

DIPLÔME NATIONAL DE DOCTORAT

(Arrêté du 25 mai 2016)

Date de la soutenance : **16 décembre 2024**

Nom de famille et prénom de l'auteur. e : **Madame Mélodie BASTIAN**

Titre de la thèse : Génomique évolutive, de la phylogénie à la génétique des populations.

Résumé



L'étude de l'évolution des génomes s'exerce à différentes échelles de temps évolutifs. Une première échelle, dite micro-évolutive (du domaine de la génétique des populations), se concentre sur les changements génétiques entre individus d'une même population. Si on étend cette échelle sur un temps plus long, on aboutit à une deuxième échelle temporelle dite macro-évolutive (du domaine de la phylogénie), basée sur l'étude des changements génétiques entre individus d'espèces différentes. Les génomes, notamment les séquences codantes, évoluent sous l'action combinée de plusieurs processus tels que la mutation, la sélection naturelle et la dérive génétique. Cette dernière, inversement proportionnelle à la taille efficace d'une population (N_e), est un processus évolutif stochastique particulièrement important, car il limite l'efficacité de la sélection naturelle. Dans un contexte où la majorité des mutations sont délétères, avec des effets allant de létaux à faibles, la sélection naturelle purifiante va éliminer une mutation délétère d'autant plus vite qu'elle est impactante. Selon la théorie quasi-neutre formulée par Ohta, plus N_e est petit et plus la dérive génétique conduit rapidement à la fixation aléatoire de mutations délétères, empêchant la sélection de purifier les mutations à faible effet. Ainsi, l'efficacité de la sélection dans une population serait proportionnelle à N_e . Cependant, cette théorie manque encore de preuves empiriques décisives sous la forme d'une étude complète qui articule de manière cohérente les échelles micro et macro-évolutives sur un jeu de données suffisamment riche à la fois en gènes et en espèces. Lors de mon premier travail de thèse, j'ai constitué un jeu de données contenant environ 150 génomes de mammifères. J'ai annoté chacun de ces génomes afin de construire un alignement de plus de 6000 gènes orthologues, globalement partagés à l'échelle du clade. Par ailleurs, puisque les individus séquencés sont diploïdes, j'ai estimé leur hétérozygotie sur ces mêmes gènes afin d'obtenir une estimation du polymorphisme de leurs populations. Dans mon deuxième travail de thèse, j'ai étudié la validité de l'approximation du polymorphisme par l'hétérozygotie d'un unique individu et j'ai

notamment montré que cette approche est valable, bien qu'elle introduise une variance supplémentaire dans l'estimation du polymorphisme. En associant les alignements de gènes interspécifiques à l'hétérozygotie intra-individu, j'ai pu estimer l'intensité de la sélection à la fois aux échelles macro-évolutive (dN/dS) et micro-évolutive (pN/pS). J'ai également intégré des estimateurs de la taille efficace (traits d'histoire de vie et pS) pour chaque échelle. Le jeu de données a été analysé par une méthode intégrative bayésienne (FastCoevol) permettant d'étudier conjointement les relations entre chaque trait étudié, le long de la phylogénie. Une attention particulière a été portée sur la qualité des données et sur la reproductibilité de l'analyse. Mon travail de thèse a permis de confirmer la corrélation positive entre dN/dS et traits d'histoire de vie, déjà connue par ailleurs, et a également permis de montrer une corrélation négative entre pN/pS et pS , un résultat inédit dans le cas des génomes nucléaires de mammifères. Prises ensemble, ces deux observations valident la relation entre intensité de la sélection et N_e séparément pour les échelles de temps micro- et macro-évolutives, chez les mammifères. Aussi, le croisement des échelles montre, bien que de façon plus faible, une concordance entre micro- et macro-évolution sur les attendus quasi-neutres. L'ensemble de ces résultats fournissent une validation empirique de la théorie quasi-neutre à l'échelle des mammifères et permettent d'envisager une reconstruction quantitative des variations de N_e aux différentes échelles ainsi que de mesurer les contributions respectives des variations de court et long terme aux changements de N_e .

Mots-clés : Génomique, Phylogénie, Génétique des populations, Évolution, Dérive génétique