



Université Claude Bernard



# DIPLÔME NATIONAL DE DOCTORAT

(Arrêté du 25 mai 2016)

Date de la soutenance : **5 décembre 2017**

Nom de famille et prénom de l'auteur : **ARELLANO DAVIN Adrian**

Titre de la thèse : « *Utilisation des transferts horizontaux de gènes pour dater des phylogénies* »



## Résumé

### Utilisation des transferts horizontaux de gènes pour dater des phylogénies

Le fait d'avoir une généalogie datée des organismes vivants est l'un des principaux objectifs de la biologie évolutive. Cette entreprise est confrontée à deux défis majeurs. Le premier est la rareté et l'incomplétude des enregistrements fossiles, pratiquement inexistantes pour les microbes et essentiels pour fournir une échelle temporelle de l'histoire de la vie. Le second est la difficulté intrinsèque d'obtenir des phylogénies d'organismes dont le génome a été largement façonné par transfert latéral de gène (TLG). L'acquisition par transfert de nouveaux gènes d'origine éloignée perturbe des arbres de gènes et rend beaucoup plus complexe la reconstruction de l'histoire des espèces. Dans ce travail de thèse, je montre comment nous pouvons utiliser ces différences entre arbres de gènes et arbres d'espèces à notre avantage pour inférer les événements anciens de TLG et comment ils peuvent fournir une nouvelle échelle de temps pour l'évolution des organismes vivants. Les transferts étant particulièrement fréquents chez les espèces dont les fossiles sont rares, ils peuvent servir de nouvelle source de datation indépendante du registre géologique pour reconstruire une phylogénie datée de la vie.

Dans la première partie, je réalise une analyse à l'échelle génomique pour montrer comment les méthodes de réconciliations phylogénétiques peuvent être utilisées pour détecter les lignées correspondant aux donneurs et aux receveurs des événements de TLG. En outre, ces méthodes fournissent également une vue détaillée de la façon dont les familles de gènes évoluent le long des arbres de l'espèce. En utilisant ALE, un logiciel de réconciliation probabiliste qui prend en compte l'incertitude dans les arbres de gènes, nous sommes en mesure de cartographier les événements de duplication, de perte et de transfert dans les phylogénies des cyanobactéries et des champignons. Nous montrons également comment les méthodes qui ignorent l'information contenue dans les arbres de gènes sous-estiment la fréquence réelle des TLG.

Dans la deuxième partie, je présente en détail comment le TLG porte un signal temporel et comment ce signal peut être utilisé pour inférer des arbres datés. J'introduis une nouvelle méthode appelée MaxTiC qui permet de trouver un ordonnancement des noeuds dans l'arbre des espèces qui maximise la cohérence temporelle entre les transferts. Par des simulations, nous montrons la robustesse de la méthode aux erreurs présentes dans l'arbre des espèces et le nombre de familles de gènes nécessaires pour obtenir des arbres datés fiables.

Enfin, pour confirmer nos résultats, je présente différentes approches permettant de comparer les temps de divergence découlant des transferts avec ceux estimés en utilisant des horloges

moléculaires. Nous effectuons une analyse phylogénomique pour détecter des milliers d'événements de TLG dans quatre groupes: les cyanobactéries, les Deltaproteobactéries, les Archées et les Champignons. Nous trouvons un large accord entre les deux méthodes de datation, ce résultat étant robuste à l'utilisation de différents prior sur les temps de divergence et différents modèles d'horloges moléculaires relâchées. Nous montrons également que certaines des dates indiquées par l'utilisation de TLG sont en désaccord avec les horloges moléculaires tout en étant soutenues par un grand nombre de TLG. Ces résultats suggèrent que l'utilisation des TLG pourrait permettre d'améliorer les méthodes de datation, notamment pour les phylogénies anciennes et ainsi conduire à d'importants changements de notre point de vue sur l'histoire de la vie.