

# Offre de stage Master 1 ou Master 2

## Modélisation statistique pour la comparaison de transcriptomes inter-espèces :

- **Durée :** 4 à 6 mois.
- **Profil du candidat :** Master 1 ou 2 disposant de compétences en statistiques et programmation R, et d'un intérêt pour l'analyse de données biologiques.
- **Contact :** Mélina Gallopin (melina.gallopin@u-psud.fr), maître de conférence à l'Université Paris Sud et membre de l'équipe bioinformatique de l'Institut de Biologie Intégrative de la Cellule (I2BC).
- **Lieu :** Campus de l'Université Paris Sud, RER B Orsay-Ville.
- **Gratification :** Environ 500 euros par mois.

## Présentation du sujet :

Les techniques de séquençage à haut débit permettent de séquencer les génomes et transcriptomes (RNA-Seq) de tous les organismes vivants. Ceci est une révolution dans le domaine de la biologie comparative, et permet d'accéder aux bases moléculaires de l'évolution et de l'adaptation des espèces. Cependant, ces études restent encore largement qualitatives car il n'existe actuellement pas de méthode statistique validée intégrant des données RNA-Seq pour comparer de manière quantitative les transcriptomes d'espèces éloignées. Pour effectuer ce type d'analyse, des groupes de gènes orthologues possédant des fonctions similaires (i.e. Cluster of Orthologous Genes ou COGs) ont été identifiés à partir de transcriptomes, produits par assemblage de novo (1). L'objectif du stage est de proposer une modélisation du niveau d'expression de ces COGs, prenant compte de tous les biais possibles introduits par les différentes étapes de la quantification de ces niveaux d'expression (2). La méthode sera développée et validée à partir de sets de données RNA-Seq cohérents déjà obtenus à partir d'espèces végétales (3), et sera facilement transposable à d'autres organismes et d'autres problématiques scientifiques. L'outil développé pourra être mis à disposition de la communauté sous la forme d'un package R CRAN ou Bioconductor.

## Bibliographie :

- (1) Pereira C, Denise A, Lespinet O (2014) A meta-approach for improving the prediction and the functional annotation of ortholog groups. *BMC Genomics*, 15 Suppl 6:S16.
- (2) Dillies MA , et al. (2013) A comprehensive evaluation of normalization methods for Illumina high-throughput RNA sequencing data analysis. *Briefings in Bioinformatics*, 14(6):671-683.
- (3) <https://bioweb.i2bc.paris-saclay.fr/evometonicks/>