



Plön, le 30 octobre 2020

Rapport d'évaluation du mémoire de thèse présenté par Thibault Latrille "Modélisation de l'articulation des mécanismes sélectifs et neutres dans l'évolution des séquences d'ADN codant pour des protéines", par Julien Y. Dutheil.

Dans ce mémoire de thèse, M. Thibault Latrille présente ses travaux ayant pour but de comprendre les mécanismes de l'évolution des séquences encodant des protéines. Cette problématique s'inscrit dans le champ disciplinaire dit de l'« évolution moléculaire », dont Thibault Latrille présente un état de l'art en introduction. Le mémoire de thèse est organisé autour de trois chapitres principaux, rédigés sous la forme d'articles scientifiques. Ces articles présentent toutes les qualités requises afin d'être soumis à des revues internationales. Les travaux comprennent des approches purement théoriques (chapitre 9), des développements de modèles de statistique inférentielle (chapitres 7 et 8), des études de simulations (chapitres 7, 8, 9) et des analyses de données (principalement chapitre 8), démontrant la très large expertise acquise par Thibault Latrille durant la préparation de son doctorat.

L'introduction de ce mémoire est organisée en cinq chapitres, couvrant l'essentiel des concepts abordés dans la thèse. Les deux premiers chapitres récapitulent les fondamentaux de la discipline, dissociant les aspects historiques et mathématiques. Il en résulte une introduction très didactique, où les concepts sont clairement énoncés avant d'être formalisés. La notion d'équilibre mutation-sélection fait l'objet d'un traitement approfondi, ce principe étant au cœur des modèles développés dans ce mémoire. Les chapitres 3 et 4 présentent des points plus techniques et plus directement liés aux travaux de cette thèse, à savoir les modèles d'évolution des séquences encodant des protéines, ainsi que les procédures d'inférence statistique, notamment Bayésienne. Ces chapitres couvrent l'essentiel des concepts les plus importants et sont exposés de manière simple et concise, mais sans pour autant masquer la complexité de leur sujet. Finalement, le chapitre 5 présente les grandes lignes de la biophysique des protéines, du moins les aspects pertinents pour ce travail. Le dernier chapitre de cette introduction est dédié à la présentation des objectifs de la thèse, qui sont énoncés clairement et articulés logiquement.

La seconde partie du mémoire comprend les trois chapitres présentant les résultats obtenus. Chaque chapitre est rédigé au format d'une publication scientifique, avec une introduction dédiée et les résultats séparés de la présentation des outils méthodologiques. L'organisation de ces résultats est logique, et illustre une démarche scientifique rigoureuse et de très grande qualité. Le premier manuscrit (chapitre 7) démontre les limites des modèles actuels ; en particulier leur inaptitude à

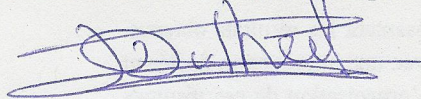
prendre en compte les biais mutationnels. A l'aide d'une étude de simulation, où des données sont générées suivant un processus évolutif biologique réaliste, Thibault Latrille démontre que les processus sélectifs et les biais mutationnels sont des facteurs confondants que les modèles simples n'arrivent pas à démêler. Il propose alors un nouveau type de modèle, plus complexe mais dont les nouveaux (nombreux) paramètres sont justifiés par des arguments théoriques. Ce nouveau modèle s'avère capable de remédier aux problèmes mis en évidence et offre donc une solution très prometteuse ainsi que des applications potentielles nombreuses.

Le second manuscrit (chapitre 8) présente une approche très ambitieuse, dont le but est d'estimer les tailles efficaces des populations ancestrales pour un échantillon d'espèces dont la phylogénie est considérée comme connue. La méthode proposée n'incorporant pas de données populationnelles, les tailles efficaces estimées représentent des moyennes sur le long terme, du moins sur les temps séparant les différents événements de spéciation représentés par l'arbre phylogénétique. Ces estimations sont rendues possibles par la modélisation orthogonale de deux composantes : la taille efficace, d'une part, qui varie entre branches, et un profil de valeurs sélectives, d'autre part, qui varie entre sites. Le modèle est implémenté au sein d'une procédure d'inférence Bayésienne et démontre de bonnes propriétés lorsqu'appliqué sur des données simulées. L'analyse de données réelles (mammifères placentaires dont primates, crustacés isopodes et insectes drosophiles) a révélé un résultat inattendu : une relative faible variance des tailles efficaces entre espèces. Thibault Latrille évoque plusieurs hypothèses permettant d'expliquer cette inconsistance avec les résultats de la littérature, notamment l'absence de prise en compte des relations épistasiques et une mis-spécification du paysage de valeur sélective. Ce travail démontre, entre autres, que les séquences protéiques contiennent d'importantes signatures de l'histoire des espèces, qui peuvent être déchiffrées par l'analyse comparative et des modèles biologiquement réalistes.

Le troisième manuscrit (chapitre 9) propose une approche théorique visant à évaluer l'impact de la taille efficace des populations sur le ratio de taux de substitutions non-synonymes et synonymes ($dN/dS = \omega$). Le ratio ω est une mesure de l'intensité (et direction) de la sélection agissant sur les séquences encodant des protéines. Les propriétés théoriques de cette mesure, bien que très fréquemment utilisée, sont relativement peu établies. Thibault Latrille dérive en premier lieu des propriétés d'un modèle simplifié, mais pouvant être traité analytiquement. Il vérifie ensuite empiriquement la robustesse des conclusions obtenues, en simulant des données violant l'une ou l'autre des hypothèses du modèle. Les résultats de Thibault Latrille vont dans le sens des conclusions du chapitre 8, suggérant que la stabilité conformationnelle d'une protéine représente une mesure incomplète de sa valeur sélective, et par conséquent, que les modèles se basant uniquement sur cet aspect ne permettent pas de rendre compte des variations de taille efficace des populations réelles. Les trois manuscrits incluent une discussion très approfondie, à la fois technique (limites des modèles, hypothèses sous-jacentes), et biologique (implications des résultats reportés).

Le chapitre 10 conclut ce mémoire de thèse en faisant la synthèse des résultats obtenus. Fort de l'expérience acquise lors de ses travaux de recherche, Thibault Latrille propose des directions pertinentes pour de futurs développements. Thibault Latrille a conduit une recherche par essence interdisciplinaire, s'inscrivant dans des problématiques évolutives, mais empruntant très largement à la biophysique des protéines. Ces travaux d'un excellent niveau démontrent qu'il est possible de maîtriser des outils mathématiques, statistiques et informatiques. Thibault Latrille argumente plus largement sur la démarche scientifique qu'il a adoptée, plaidant pour une approche mécanistique lors du développement de modèles biologiques. Ses travaux de thèse constituent assurément une démonstration convaincante de la pertinence d'une telle approche.

En conclusion, **je suis donc très favorable à la soutenance par M. Thibault Latrille de son travail de thèse en vue de l'obtention du diplôme de doctorat.**



Julien Yann Dutheil,
Group Leader au Max Planck Institute for Evolutionary Biology, Plön, Allemagne.