



Université Claude Bernard



DIPLÔME NATIONAL DE DOCTORAT

(Arrêté du 25 mai 2016)

Date de la soutenance : **04 juillet 2022**

Nom de famille et prénom de l'auteur : **Monsieur MENET Hugo**

Titre de la thèse : « *Approches phylogénétiques multi-niveaux pour l'évolution d'un holobionte* »

Résumé

Les systèmes biologiques sont constitués d'entités à plusieurs niveaux d'organisation (macro-organismes, micro-organismes, gènes...), qui partagent une histoire commune par leur dépendance, mais sont aussi guidés par leurs intérêts individuels. La réconciliation phylogénétique est une façon d'aborder l'évolution d'un tel système en décrivant la coévolution de deux niveaux différents, gènes et espèces, ou hôtes et symbiotes par exemple. La réconciliation est cependant limitée à deux niveaux, et s'inscrit ainsi soit dans un contexte d'évolution moléculaire avec des arbres de gènes et d'espèces, soit dans un contexte écologique d'évolution des associations hôte-symbiote. Le concept d'holobionte, la prise en compte comme un tout d'un macro-organisme (plante ou animal notamment) et de tous les micro-organismes qui vivent et fonctionnent avec lui, est l'occasion de rassembler toutes ces échelles en modélisant des interdépendances à plusieurs niveaux. Le but de cette thèse est d'explorer et d'étendre la réconciliation pour modéliser de tels systèmes.

La réconciliation est une méthode phylogénétique née à l'intersection de deux communautés, l'une qui s'intéresse à la coévolution d'espèces en symbiose, et l'autre qui compare des arbres de gènes et d'espèces. Malgré ce développement initial commun, ces deux communautés ont tendance à peu interagir, même si elles ont beaucoup à apprendre l'une de l'autre.

Nous proposons dans cette thèse un état de l'art de la réconciliation phylogénétique en adoptant un point de vue générique et en soulignant les avancées vers des modèles plus intégratifs, qui tendent vers des méthodes intégrant plus de deux niveaux.

Parmi ces nouvelles méthodes, certaines proposent de modéliser ensemble l'évolution des espèces, des gènes et des domaines de gènes, ou encore d'imposer des contraintes géographiques à un système hôte symbiote. Cependant, aucune ne s'est pour le moment intéressée aux niveaux qui sont au cœur du concept d'holobionte : hôte, symbiote et gènes. D'un point de vue méthodologique, aucune ne fait appel à un cadre probabiliste permettant des transferts horizontaux.

Nous avons réimplémenté ALE, un logiciel probabiliste de réconciliation modélisant les événements de Duplication, Transfert horizontal et perte (Loss) (modèle DTL), et l'avons étendu pour considérer la réconciliation de trois niveaux : hôte, symbiote et gène. Ce nouveau modèle probabiliste, qui prend en entrée trois arbres, combine un modèle DTL pour la coévolution de l'hôte et du symbiote, et un modèle DTL pour l'évolution des gènes du symbiote. Nous avons conçu un algorithme de Monte Carlo pour construire des scénarios couplés et calculer leurs probabilités dans le modèle, en tenant notamment compte de la

dépendance des taux de transfert de gènes à la réconciliation entre symbiotes et hôtes, ainsi que de l'impact des lignées fantômes sur ces taux. Comme avec ALE, nous utilisons l'amalgamation pour tenir compte de l'incertitude dans les arbres de gènes, mais aussi pour inférer l'arbre de symbiote en utilisant les arbres des familles de gènes universels et unicopies comme une distribution de la topologie de l'arbre de leur génome.

Nous avons évalué cette méthode sur un jeu de données simulées, sur lequel nous avons montré qu'il était possible de distinguer les modèles de coévolution à 2 et 3 niveaux en utilisant la vraisemblance. La méthode est également capable sur des phylogénies de pucerons et de leurs entérobactéries de mieux retrouver les transferts de gènes que la méthode ignorant l'arbre d'hôte.

Il peut être difficile d'interpréter la sortie d'une méthode de réconciliation, notamment lorsque l'on considère plusieurs scénarios échantillonnés, plusieurs familles de gènes ou lorsqu'on s'intéresse à des systèmes à plusieurs niveaux. Peu de logiciels proposent une sortie graphique générique, utilisable avec plusieurs logiciels de réconciliation. Un premier pas dans cette direction est l'utilisation de RecPhyloXML, un format de scénario adopté par une partie importante de la communauté gène espèce. C'est ce format que nous utilisons dans notre implémentation.

Thirdkind, un logiciel que nous avons développé, est capable de produire une sortie graphique en SVG à partir d'un scénario de réconciliation en RecPhyloXML. Il est facile à utiliser et à installer, il peut afficher différentes vues représentant la réconciliation de trois niveaux, et résumer l'évolution de plusieurs familles de gènes ou de plusieurs scénarios échantillonnés dans une seule figure en agrégeant les transferts redondants.

Un exemple fascinant d'histoire complexe de coévolution est la relation entre *Helicobacter pylori* et son hôte humain. *Helicobacter pylori* est une bactérie pathogène qui aurait suivi *Homo sapiens* lors de ses migrations ancestrales : colonisation de l'Afrique, de l'Asie, de l'Europe, de l'Océanie et de l'Amérique.

Les souches bactériennes sont structurées en populations dont la répartition géographique est le plus souvent congruente avec celle de leur hôte.

L'une des exceptions significatives est la population européenne, qui semble résulter de l'introgession entre deux populations ancestrales, l'une apparentée à une population africaine moderne, l'autre à une asiatique.

Les études précédentes de ce système reposent sur des modèles bayésiens d'attribution de SNP à des populations, pour des génomes entiers ou un sous-ensemble de gènes via une approche MLST.

J'ai pris un point de vue phylogénétique sur cette question, en utilisant un jeu de données construit dans l'équipe. Ce jeu de données est constitué de la phylogénie de 120 souches, comprenant la souche ancienne séquencée chez Ötzi, une momie trouvée dans les Alpes et datée à plus de 5 000 ans, et de 1 034 arbres de gènes. Nous avons appliqué la réconciliation aux arbres de gènes et aux arbres de population pour mieux comprendre les origines des gènes de la population européenne. Cette nouvelle approche, qui repose sur l'appariement uniforme de certaines feuilles des arbres de gènes (ici les européennes) à toutes les feuilles de l'arbre du dessus puis sur la probabilité a posteriori d'appariement obtenue en échantillonnant des scénarios, pourrait être facilement transposée à d'autres problèmes. Nous avons également utilisé notre approche de réconciliation à 3 niveaux pour comparer différents arbres de population.