

Intitulé du poste

M2 – Mise au point d'un pipeline bioinformatique de détection de méthylation de l'ADN chez deux espèces d'arbres : le peuplier et le chêne

Type de poste

Stage M2

Dates

Durée du poste

6 mois

Date de prise de fonction

Début 2021

Date de fin de validité de l'annonce

01/12/2020

Localisation

Nom de la structure d'accueil

CEA - Institut de Biologie Francois Jacob - Centre National de la Recherche en Génomique Humaine - Laboratory for Epigenetics and Environment (LEE)

INRAe – Unités **BioForA** (Biologie intégrée pour la valorisation de la diversité des arbres et de la forêt) et **BioGeCo** (Biodiversité, Gènes et Communautés)

Adresse

Centre National de la Recherche en Génomique Humaine

2 rue Gaston Crémieux

CP5721

91057 Evry Cedex

Unité INRAe BioForA

2163, avenue de la pomme de pin

CS 40001 Ardon

45075 Orléans Cedex 2

Contacts

Odile Rogier (BioForA)

Isabelle Lesur Kupin (BioGeCo)

Email du/des contacts

odile.rogier@inrae.fr

Isabelle.LesurKupin@inrae.fr

Merci d'envoyer un CV et une lettre de motivation à ces 2 adresses le plus rapidement possible (deadline : 20/11/2020).

Description

<https://www6.inrae.fr/epitree-project/Le-projet-EPITREE>



Description du poste

Contexte scientifique et enjeux

Pour comprendre les impacts des changements climatiques sur les écosystèmes et en particulier les systèmes forestiers, nous souhaitons étudier les impacts évolutifs et fonctionnels des variations épigénétiques chez le chêne (*Quercus robur*) et chez le peuplier (*Populus nigra*). C'est un objectif majeur du projet ANR [EpiTree \(ANR-17-CE32-0009\)](#) dans lequel s'inscrit ce stage et qui vise à étudier l'impact de la méthylation de l'ADN, de l'expression des gènes et de la variation allélique dans les mécanismes d'adaptation des arbres à l'environnement local.

Objectif et plan de recherche

L'objectif du stage est d'étudier les variations du niveau de la méthylation de l'ADN. Un séquençage en WGBS (Whole Genome Bisulfite Sequencing) a été réalisé sur 30 individus. Il a permis l'identification de 24Mb de régions d'intérêt (zones hypo- hyperméthylées, gènes candidats) pour chacune des espèces (chêne et peuplier). Ces régions seront, par la suite, séquencées chez 300 peupliers et 250 chênes selon un protocole de capture methyl-seq. Elles devront être analysées dans les 3 contextes de méthylation (CG, CHG et CHH) comme elles l'ont été sur le génome entier. L'enjeu du stage proposé sera de développer et automatiser ce workflow bioinformatique afin de pouvoir le reproduire sur un grand nombre d'individus.

Un test pilote (sur un individu avec différentes conditions techniques) pour valider notre approche est en cours. Il conviendra, dans un premier temps, de vérifier les résultats obtenus pour valider la stratégie en comparant les résultats obtenus par WGBS (génomique entière) et par capture (24Mb d'intérêt). Ensuite, l'objectif principal du travail proposé sera la mise au point d'un pipeline de détection de méthylation de l'ADN utilisant les outils déjà disponibles qui auront été validés par les partenaires du projet spécialistes dans ce domaine. Une fois les outils choisis, le recours à un gestionnaire de workflow sera nécessaire pour la mise en place du pipeline d'analyse. Enfin, il conviendra de valider et de tester ce pipeline avec les données acquises au cours du projet.

Techniques

Le travail consistera principalement à développer un pipeline sur un cluster de calcul bioinformatique sous SLURM à l'aide d'outils bioinformatiques existants pour la détection de méthylation (Bismark, BSMAP, MethylKit, ...). En particulier, l'étudiant(e) sera en contact avec les chercheurs du projet qui connaissent bien ces outils ainsi qu'un ingénieur informaticien du CEA. L'étudiant(e) devra connaître l'environnement Linux et être familier avec un gestionnaire de workflows (de préférence Snakemake) ainsi que les différents formats classiques de fichiers bioinformatiques associés (FastQ, BAM/SAM, BED, VCF, ...). Ce travail nécessitera l'utilisation de langages de programmation pour la manipulation (Shell, Python (2 et 3), Perl) et l'analyse statistique (R) des données. Enfin, dans un souci de reproductibilité, l'étudiant(e) devra participer à l'alimentation du GitLab du projet.

Enfin, l'encadrement de ce stage étant réalisé sur 2 sites distincts (BioForA et BioGeCo) et le lieu du stage se situant au CNRGH (CEA) à Évry, de bonnes capacités de communication sont essentielles tout comme une capacité à travailler en collaboration.

Résultats attendus

Ce travail devrait permettre d'identifier les liens entre l'épigénome de deux espèces d'arbres (peuplier et chêne) et les caractéristiques de ces espèces telles que la tolérance à la sécheresse et la phénologie de croissance. L'automatisation de l'analyse d'une marque épigénétique (méthylation de l'ADN) sur un nombre limité d'individus devrait aboutir à un pipeline d'analyse robuste et reproductible qui pourra, par la suite, être déployé sur l'ensemble des 300 peupliers et 250 chênes prévus dans le projet. Il est également envisagé une possible valorisation par une publication.