

## STAGE DE RECHERCHE M2 ECOLOGIE EVOLUTION GENOMIQUE Rentrée 2017

---

### Epissage alternatif dans les génomes animaux : diversité fonctionnelle ou fardeau génétique ?

Université Claude Bernard Lyon 1  
Laboratoire de Biométrie et Biologie Evolutive UMR CNRS 5558 Villeurbanne  
Encadrant : Laurent Duret ([laurent.duret@univ-lyon1.fr](mailto:laurent.duret@univ-lyon1.fr))

#### Description du projet :

Chez les eucaryotes, la majorité des gènes sont soumis à des phénomènes d'épissage alternatif. Ce processus contribue à augmenter le répertoire protéique codé par les génomes, et permet également de réguler le niveau d'expression de certains gènes. Cependant, les variants d'épissage ne sont pas tous fonctionnels : une fraction d'entre eux correspond à des erreurs de la machinerie d'épissage qui sont coûteuses pour l'organisme. La question de la signification fonctionnelle des variants d'épissage est l'objet d'un débat intense au sein de la communauté scientifique : dans quelle proportion ces variants d'épissage constituent-ils un fardeau ou bien sont-ils bénéfiques pour le développement des organismes ?

Ces deux modèles font des prédictions distinctes : sous l'hypothèse que les variants d'épissage sont majoritairement des erreurs, résultant de l'accumulation de mutations délétères, alors on s'attend à ce que ce phénomène soit plus fréquent chez les organismes qui sont plus fortement soumis aux effets de la dérive génétique (i.e. ceux dont la taille efficace des populations est faible). Inversement, si le rôle de l'épissage alternatif est d'accroître le répertoire fonctionnel des génomes alors on s'attend à ce que ce phénomène soit plus fréquent chez les organismes dont le développement est plus complexe.

Pour trancher entre ces hypothèses nous proposons d'analyser les transcriptomes d'un grand nombre d'espèces animales, afin d'analyser les relations entre fréquence d'épissage alternatif, complexité des organismes (mesurée par le nombre de types cellulaires différents) et taille efficace des populations (estimée à partir du polymorphisme génétique).

Nous disposons de séquences RNAseq provenant des transcriptomes de 90 espèces animales, présentant des tailles efficaces et des degrés de complexité contrastés. L'étudiant devra mettre en œuvre des pipelines bioinformatiques pour quantifier l'épissage alternatif au sein de ces transcriptomes. Il devra également développer des scripts pour gérer et analyser les données, et réaliser l'analyse statistique des données (R). Ce projet permettra donc à l'étudiant d'acquérir de l'expérience en analyse de données massives de séquences (bioinformatique, statistique), et de se former aux concepts de génomique évolutive, génomique des populations et génomique fonctionnelle.

#### Références :

1. Romiguier J, Gayral P, Ballenghien M, Bernard A, Cahais V, Chenuil A, et al. Comparative population genomics in animals uncovers the determinants of genetic diversity. *Nature*. 2014 ; 515 : 261–263. Doi : 10.1038/nature13685
2. Chen L, Bush SJ, Tovar-Corona JM, Castillo-Morales A, Urrutia AO. Correcting for differential transcript coverage reveals a strong relationship between alternative splicing and organism complexity. *Mol Biol Evol*. 2014 ; 31 : 1402–13. Doi : 10.1093/molbev/msu083