



Proposition de stage
2019-2020

**Analyses multivariées de cubes de données omiques appareillées :
Application à des données transcriptomiques multi-tissus**

Denis Laloë et Andrea Rau

Génétique Animale et Biologie Intégrative (GABI)

Domaine de Vilvert 78352 Jouy-en-Josas Cedex

Téléphone : 01 34 65 28 15

E-mail : andrea.rau@inra.fr ; denis.laloe@gmail.com

Site internet : <https://www6.jouy.inra.fr/gabi/>

Directrice du Laboratoire ou de l'Unité : Claire Rogel-Gaillard

Nom de l'équipe d'accueil : GIBBS

Responsable de l'équipe: Denis Laloë

Thèmes de recherche de l'équipe

L'équipe "Génomique, Biodiversité, Bioinformatique et Statistique" (GIBBS) conduit des recherches sur l'histoire et la diversité génétique des populations animales domestiques, sur les analyses statistiques intégrant les variabilités génétique et transcriptomique en lien avec l'adaptation animale aux stressés abiotiques. Des approches variées sont développées pour caractériser la variabilité génétique des populations, gérer les populations et valoriser les ressources génétiques. L'équipe connecte génomique de la population et les bases biologiques de l'adaptation. Enfin, elle propose une cellule d'appui bioinformatique et biostatistique pour répondre aux besoins des biologistes de l'unité GABI et au-delà.



Mots clés: Intégration de données, transcriptomique, analyses triadiques

Responsables du stage :

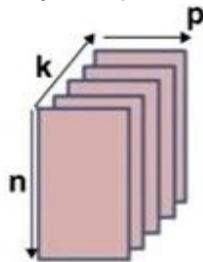
- Denis Laloë Email: denis.laloe@inra.fr
- Andrea Rau Email: andrea.rau@inra.fr

Projet de stage :

Contexte du projet

Les approches omiques cherchent à appréhender le vivant dans sa complexité, en intégrant des données de sources et de niveaux biologiques différents, comme la génomique, la transcriptomique, la protéomique et la métabolomique. Ces approches sont essentielles pour l'identification de biomarqueurs ou la mise en évidence de mécanismes impliqués dans des stress ou des maladies.

Jusqu'à récemment, le coût élevé des technologies à haut-débit pour mesurer l'activité transcriptomique a fortement limité les plans d'expérience possibles, et la plupart des études réalisées se limitait à des comparaisons simples entre deux groupes (e.g., des individus sains versus malades). Aujourd'hui, les réductions de coûts et améliorations technologiques commencent à permettre des plans d'expérience de plus en plus sophistiqués pour répondre à des questions biologiques complexes. Un exemple notable est la réalisation des études transcriptomiques sur les mêmes individus à partir de différents organes, tissus, ou à différents temps, afin de pouvoir étudier la variabilité transcriptomique à des échelles spatiales, temporelles, et intra-individuelles. En particulier, ce type d'étude doit permettre d'appréhender les patrons d'expression des gènes, de leur dynamique et de leur complexité.



Du point de vue statistique, les données correspondantes à ce plan d'expérience sont organisées selon un cube, ou un tableau à trois dimensions (cf illustration ci-dessus): individus (n) \times gènes (p) \times tissus (k). L'analyse triadique partielle est une extension de l'analyse en composantes principales (ACP), qui a été développée pour l'analyse spécifique des cubes de données. Il s'agit d'une approche multivariée exploratoire, principalement utilisée en écologie ou en hydrologie, et dont l'application en génomique représente un point novateur. Le but de ce stage est d'étudier le potentiel de cette méthode pour évaluer la variabilité du transcriptome (notamment ressemblances et dissemblances entre tissus), et sa capacité à établir une typologie des tissus. Une attention particulière sera portée à l'étude de la plus-value de cette approche comparée à l'état de l'art en analyses génomiques (e.g., ACP, analyse factorielle multiple, analyses différentielles).

On s'appuiera sur deux ensembles de données existantes : (1) le transcriptome de 21 tissus chez la poule, en collaboration avec une biologiste dans notre équipe (T. Zerjal); et (2) le transcriptome de plusieurs dizaines de tissus différents chez l'humain du projet Genotype-Tissue Expression (GTEx ; données publiques).

**Compétences requises :**

- Aptitude à l'analyse des données omiques
- Connaissance des analyses multivariées (ACP et extensions)
- Connaissance du logiciel R
- Bon niveau d'anglais (capacité à lire des articles scientifiques)
- Motivation et intérêt pour les objectifs du projet.

Le master se déroulera dans un environnement de recherche associant bioinformaticiens, biostatisticiens et biologistes.

La supervision sera assurée par Denis Laloë et Andrea Rau.

Rémunération :

- Prime de stage pour un Master 2 (environ 520 € net/mois)