

Analyse de données métabolomiques

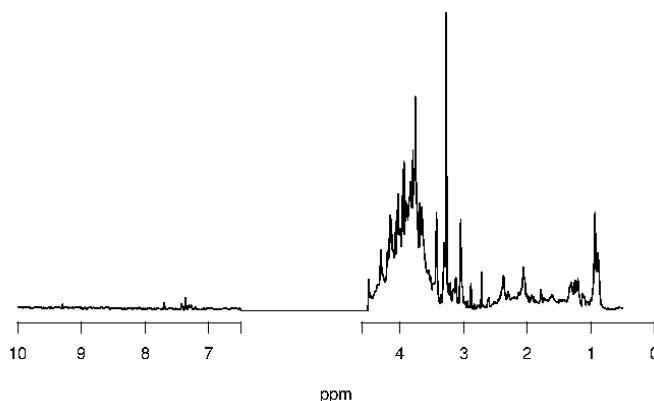
Magali San Cristobal, Laurence Liaubet & Nathalie Villa-Vialaneix
Unité GenPhySE, INRA de Toulouse

Proposition de stage niveau M2

Problématique du stage

Ce stage s'inscrit dans le cadre d'un projet dont le but est d'explorer l'impact sur l'organisme de porcs d'élevage d'une augmentation de la température ambiante. La problématique du stage est en relation directe avec des questions relatives à la robustesse des animaux face à l'augmentation de la température et a donc une importance cruciale en élevage dans un contexte de changement climatique.

Pour ce faire, des données numériques correspondant à la mesure de la quantité d'un grand nombre de métabolites¹ ont été obtenues à l'aide d'une technologie haut débit, pour plusieurs animaux, dans des conditions de température diverses (tempéré / tropical et normal / vague de chaleur). Les données se présentent sous la forme d'une fonction présentant de nombreux pics comme dans l'exemple ci-dessous :



Dans cette figure, les pics peuvent être interprétés comme correspondant à une quantité mesurée d'un métabolite donné pour un animal.

La problématique de ce stage sera d'analyser, dans ce contexte, de mettre en œuvre des méthodes statistiques classiques d'exploration et de nettoyage des données. Ensuite, le sujet s'orientera sur la mise en relation entre ces données métabolomiques et des phénotypes (poids, températures corporelles, ...) observés sur les animaux sous l'angle de la recherche de métabolites expliquant les différences et aussi sous l'angle de la recherche d'un prédicteur permettant d'expliquer un phénotype d'intérêt à partir de l'observation des données métabolomiques.

Travail à effectuer

Le stagiaire travaillera sous le logiciel R et commencera par une étude exploratoire des données (en utilisant, par exemple, des ACP, des PLS), pour détecter d'éventuels biais techniques dans les données, qu'il devra corriger avec les méthodes classiques utilisées dans ce domaine, et pour explorer les relations entre métabolome et données phénotypiques.

La deuxième partie du travail consistera à identifier des pics en lien avec des phénotypes d'intérêt. Diverses approches pourront être utilisées, sur les données brutes ou sur une transformation par ondelettes de ces données. Enfin, des méthodes de régression adaptées seront utilisées pour construire des modèles prédictifs de phénotypes d'intérêt.

1. Un métabolite est une petite molécule, comme le glucose, que l'on trouve dans les tissus d'un organisme vivant.

Conditions du stage

durée 4 à 6 mois

localisation Unité GenPhySE, INRA de Toulouse (Castanet-Tolosan)

rémunération taux légal

encadrement Magali San Cristobal, Laurence Liaubet, Nathalie Villa-Vialaneix

Contact : envoyer CV et lettre de motivation aux deux adresses e-mails : mhc@toulouse.inra.fr et nathalie.villa@toulouse.inra.fr

Profil recherché

- niveau équivalent Master 2 en mathématiques appliquées ;
- très bonne connaissance du logiciel de calcul scientifique R ;
- une spécialisation en biostatistique ou des connaissances sur l'analyse de données biologiques serait un plus mais n'est pas indispensable : une formation basique aux problématiques dans ce domaine est prévue en début de stage.