

Proposition de sujet de thèse 2025

Laboratoire de mathématiques de Besançon
Université Marie et Louis Pasteur

Intitulé : Limites de graphes aléatoires pour la modélisation de l'évolution réticulée.

Encadrants :

Bruno SAUSSEREAU
François BIENVENU
Jean-Jil DUCHAMPS

bruno.saussereau [at] univ-fcomte.fr
francois.bienvenu [at] univ-fcomte.fr
jean-jil.duchamps [at] univ-fcomte.fr

Contexte. Depuis l'acceptation de la théorie de l'Évolution au sein de la communauté scientifique et l'avènement de la cladistique, l'histoire évolutive du vivant est représentée au moyen d'arbres appelés phylogénies. L'étude mathématiques des arbres joue donc un rôle central en biologie évolutive — aussi bien dans le développement de la théorie que dans celui de méthodes concrètes d'analyse de données. En retour, la modélisation des arbres phylogénétiques s'est révélée une formidable source d'inspiration pour les mathématiques.

Toutefois, les limitations inhérentes à l'utilisation d'arbres pour décrire les phylogénies sont de plus en plus apparentes. En particulier, la « révolution génomique » a montré que des phénomènes comme le transfert horizontal de gènes et l'hybridation — phénomènes connus depuis longtemps, mais que l'on pensait anecdotiques — sont omniprésents et jouent un rôle majeur en macroévolution. Par exemple, notre propre espèce, *Homo sapiens*, s'est récemment hybridée avec *Homo neanderthalensis*.

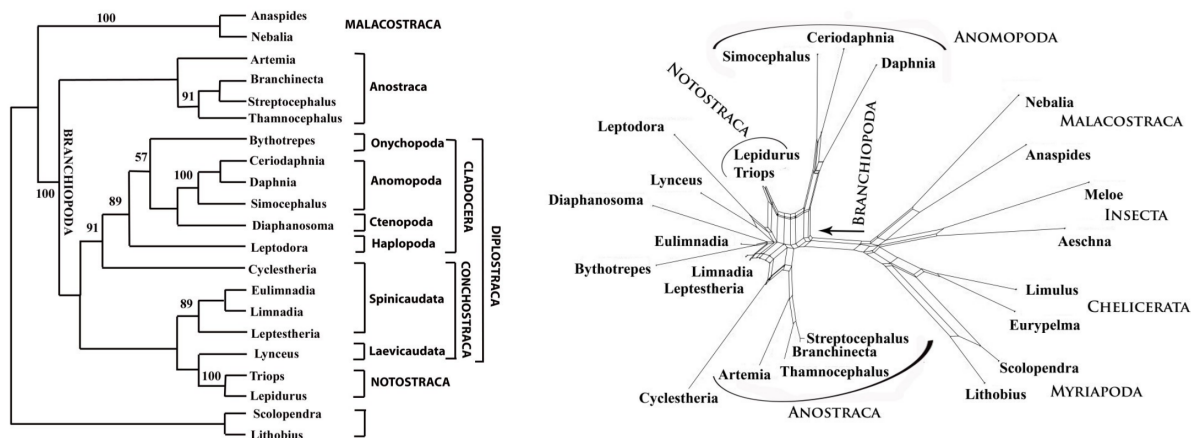


FIGURE 1 – À gauche, une phylogénie des brachiopodes basée sur le gène de l'ARN ribosomique 18S, représentée sous forme d'arbre. À droite, un exemple de réseau phylogénétique reconstitué à partir des mêmes données (algorithme Neighbor-Net). Ce réseau phylogénétique contient des informations sur l'histoire évolutive des brachiopodes qui sont perdues avec la représentation sous forme d'arbre. Figures tirées de [Wägele and Mayer \[6\]](#).

Le transfert horizontal de gènes et l'hybridation sont deux exemples d'évolution réticulée — terme générique regroupant les situations où le génome d'une espèce contient des éléments hérités de plusieurs espèces ancestrales distinctes. L'évolution réticulée ne pouvant être représentée par des arbres, les appels à privilégier l'utilisation de réseaux pour décrire les phylogénies se sont multipliés au cours des dix dernières années, conduisant à l'essor d'un nouveau domaine de recherche : l'étude des réseaux phylogénétiques (cf. Figure 1). Ce domaine se compose en particulier d'une communauté très active de

bioinformaticien·nes et de combinatoricien·nes qui étudient les propriétés algorithmiques et combinatoires de diverses classes de graphes dirigés acycliques.

Malgré cela, l'utilisation de réseaux phylogénétiques reste à ce jour marginale en biologie évolutive. Ceci s'explique par un manque d'outils mathématiques et informatiques adaptés, avec en particulier un manque (1) de mesures quantitatives permettant de décrire la structure des réseaux phylogénétiques et (2) de modèles de réseaux phylogénétiques aléatoires. Ces deux éléments sont pourtant indispensables à de multiples points de vue : développement de tests statistiques rigoureux ; test et calibrage des méthodes de reconstitution de phylogénies ; développement de modèles théoriques permettant d'étudier les éléments façonnant la structure et la géométrie des phylogénies.

Objectifs. Ce projet de thèse a pour but de contribuer à combler le manque d'outils mathématiques nécessaires à l'étude de l'évolution réticulée, en développant et en étudiant :

- des mesures quantitatives sur les réseaux (mesure d'équilibre et de diversité phylogénétique, notamment) ;
- des modèles de réseaux phylogénétiques aléatoires à la fois pertinents pour la biologie évolutive et intéressants du point de vue mathématique.

Ceci passera par l'exploitation d'outils probabilistes modernes issus du domaine de la géométrie aléatoire (limites locales [5] et limites de Gromov–Hausdorff [4], notamment). Nous étudierons en particulier les graphes aléatoires qui apparaissent comme généalogies de modèles de particules en interaction, dont l'évolution stochastique est constituée d'événements de branchement et d'événements de coalescence. La Figure 2 montre un exemple d'un tel processus dont nous avons étudié la généalogie dans un cas particulier [1], et qui servira de base à cette thèse. Pour l'étude de ces processus, les techniques de couplages avec des objets plus classiques (e.g. des processus de branchement : arbres de Galton–Watson et Crump–Mode–Jagers [3]) seront primordiales.

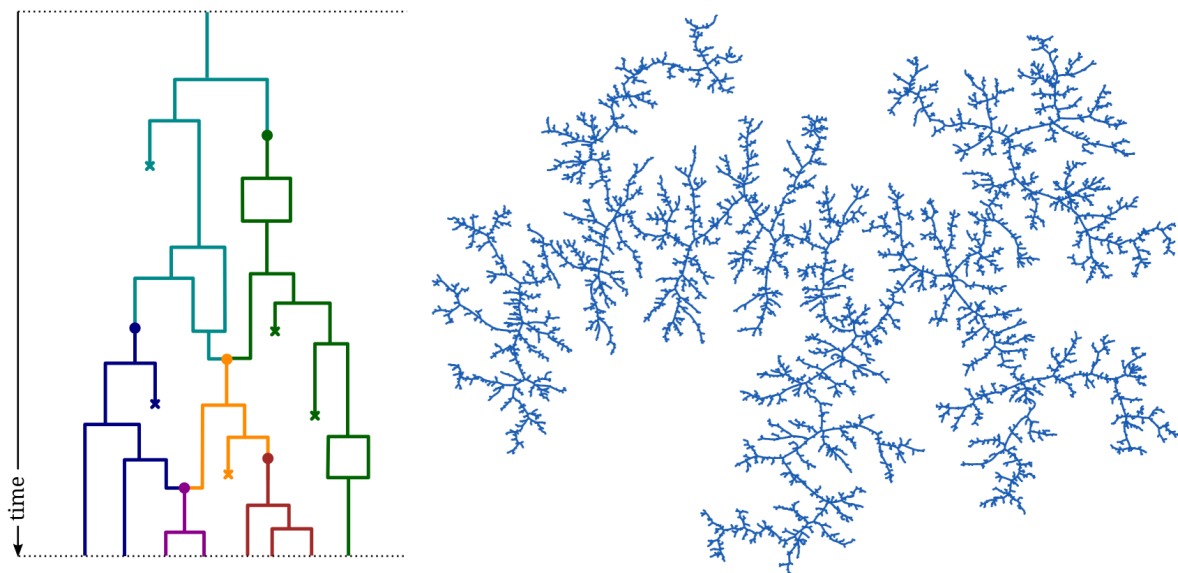


FIGURE 2 – À gauche, la généalogie d'un modèle de population dans lequel chaque individu se reproduit à taux 1, meurt à taux α et mute (change de couleur) à taux μ ; enfin, chaque paire d'individus coalesce à un taux β_k dépendant de k , une mesure de distance génétique (nombre de mutations nécessaires pour passer d'une couleur à l'autre). À droite, l'arbre brownien, espace métrique aléatoire identifié comme limite d'échelle de ce modèle dans le cas $(\beta_0, \beta_1, \beta_2 \dots) = (\beta, 0, 0, \dots)$.

Dans un premier temps, le projet consistera à identifier, parmi les mesures de structure et les modèles de réseaux aléatoires existants, ceux pour lesquels des méthodes récemment développées dans [2] peuvent être adaptées. Dans une seconde phase plus exploratoire, de nouvelles mesures de structure seront développées, en se concentrant particulièrement sur les deux points suivants : mesures d'équilibre phylogénétique (“*balance indices*”) et mesures de diversité phylogénétique. Enfin, la troisième phase du projet tournera vers la conception et l'étude de modèles de réseaux phylogénétiques aléatoires qui soient pertinents pour la biologie tout en se prêtant à l'analyse mathématique.

Ce projet de thèse bénéficiera de l'appui d'un postdoc recruté en 2025 dans le cas de l'ANR “GARP” portée par JJ. Duchamps, et sera réalisé en étroite collaboration avec l'équipe de M. Fuchs à Taïwan. En parallèle, une collaboration pourra être menée avec l'équipe de C. Scornavacca à l'université de Montpellier. Cette collaboration aura pour but d'implémenter certains des outils développés dans des bibliothèques informatiques destinés aux biologistes. Pour cette étape, nous aurons principalement un rôle consultatif, appuyant les bioinformaticien·nes dans la compréhension des subtilités mathématiques des modèles et dans le développement d'algorithmes efficaces.

Références

- [1] F. Bienvenu and J.-J. Duchamps. A branching process with coalescence to model random phylogenetic networks. *Electronic Journal of Probability*, 29:1–48, 2024. DOI: [10.1214/24-EJP1088](https://doi.org/10.1214/24-EJP1088).
- [2] F. Bienvenu, J.-J. Duchamps, M. Fuchs, and T.-C. Yu. The B_2 index of galled trees. *preprint*, ([arXiv:2407.19454](https://arxiv.org/abs/2407.19454)), 2025+.
- [3] P. Jagers. *Branching Processes with Biological Applications*. Wiley, London ; New York, 1975. ISBN 978-0-471-43652-2.
- [4] G. Miermont. Tessellations of random maps of arbitrary genus. *Annales scientifiques de l'École normale supérieure*, 42(5):725–781, 2009. DOI: [10.24033/asens.2108](https://doi.org/10.24033/asens.2108).
- [5] R. van der Hofstad. *Random Graphs and Complex Networks: Volume 2*. Cambridge Series in Statistical and Probabilistic Mathematics Series. Cambridge University Press, New York, 1st edition, 2024. ISBN 978-1-107-17400-9.
- [6] J. W. Wägele and C. Mayer. Visualizing differences in phylogenetic information content of alignments and distinction of three classes of long-branch effects. *BMC Evolutionary Biology*, 7(1):147, 2007. DOI: [10.1186/1471-2148-7-147](https://doi.org/10.1186/1471-2148-7-147).