

Sujet de thèse Unité Evo-Eco-Paléo – campagne 2021

Université : Université de Lille, Faculté des Sciences et Technologies

Ecole doctorale: ED104 Sciences de la matière, du rayonnement et de l'environnement (SMRE)

Domaine scientifique, Spécialité : Biologie de l'environnement, des organismes, des populations

Titre de la thèse : Origine, évolution et rôle des ORFs dans les mécanismes d'apparition de nouveaux gènes et éléments régulateurs - *Origin, evolution and role of ORFs in the birth of new genes and regulatory elements*

Direction de thèse : Vincent Castric (directeur de thèse), Eléonore Durand et Sylvain Legrand (co-encadrants)

Email de contact: eleonore.durand@univ-lille.fr

Laboratoire(s) de Rattachement : Unité Evolution, Ecologie et Paléontologie, UMR CNRS 8198

Programme(s) de Rattachement : ANR-TE-MOMA, overhead ERC NOVEL

Financement acquis ou demandé : Université de Lille (demandé)

SUJET DE THESE

Un des enjeux majeurs en biologie évolutive est de comprendre les mécanismes impliqués dans l'innovation rapide des organismes. Dans ce contexte, nous étudions la dynamique d'émergence de nouveaux gènes et éléments régulateurs au sein du genre *Arabidopsis*. Les régions non-géniques sont riches en cadres de lecture ouverts (Open Reading Frames, ORFs) et certains sont exprimés et produisent des polypeptides. Ces ORFs ont plusieurs origines possibles : ils peuvent être apparus *de novo* dans une région intergénique, ou représenter des vestiges d'anciennes régions géniques en cours de pseudogénéisation. Ces ORFs intergéniques peuvent ainsi évoluer de façon neutre ou acquérir des caractéristiques qui pourront ensuite être sélectionnées. Nous proposons ici d'étudier l'évolution du répertoire en ORFs entre espèces proches du genre *Arabidopsis*, ainsi qu'au sein de ces espèces dans des populations, afin de suivre au plus près les caractéristiques récentes des ORFs associés à l'émergence de nouvelles fonctions, ou modes de régulation.

Ce projet consistera à i) caractériser le répertoire en ORFs et leur niveau d'expression chez plusieurs espèces d'*Arabidopsis*, ii) étudier leur conservation et les mécanismes mutationnels à l'origine de leur apparition, et enfin iii) identifier des gènes candidats apparus récemment et les caractéristiques associées à cette origine récente. Nous nous intéresserons en particulier à une nouvelle catégorie d'ORFs localisés sur des transcrits de microARNs et impliqués dans la régulation de l'expression du transcrit hôte.

Compétences recherchées : Formation en génétique ou en génomique évolutive, analyse de jeux de données génomiques (langage Python ou R)

Mots clefs : Évolution des génomes, ORFs, naissance des gènes, microARNs, *Arabidopsis*, populations

Publications du laboratoire sur le sujet :

- Durand, É. *et al.* Turnover of ribosome-associated transcripts from *de novo* ORFs produces gene-like characteristics available for *de novo* gene emergence in wild yeast populations. *Genome Res* Jun 29; 932-943 (2019)
- Legrand, S. *et al.* Differential retention of transposable element-derived sequences in outcrossing *Arabidopsis* genomes. *Mobile DNA*, 10:30 (2019)
- Durand, E., Meheust, R. *et al.* Dominance hierarchy arising from the evolution of a complex small RNA regulatory network. *Science* 346, 1200–1205 (2014)

PhD SUBJECT

Origin, evolution and role of ORFs in the birth of new genes and regulatory elements

Understanding the genetic basis for the origin and evolution of novel phenotypes and, ultimately, biological diversity is one of the most fundamental questions in evolutionary biology. In this context, we are studying the dynamics of emergence of new genes and regulatory elements in the *Arabidopsis* genus. Open Reading Frames (ORFs) are frequently found in non-genic regions and some of them are known to be expressed and produce polypeptides. These ORFs may have different origins: they may emerge *de novo* from an intergenic region, or they may arise from an ancient gene undergoing pseudogenization. The repertoire of non genic ORFs is very dynamics, they can evolve randomly or acquire some characteristics which can subsequently be selected. Here we propose to study the evolution of the ORFs repertoire in closely related *Arabidopsis* species or populations, to closely examine the early characteristics of newly emerging genes or regulatory elements.

We will i) characterize the ORFs repertoire and their expression level in closely related *Arabidopsis* species, ii) study

their conservation and their mutational origins, and finally iii) identify candidate newly emerging genes and the characteristics associated with this recent origin. We are particularly interested in a new class of ORFs localized on microRNA transcripts and involved in the regulation of its own transcript.

Requested skills: Training in evolutionary genetics or genomics, analysis of genomic datasets (Python or R)

Key words: Evolutionary genomics, ORFs, gene birth, miRNAs, Arabidopsis, populations