

SUJET DE STAGE

NOM, prénom de la personne proposant le stage :

Francillonne Nicolas

Johann Confais,

Adresse Professionnelle :

INRAE, Centre de recherche de Versailles, bat.18 RD10, Route de Saint Cyr
78026 Versailles
France

Adresse électronique : johann.confais@inrae.fr et nicolas.francillonne@inrae.fr

Entreprise/Unité d'appartenance : Institut National de recherche pour l'agriculture, l'alimentation et l'environnement INRAE

Unité de Recherche en Génomique-Info (URGI) - UR 1164 INRAE Versailles

Domaine d'expertise de l'entreprise / laboratoire : Développement d'outils et acquisition de connaissances sur la structure, l'évolution et le fonctionnement du génome.

Titre du stage : Exploitation de données hétérogènes dans une base de données orientée graphe : cas d'usage pour l'adaptation des plantes au changement climatique.

Mots clés :

Génomique, pangénomique, éléments transposables, système d'information, base graphe, traitement de données, pipeline

Description du sujet (1 page maximum)

Contexte :

Une base de données orientée graphe a été construite au sein de l'URGI sur 2 espèces modèles de plante : une monocotylédone (*Brachypodium distachyon*) et une dicotylédone (*Arabidopsis thaliana*). Elle intègre de nombreuses données hétérogènes en génomique: annotations de génome (annotations structurales et fonctionnelles de gènes, éléments transposables (ET), séquences non codantes conservées (CNS), site de fixation de facteur de transcription (TFBS), relations d'homologie et d'orthologie.

Les TFBS (Transcription Factor Binding Site) sont les éléments clés de la régulation des gènes. Tout comme les gènes, les Certains ET TE portent aussi des TFBS pour leur régulation. Certains qui portent des TFBS sont connus pour être activés en réponse à certains stress abiotiques comme ONSEN/ATCOPIA78 qui est activé par la température chez *Arabidopsis thaliana* :

<https://doi.org/10.1371/journal.pgen.1004115>

<https://genomebiology.biomedcentral.com/articles/10.1186/s13059-016-1072-3>

La base graphe permet d'identifier quelles sont les familles d'ET spécifiquement liées à des TFBS. Une analyse approfondie sur ces familles d'ET et leurs relations avec les gènes et réseaux de gènes impliqués dans des traits d'adaptation permettra d'améliorer la prédiction de leur impact fonctionnel chez leur hôte.

Ce stage s'inscrit dans cette dynamique et plus particulièrement sur le rôle des éléments transposables dans l'adaptation au changement climatique.

Objectifs :

Construction d'un catalogue de TFBS liés à des familles d'éléments transposable.

Prédiction de l'impact fonctionnel des copies d'ET qui portent de TFBS sur les gènes à proximité, à l'aide de données de méthylome et de transcriptome.

Travail demandé :

Le(a) candidat(e) devra analyser des données -omiques disponibles pour notamment extraire des informations de co-localisation entre différentes sources de données.

Il(elle) travaillera à insérer de nouvelles données qui pourraient être nécessaires dans la base graphe pour répondre à la problématique posée.

Le(a) stagiaire acquerra des compétences en développement de pipelines ainsi que sur les systèmes de gestion de données basé sur les graphes.

Compétences techniques recherchées :

- Maîtrise des commandes UNIX (shell) et de la programmation python.
- Connaissance en SGBD souhaitable notamment NoSQL (neo4J).
- Connaissance de la technologie Docker souhaitable

Ce sujet constitue un premier pas vers un travail de thèse : Oui

Date de début du stage et durée estimée du stage : A partir de début 2023 de 6 mois.

Montant (brut mensuel) de la rémunération proposée :

Indemnité de stage selon barème en vigueur (environ 550 euros net par mois)

Date de la proposition de stage et date limite de candidature :

Proposition de stage fin septembre 2022

Date limite fin décembre 2022