

STAGE DE RECHERCHE M2 ECOLOGIE EVOLUTION GENOMIQUE

Rentrée 2021

Evolution moléculaire de protéines de la reproduction chez les Mammifères

Encadrants :

Pascale Chevret, Equipe Ecologie et Evolution des Populations (Pascale.Chevret@univ-lyon1.fr)

Nicolas Lartillot, Equipe Bioinformatique, Phylogénie et Génomique Evolutive (Nicolas.Lartillot@univ-lyon1.fr) Laboratoire de Biométrie et Biologie Évolutive, UMR CNRS 5558, Université Lyon 1

Les gènes impliqués dans la reproduction sont connus pour présenter des taux d'évolution plus rapides que ceux exprimés dans d'autres organes (Meslin et al., 2012 ; Wilburn and Swanson, 2016). Cette évolution rapide est probablement causée par la compétition spermatique et la sélection sexuelle, qui résulte en une 'course aux armements', au niveau génétique et moléculaire, autour de la biologie de la reproduction. Ces gènes offrent de ce fait une opportunité d'investiguer en détail les modalités de la sélection positive au niveau moléculaire, dans un cas où la biologie est par ailleurs bien connue. L'objectif du stage est de conduire une analyse détaillée en ce sens, en recrutant les développements récents en modélisation de l'évolution des familles multigéniques et de la sélection à l'échelle phylogénétique. Dans un premier temps, le travail se focalisera sur les protéines de la zone pellucide, pour être ensuite élargi à d'autres protéines impliquées dans l'interaction ovocyte-spermatozoïde.

La zone pellucide (ZP) est une matrice enveloppant les oocytes de mammifères. Elle participe aux interactions spermatozoïdes-oocyte pendant les premières étapes de la fertilisation et joue un rôle important dans la mise en place de barrières interspécifiques. Dans une analyse phylogénomique, Goudet et al. (2008) ont montré que si de nombreux mammifères possédaient 4 protéines (ZP1, ZP2, ZP3 et ZP4), certains avaient perdu la ZP1 ou la ZP4. Ils ont également proposé que l'évolution des ZP et la composition de la zone pellucide soient le résultat de plusieurs événements de duplication et de disparition de gènes (Goudet et al., 2008). D'autres études ont également mis en évidence de la sélection positive et de l'évolution adaptative chez une ou deux de ces glycoprotéines comme ZP2 et/ou ZP3 chez les rongeurs (Turner and Hoekstra, 2006; Swann et al., 2017). Toutefois, peu d'études détaillées de l'évolution moléculaire et de la possible co évolution de ces protéines de façon plus large chez les mammifères ont été réalisées à ce jour.

Dans le cadre d'une collaboration entre Pascale Chevret et Manuel Aviles (Murcia, Espagne), différentes études ont mis en évidence l'existence de pseudogènes pour ZP1 (Moros-Nicolás et al., 2018) et ZP4 (Izquierdo-Rico et al., 2021) chez différentes espèces de mammifères. Une histoire plus complexe a également été décrite chez les marsupiaux qui possèdent jusqu'à 7 protéines et les monotrèmes jusqu'à 8 (Moros-Nicolás et al., 2021). Ces études ont été réalisées avec des approches de séquençage (ADN et ARN), de protéomique et de bioinformatique combinant des données originales et des données disponibles dans les banques. Les analyses bioinformatiques étaient restées assez simples : reconstructions phylogénétiques et recherche de sélection positive en utilisant des modèles phylogénétiques à codons standards (logiciel PaML, (Yang, 2007)).

Un article publié récemment augmente de manière très importante le nombre de génomes de mammifères disponibles, de 110 à 240 (Genereux et al., 2020). Par ailleurs, de nouveaux modèles à codons ont été développés ces dernières années (Pond et al., 2011; Murrell et al., 2015; Rodrigue and Lartillot, 2017; Rodrigue et al., 2021). Ces modèles visent à caractériser de façon plus fine les modalités de la sélection (purificatrice et positive) agissant sur les séquences codantes.

Ces nouvelles données et ces nouveaux modèles offrent une opportunité pour réaliser des analyses plus complètes de l'évolution moléculaire des protéines ZP chez les mammifères et répondre ainsi à des questions importantes : comment cette famille multi-génique s'est diversifiée à travers les mammifères ? Y-a-t-il eu des duplications et des pseudogénisations répétées ? Est-ce que cette diversification est causée par la compétition spermatique ou plus généralement la sélection sexuelle ? Comment s'articule cette dynamique évolutive au niveau de la famille multi-génique avec la sélection à une échelle moléculaire plus fine (substitutions ponctuelles dans les séquences en acides aminées) ? Par ailleurs, le stage sera aussi l'occasion d'étudier d'autres gènes candidats impliqués dans les interactions entre gamètes, comme par exemple les gènes IZUMO1 et FIMP pour le spermatozoïde et JUNO et CD9 pour l'ovocyte. L'étude de ces gènes permettra d'élargir le questionnement autour des problématiques d'évolution moléculaire liée à la sélection sexuelle.

Le stage de master sera co-encadré par Pascale Chevret et Nicolas Lartillot. Pascale Chevret apportera sa connaissance de la phylogénie des mammifères et des gènes d'intérêt (avec des jeux de données déjà disponibles). Nicolas Lartillot contribuera une expertise sur la gamme de méthodes et de modèles disponibles, à la fois pour inférer les phylogénies et caractériser les modalités sélectives agissant sur les séquences codantes. Sur ce dernier point, il a contribué au développement de nouvelles approches de modèles à codons (Rodrigue and Lartillot, 2017; Parto and Lartillot, 2017; Rodrigue et al., 2021) qui seront testées et comparées aux alternatives existantes (Yang, 2007; Pond et al., 2011; Murrell et al., 2015) dans le contexte de l'évolution de protéines impliqués dans la reproduction.

Dans le cadre de ce stage, l'étudiant.e recherchera les séquences des gènes d'intérêt dans les différents génomes de mammifères disponibles après en avoir évalué la qualité. Cette première étape permettra de compléter les jeux de données existants. Les alignements obtenus seront ensuite analysés afin 1) de mettre en évidence et dater les principaux événements de duplication et pseudogénisation le long de la phylogénie des Mammifères 2) d'étudier les patrons de sélection à l'œuvre sur les différents gènes en utilisant toute la gamme de modèles actuellement disponibles.

Ce stage permettra à l'étudiant.e d'acquérir des bases solides en bioinformatique, et plus particulièrement autour de l'analyse des séquences codantes, dans le contexte d'une question biologique de grande importance: mieux comprendre, au niveau moléculaire, les mécanismes de sélection au niveau génomique opérant autour de la reproduction. Pour effectuer ce stage, une connaissance préalable de la programmation ou de la bioinformatique n'est pas requise. Toutefois, il est souhaitable d'avoir suivi les cours de phylogénomique et de statistique bayésienne durant le premier semestre du M2.

Références

- Genereux DP, Serres A, Armstrong J, Johnson J, Marinescu VD, Murén E, et al. (2020). A comparative genomics multitool for scientific discovery and conservation. *Nature* 587: 240–245.
- Goudet G, Mugnier S, Callebaut I, Monget P (2008). Phylogenetic analysis and identification of pseudogenes reveal a progressive loss of zona pellucida genes during evolution of vertebrates. *Biol Reprod* 78: 796–806.

Izquierdo-Rico MJ, Moros-Nicolás C, Pérez-Crespo M, Laguna-Barraza R, Gutiérrez-Adán A, Veyrunes F, et al. (2021). ZP4 Is Present in Murine Zona Pellucida and Is Not Responsible for the Specific Gamete Interaction. *Front Cell Dev Biol* 8.

Meslin C, Mugnier S, Callebaut I, Laurin M, Pascal G, Poupon A, et al. (2012). Evolution of genes involved in gamete interaction: evidence for positive selection, duplications and losses in vertebrates. *PLoS One* 7: e44548.

Moros-Nicolás C, Chevret P, Jiménez-Movilla M, Algarra B, Cots-Rodríguez P, González-Brusi L, et al. (2021). New Insights into the Mammalian Egg Zona Pellucida. *Int J Mol Sci* 22.

Moros-Nicolás C, Leza A, Chevret P, Guillén-Martínez A, González-Brusi L, Boué F, et al. (2018). Analysis of ZP1 gene reveals differences in zona pellucida composition in carnivores. *Reprod Fertil Dev* 30: 272–285.

Murrell B, Weaver S, Smith MD, Wertheim JO, Murrell S, Aylward A, et al. (2015). Gene-Wide Identification of Episodic Selection. *Mol Biol Evol* 32: 1365–1371.

Parto S, Lartillot N (2017). Detecting consistent patterns of directional adaptation using differential selection codon models. *BMC Evol Biol* 17: 147.

Pond SLK, Murrell B, Fourment M, Frost SDW, Delport W, Scheffler K (2011). A random effects branch-site model for detecting episodic diversifying selection. *Mol Biol Evol* 28: 3033–43.

Rodrigue N, Lartillot N (2017). Detecting Adaptation in Protein-Coding Genes Using a Bayesian Site-Heterogeneous Mutation-Selection Codon Substitution Model. *Mol Biol Evol* 34: 204–214.

Rodrigue N, Latrille T, Lartillot N (2021). A Bayesian Mutation–Selection Framework for Detecting Site-Specific Adaptive Evolution in Protein-Coding Genes (T Pupko, Ed.). *Mol Biol Evol* 38: 1199–1208.

Swann CA, Cooper SJB, Breed WG (2017). The egg coat zona pellucida 3 glycoprotein – evolution of its putative sperm-binding region in Old World murine rodents (Rodentia: Muridae). *Reprod Fertil Dev* 29: 2376.

Turner LM, Hoekstra HE (2006). Adaptive evolution of fertilization proteins within a genus: variation in ZP2 and ZP3 in deer mice (*Peromyscus*). *Mol Biol Evol* 23: 1656–69.

Wilburn DB, Swanson WJ (2016). From molecules to mating: Rapid evolution and biochemical studies of reproductive proteins. *J Proteomics* 135: 12–25.

Yang Z (2007). PAML 4: phylogenetic analysis by maximum likelihood. *Mol Biol Evol* 24: 1586–91