

Intégration de données 'omiques multi-niveaux pour expliquer l'efficacité alimentaire chez les agneaux

Proposition de stage niveau M2 (printemps 2020)

Unité MIAT, INRAE Toulouse

Pour postuler envoyer CV, lettre de motivation et dernier relevé de notes à nathalie.vialaneix@inrae.fr et jerome.mariette@inrae.fr

Problématique du stage

Ce stage s'inscrit dans le cadre du projet MILAGE (financement métaprogramme MEM d'INRAE) qui s'intéresse à comprendre et prédire l'efficacité alimentaire chez les agneaux, c'est-à-dire, la capacité à bien grandir/grossir avec une alimentation réduite. Pour cela, un dispositif a été mis en place permettant de collecter des données à plusieurs niveaux de l'échelle du vivant sur deux lignées divergentes pour l'efficacité alimentaire. Ces données mesurent des informations génotypiques, métabolomiques, métagénomique et phénotypique chez un grand nombre d'animaux pour deux types de régimes alimentaires. L'objectif du stage sera de réaliser des modèles prédictifs intégrant ces différents omiques et permettant de prédire la lignée divergente, d'expliquer le régime alimentaire et, in fine, de prédire l'efficacité alimentaire.

Travail à effectuer

L'étudiant-e aura pour première tâche de s'approprier les données (en particulier, génotypes, métabolomes et métagénome) et de réaliser des analyses exploratoires (ACP, par exemple) pour identifier d'éventuels biais techniques indésirables et les corriger.

La deuxième partie du stage consistera à mettre en œuvre des méthodes de sélection de variables, au sein de chaque omique dans un premier temps, puis multi-omiques, pour identifier les groupes de variables les plus prédictifs de la divergence entre les deux lignées d'animaux. Les méthodes utilisées pour cette partie pourront inclure, par exemple, des approches de type « Lasso ». Une attention particulière sera portée, entre l'étape intra-omique et l'étape d'intégration, à comprendre les relations existants entre les variables sélectionnées au sein de deux omiques différents. Pour cela, des approches factorielles de type Analyse Canonique des Correspondances (CCA) ou bien de type PLS pourront être utilisées.

L'implémentation des méthodes ainsi que le travail d'analyse seront effectués à l'aide du langage de programmation R (principalement) et également (de manière secondaire) de Python. L'étudiant-e pourra être amené à utiliser des outils développés par les encadrants du stage.

Conditions du stage

durée 4 à 6 mois

localisation Unité MIA-T, INRAE Toulouse (Castanet-Tolosan)

rémunération taux légal (environ 550 euros par mois)

encadrement Nathalie Vialaneix et Jérôme Mariette

Contact : nathalie.vialaneix@inrae.fr et jerome.mariette@inrae.fr

Profil recherché

- Master 2 en statistique ou bioinformatique ou équivalent en école d'ingénieur ;
- maîtrise de R ;
- bonnes connaissances en statistique, machine learning ;
- des connaissances en biologie moléculaire seraient un plus.