



DIPLÔME NATIONAL DE DOCTORAT

(Arrêté du 25 mai 2016)

Date de la soutenance : 15 février 2017

Nom de famille et prénom de l'auteur : LOPEZ Hélène

Titre de la thèse : « Analyses et méthodes pour les données transcriptomiques issues d'espèces non modèles : variation de l'expression des éléments transposables (et des gènes) et variants nucléotidiques. »



RESUME DE LA THESE

Le développement de la seconde génération de séquenceurs haut débit a généralisé l'accès à l'étude du transcriptome via le protocole RNAseq. Celui-ci permet d'obtenir à la fois la séquence et l'abondance des transcrits d'un échantillon. De nombreuses méthodes bioinformatiques ont été et sont encore développées pour permettre l'analyse des données issues du RNAseq et en tirer le maximum d'information. Ce type d'analyse est notamment possible sans utiliser de génome de référence, et donc pour les espèces modèles ou non-modèles, grâce à des méthodes d'assemblage.

Durant ma thèse, j'ai principalement travaillé à partir de données RNA-seq issues d'espèces non modèles. Je me suis intéressée dans un premier temps à l'impacte de l'hybridation inter spécifique sur la stabilité des génomes chez les hybrides issus des croisements réciproques de *D. mojavensis* et *D. arizonae*. Nos résultats ne montrent pas une dérégulation globale, mais plutôt quelques gènes et éléments transposables qui sont spécifiquement dérégulés. La pipeline d'analyse mis en place ici sera réutilisée pour l'étude des niveaux d'expression des transcrits chez les mâles ainsi que pour les croisements issus d'autres lignées de *D. mojavensis* avec *D. arizonae*, conduisant à une fertilité variable chez les hybrides.

Dans un second temps, j'ai participé à la validation du logiciel KisSplice pour la détection de SNP dans des données RNA-seq sans génome de référence. Celui-ci permet de trouver différents types de variants (épissage, indels) directement dans le graphe de de Bruijn construit à partir des lectures séquencées. J'ai également participé au développement d'outils de post-traitement permettant de prédire l'impact des SNP sur les protéines.