

Modélisation et estimation statistique pour l'analyse de la variabilité de la réponse au virus de la sharka chez l'abricotier dans le cadre d'études de génétique d'association génotype/phénotype incluant des marqueurs de grande dimension

Stage niveau Master 2, début 2023

Unité Mathématiques et informatique appliquées
du génome à l'environnement (MaIAGE)
INRAE, Jouy-en-Josas

Pour postuler, merci d'envoyer un dossier complet contenant votre CV, votre lettre de motivation et vos derniers relevés de notes de Master 1 ou équivalent à Estelle Kuhn (estelle.kuhn@inrae.fr) et à Maud Delattre (maud.delattre@inrae.fr).

Contexte

La sharka est une maladie de quarantaine touchant les arbres fruitiers de la sous-famille des Prunoideae, dont les abricotiers, pêchers, pruniers, amandiers. Elle est due à l'infection par un virus, le Plum Pox Virus (noté PPV). L'abricotier cultivé (*Prunus armeniaca*) en Europe est aujourd'hui isolé des populations naturelles à partir desquelles il a été domestiqué, qui sont situées en Asie Centrale. Dans le cadre d'un projet européen STONE, ces populations sauvages ont été échantillonnées. Une collection d'abricotiers sauvages a été placée en collection dans l'unité INRAE Biologie du Fruit et Pathologie à Bordeaux, génotypés au niveau du génome complet par re-séquençage ILLUMINA et phénotypés pour leur résistance au virus de la sharka. Des tests de phénotypage dits « sévères » ont été réalisés par greffage sur porte-greffe malades de 3 à 4 réplicas par génotype. Les évaluations ont consisté à réaliser des tests sérologiques semi-quantitatifs de type ELISA à raison de 2 tests successifs par cycle et ce pour 3 cycles végétatifs consécutifs. Un cycle végétatif correspond à une succession de 3 mois de dormance, en chambre froide, et de 3 à 4 mois de croissance en serre de confinement. La réponse à l'infection virale chez l'abricotier sauvage se caractérise par :

- Son taux relativement élevé de génotypes résistants en comparaison de l'abricotier cultivé
- Sa grande hétérogénéité de réponse selon les réplicas, les tests sérologiques et les cycles végétatifs

Des matrices de polymorphismes nucléotidiques (marqueurs SNP) ont par ailleurs été obtenues par l'INRAE de Bordeaux, à partir des séquences de génome complet et après alignement sur une séquence de référence abricotier. Sur la base des données quantitatives générées lors du phénotypage, il convient à présent de modéliser la réponse au virus chez cette plante hôte pérenne. Il est attendu que ceci permettra de proposer des modèles d'études statistiques adaptés à ce pathosystème pour les approches de génétique d'association futures. Les données de génotypage et phénotypage de plus d'une centaine d'abricotiers sauvages sont d'ores et déjà disponibles.

Objectifs du stage

- modéliser la dynamique de réponse à l'infection en intégrant l'hétérogénéité des réponses au sein d'un modèle de mélange
- effectuer un clustering des individus sur la base de leurs réponses au virus
- prédire les probabilités a posteriori d'appartenance aux classes du mélange
- intégrer des covariables de grande dimension (marqueurs SNP) dans le modèle de mélange et sélectionner les variables pertinentes
- intégrer la structuration de la population au modèle via la matrice d'apparement
- intégrer les informations obtenues lors d'approches de type Genome Wide Association Study (GWAS) menées par l'unité INRAE BFP de Bordeaux.

L'approche envisagée reposera dans un premier temps sur des modèles de mélange et des modèles à effets mixtes pour prendre en compte les variabilités présentes dans les données. L'estimation des paramètres sera réalisée via des algorithmes numériques adaptés et efficaces. Des modèles à effets mixtes intégrant des covariables de grande dimension seront également utilisées pour analyser les effets des marqueurs SNP via une estimation pénalisée type LASSO.

Profil recherché

Formation niveau BAC+5 (Master 2 ou école d'ingénieurs), connaissance nécessaire en mathématiques appliquées et statistiques, connaissance souhaitée en génétique, ayant un intérêt pour le travail à l'interface et les applications en sciences du vivant; maîtrise d'un langage de programmation indispensable; rigueur scientifique, curiosité intellectuelle, facilité de communication.

Modalités pratiques

Le stage se déroulera principalement au centre INRAE de Jouy-en-Josas dans l'unité MaIAGE et de façon ponctuelle dans l'unité INRAE Biologie du fruit et Pathologie (BFP) de Bordeaux. La durée du stage sera de cinq ou six mois, à partir de janvier 2023. La gratification mensuelle est d'environ 550 euro (taux légal). L'encadrement sera réalisé par Estelle Kuhn et Maud Delattre (MaIAGE, INRAE), Véronique Decroocq, Stéphane Decroocq (BFP, INRAE) et Benjamin Brachi (BIOGECO, INRAE).

Références bibliographiques

- [1] Groppi A., Liu S., Cornille A., Tricon D., Bui Q. T., Cruaud C., Arribat S., Belser C., Marande W., Salse J., Huneau C., Rodde N., Rhalloussi W., Cauet S., Istace B., Denis E., Carrère S., Audergon J.-M., Roch G., Lambert P., Zhebentyayeva T., Liu W.-S., Bouchez O., Lopez-Roques C., Serre R.-F., Debuchy R., Tran J., Wincker P., Chen X., Pétriacq P., Barré A., Nikolski M., Aury J.-M., Abbott A., Giraud T., Decroocq V. (2021). Population genomics of apricots unravels domestication history and adaptive events. *Nature Communications*, 12 (1), 3956.
- [2] Mariette S., Wong Jun Tai F., Roch G., Barré A., Chague A., Decroocq S., Groppi A., Laizet Y., Lambert P., Tricon D., Nikolski M., Audergon J.-M., Abbott A., Decroocq V. (2015). Genome-wide association links candidate genes to resistance to Plum Pox Virus in apricot (*Prunus armeniaca*). *New Phytologist*, 209 (2), DOI : 10.1111/nph.13627
- [3] Onogi, A. (2020) Connecting mathematical models to genomes : joint estimation of model parameters and genome-wide marker effects on these parameters. *Bioinformatics*, 36(10), 3169-3176.
- [4] Taghipoor, M., Delattre, M., Giger-Reverdin, S. (2020). A novel modelling approach to quantify the response of dairy goats to a high-concentrate diet. *Scientific reports*, 10(1), 1-11.
- [5] Kuhn E., Matias C., Rebafka T. (2020) Properties of the Stochastic Approximation EM Algorithm with Mini-batch Sampling, *Statistics and Computing*, Vol. 30, pp 1725–1739.