

## Stage Master 2

# Méthodes d'apprentissage pour l'identification de QTLs impliqués dans les mécanismes de résistance des plantes aux pathogènes

### Contexte

La compréhension des mécanismes génétiques de résistance des plantes aux pathogènes est un enjeu crucial pour le développement de variétés résistantes dans l'objectif de réduire l'utilisation des pesticides. On souhaite par ailleurs identifier des mécanismes de résistance robustes, i.e. susceptibles de perdurer en condition de température élevée correspondant à un scénario de changement climatique. Toutefois il n'existe actuellement que peu de méthodes dédiées à l'identification conjointe de régions génomiques impliquées dans la réponse immunitaire chez le pathogène et la plante.

### Objectifs du stage

Les objectifs du stage sont les suivants :

- i) identifier simultanément, chez les deux partenaires de l'interaction plante-pathogène les régions génomiques impliquées dans la réponse immunitaire des plantes et dans les mécanismes de virulence des pathogènes,
- ii) déterminer si l'efficacité des mécanismes de résistance aux pathogènes précédemment identifiés est stable lorsque la plante (et le pathogène) est soumise à des conditions de températures élevées.

Le candidat devra développer des procédures de sélection de variables d'une part dans le cadre d'un modèle linéaire généralisé régularisé (type sélection de variables pour la régression ridge), et d'autre part dans le cadre d'un modèle de réseau de neurones profond multi-classes. Les procédures développées devront être implémentées de manière efficace afin de passer à l'échelle et d'être appliquées à de gros jeux de données.

Les procédures développées seront appliquées à l'analyse d'un jeu de données expérimentales constitué

- i) d'une collection constituée de 24 variétés d'une espèce sauvage de tomate (*Solanum pimpinellifolium*), chaque variété étant génotypée pour 3,7 millions de SNPs (i.e. la séquence de chaque variété a été lue en 3,7 millions de positions couvrant l'ensemble du génome);
- ii) d'un panel de 96 souches de la bactérie phytopathogène *Ralstonia solanacearum*, chaque souche de bactérie étant génotypée pour 120,000 SNPs.

Des mesures de résistance des 24 variétés de tomate ont été obtenues par inoculation de chaque variété par chacune des 96 souches de la bactérie à deux températures (28°C et 32°C) lors de 3 répétitions biologiques.

### **Compétences recherchées**

Solides compétences en apprentissage statistique

Bonne maîtrise des méthodes d'analyse multivariée, du modèle linéaire et de ses extensions.

Connaissances des modèles de réseau de neurones.

Programmation en R / Python.

Intérêt pour les applications statistiques en génétique.

### **Unité d'accueil, ressources mises à disposition**

Ce stage s'inscrit dans le cadre d'une collaboration entre l'équipe SOLsTIS de l'UMR MIA-Paris et l'équipe REACH (plant REsistance pathways dynamics and Adaptation to Climate cHange) du Laboratoire des Interactions Plantes Microorganismes Environnement (LIPME) du centre INRAE Occitanie. Il sera encadré par Céline Lévy-Leduc (Professeur AgroParisTech en statistique), Tristan Mary-Huard (Chercheur INRAE en statistique) et Richard Berthomé (Chercheur INRAE).

Le stage se déroulera dans l'UMR de statistique d'AgroParisTech (Campus Paris Saclay). Le stagiaire disposera d'un ordinateur personnel et pourra utiliser les ressources informatiques (serveurs + cluster de calcul) de l'unité. Le stagiaire percevra la gratification INRAE. La durée du stage (entre 5 et 6 mois) et la date de commencement peuvent être adaptées en fonction des contraintes du candidat.

### **Poursuite en thèse**

Le stage pourra être suivi par une thèse, qui fera l'objet d'une collaboration avec l'entreprise Syngenta.

### **Contacts**

Tristan Mary-Huard, maryhuar@agroparistech.fr

Céline Levy-Leduc, celine.levy-leduc@agroparistech.fr

### **Références**

[1] Wang, Miaoyan and Roux, Fabrice and Bartoli, Claudia and Huard-Chauveau, Carine and Meyer, Christopher and Lee, Hana and Roby, Dominique and McPeck, Mary Sara and Bergelson, Joy. Two-way mixed-effects methods for joint association analysis using both host and pathogen genomes, PNAS, vol. 115, n. 24, p. E5440--E5449, 2018, doi 10.1073/pnas.1710980115.

[2] George, Andrew W., Arunas Verbyla, and Joshua Bowden. "Eagle: multi-locus association mapping on a genome-wide scale made routine." *Bioinformatics* 36.5 (2020): 1509-1516.