantitative liées à la méthodologie de la sélection et à la gestion de la variabilité génétique. Elle vise à comprendre l'organisation de la diversité du maïs, étudier le déterminisme des caractères quantitatifs et optimiser les programmes de sélection.

Rémunération : Le/la stagiaire recevra une gratification d'environ 570-600€/mois.

Candidatures : envoyer CV, lettre de motivation et relevés de notes par courriel à patricia.favre-rampant@inrae.fr, damien.hinsinger@inrae.fr