



UNIL | Université de Lausanne

Faculté de biologie
et de médecine

Soutenance de thèse

Romain Feron

Master in Bioinformatics
Université de Rennes I, France

Soutiendra en vue de l'obtention du grade de
Doctorat ès sciences de la vie (PhD)
de l'Université de Lausanne

sa thèse intitulée :

**Investigating relationships between
sequence conservation and function using
multispecies whole genome alignments**

Directeur·trice de thèse
Prof. Marc Robinson-Rechavi

Codirecteur·trice de thèse
Dr Robert Waterhouse

Cette soutenance aura lieu

**Lundi 1 juillet 2024
à 16h00**

Amphithéâtre, Biophore, quartier UNIL-Sorge, 1015 Lausanne

L'entrée est publique

Prof. Niko GELDNER
Directeur de l'École Doctorale

17.06.24

Les travaux présentés dans cette thèse s'intéressent au lien entre la conservation évolutive des séquences génétiques et la fonction biologique chez les arthropodes, un groupe qui regroupe les insectes, arachnides, crustacés, et certains autres invertébrés. En s'appuyant sur les progrès remarquables dans le domaine du séquençage de l'ADN et les initiatives de séquençage à grande échelle, elle vise à éclairer les schémas évolutifs des génomes d'arthropodes et à caractériser de nouveaux éléments génomiques fonctionnels grâce à des approches de génomique comparative multi-espèces.

Le principe sous-jacent à cette approche est fascinant : les séquences génétiques qui persistent à travers de nombreuses espèces, souvent sur des millions d'années d'évolution, le font probablement en raison de pressions évolutives exercées par leur fonction biologique. Par conséquent, une meilleure compréhension de cette relation entre la conservation des séquences et la fonction biologique associée à ces séquences pourrait grandement faciliter l'identification à grande échelle de nouveaux éléments fonctionnels dans les génomes, et enrichir considérablement notre compréhension de la manière dont les génomes codent pour des fonctions biologiques.

Cependant, malgré les succès initiaux d'approches similaires, la génération des données secondaires nécessaires aux analyses répondant à ces questions - les *Multi-species Whole Genome Alignments* (MWGA) - reste un défi important. Avant le début de ce projet de thèse, il n'existait pas d'implémentation fiable des outils bioinformatiques nécessaires pour générer ces alignements. De plus, alors que les progrès technologiques ont permis une accumulation de séquences génomiques à un rythme sans précédent, la qualité de ces données et leur représentativité taxonomique demeurent inégales et difficiles à explorer.

Pour surmonter ces obstacles, cette thèse a adopté une approche méthodique. Tout d'abord, elle a entrepris d'évaluer précisément la qualité et la distribution taxonomique des ressources génomiques disponibles pour les arthropodes. Ce travail a abouti à la création d'un outil accessible en ligne, l'Arthropoda Assembly Assessment Catalogue (A³Cat), qui constitue une ressource inestimable pour la communauté scientifique en regroupant des informations sur les génomes d'arthropodes et en fournissant des estimations de leur qualité.

Ensuite, cette thèse s'est attelée à développer des outils bioinformatiques pour la création et l'analyse de MWGA pour les arthropodes. Ces outils ont été conçus pour répondre aux normes modernes de la science informatique en matière de portabilité, de scalabilité et de reproductibilité des analyses. Le développement de ces outils a nécessité des efforts considérables, mais leur création représente une avancée majeure dans le domaine de la génomique comparative.

Finalement, cette thèse a réalisé des analyses exploratoires à l'aide des MWGA générés, en se concentrant notamment sur l'étude de la conservation des séquences génomiques chez les moustiques, qui sont d'importants vecteurs de maladies humaines. Ces analyses ont permis d'explorer comment la conservation des séquences est liée à la fonction biologique des gènes, en mettant particulièrement en lumière le rôle des gènes impliqués dans le système immunitaire des moustiques.

En conclusion, cette thèse représente une contribution technique significative à la recherche future en génomique comparative au travers d'outils qui avanceront notre compréhension de la relation entre la conservation des séquences génétiques et la fonction biologique. Elle ouvre également de nouvelles perspectives pour l'étude de l'évolution des génomes et de la biologie des arthropodes.