



Département de Statistique et Informatique



UMR 6625 CNRS



Offre de stage de niveau M2 à l'Institut Agro Rennes Angers, France

Sujet: Apprentissage de modèles d'association entre épigénome et transcriptome

Mots-clés: Données fonctionnelles, Epigénétique, Grande dimension, Machine learning, Tests multiples.

Contexte

Dans le contexte actuel de réchauffement global, les organismes qui coexistent au sein des agro-écosystèmes et plus particulièrement les pathogènes et ravageurs des plantes doivent faire face à d'importantes variations climatiques avec des conséquences potentiellement néfastes sur leur physiologie et leur potentiel de nuisibilité. Face à ces épisodes dont la fréquence et la durée sont difficiles à anticiper, les organismes doivent répondre rapidement afin de mettre en place des phénotypes leur permettant de survivre en s'acclimatant.

A des échelles de temps longues, la création de variabilité génétique par recombinaison constitue une option pour pérenniser de nouveaux génotypes et phénotypes adaptés à ces conditions extrêmes. En revanche, à des échelles de temps plus courtes, des mécanismes épigénétiques (changements d'accessibilité de la chromatine, marques de modifications post-traductionnelles des histones et méthylation de l'ADN) constituent un moyen efficace et flexible pour modifier rapidement l'expression des gènes et produire de nouveaux phénotypes permettant de répondre à ces nouvelles contraintes environnementales. Ce code épigénétique est souvent étudié par des approches de séquençage haut-débit, générant de gros volumes de données de nature hétérogènes, pour lesquelles les méthodes d'analyses actuelles en limitent la compréhension.

Le projet EPIPREDICT propose de développer des approches statistiques et mathématiques innovantes afin d'identifier dans ces données les éléments qui permettent de décrire les variations de l'expression des gènes (notamment ceux responsables de la capacité de la virulence et de l'agressivité des pathogènes et ravageurs), en considérant les caractéristiques spatiales des génomes.

In fine, décoder comment les gènes s'expriment en réponse à l'environnement pourrait permettre de fournir un modèle d'aide à la décision pour développer des agro-écosystèmes résilients et économiquement viables.

Afin d'appuyer la réflexion sur ces approches, nous disposons de données transcriptomiques (RNA-seq) et épigénomiques (ChIP-seq ou CUT&RUN), issues de technologies de séquençage à haut débit, obtenues sur un champignon filamenteux pathogène du blé et producteur de mycotoxines, *Fusarium graminearum*. Notre premier objectif est d'explorer les relations à l'échelle du génome entre expression de gènes et diverses

marques histones afin de caractériser finement à l'échelle moléculaire l'état épigénomique et transcriptionnel du génome de cet organisme.

Les signaux obtenus par ces approches sont par nature hétérogènes et complexes à décrire. L'étude de l'association entre l'expression d'un gène et les données épigénomiques s'y rapportant doivent tenir compte de connaissances biologiques préalables mais aussi d'un modèle pour données fonctionnelles rendant compte de la forte dépendance, à la fois entre les gènes dont les expressions sont mutuellement régulées (on parle de réseau) mais aussi au sein même des signaux épigénomiques et d'expression. La grande dimension des données couplée à cette double structure de dépendance définit un nouveau paradigme pour élaborer des méthodes de tests multiples et de prédiction optimales.

Département de Statistique et Informatique

L'équipe de recherche en Statistique (7 membres) de l'institut Agro (Rennes) est intégrée à l'équipe Statistique (environ 40 membres) de l'Institut de Recherche en Mathématique de Rennes (Irmr, UMR 6625 CNRS). La recherche en statistique à l'Institut Agro Rennes Angers est particulièrement orientée vers l'apprentissage de données de grande dimension, l'analyse de données génomiques, écologiques et de signaux biophysiques.

Plus d'information : <https://www.institut-agro-rennes-angers.fr/> et <https://www.irmar.univ-rennes1.fr>

Détails

Encadrement : David Causeur (Irmr, Rennes), Nadia Ponts (UR MycSA, INRAE Bordeaux), Gaël le Trionnaire (UMR IGEPP, INRAE Rennes)

Durée : 6 mois (Début : Février ou Mars 2023)

Date limite de candidature (CV + lettre de motivation) : 31 Janvier 2023

Lieu : Institut Agro Rennes Angers, 65 rue de St-Brieuc, CS84215, 35042 Rennes cedex

Rémunération : 550 €/mois

Compétences requises : Expérience en R et intérêt marqué pour les statistiques appliquées

Contact: David Causeur,

Email: david.causeur@agrocampus-ouest.fr

Url: <https://dcauseur.netlify.app>
