

## Sujet M2 2019-2020

### **Intitulé du sujet : Caractérisation des points chauds de recombinaison dans les génomes animaux.**

(english version below)

**Encadrants:** Laurent DURET ([laurent.duret@univ-lyon1.fr](mailto:laurent.duret@univ-lyon1.fr)), Nicolas LARTILLOT

([nicolas.lartillot@univ-lyon1.fr](mailto:nicolas.lartillot@univ-lyon1.fr)) LBBE - UMR 5558.

**Lieu du stage:** Laboratoire de Biométrie et Biologie Evolutive (LBBE), 43 boulevard du 11 novembre 1918, Villeurbanne

#### **Résumé du projet:**

La recombinaison est un mécanisme fondamental de l'évolution des génomes et des traits; le brassage génétique qu'elle induit augmente notamment l'efficacité de la sélection naturelle multi-locus, affecte le niveau de polymorphisme, la composition en bases des génomes, et les patrons d'introgession [1-4]. La recombinaison est omniprésente chez les eucaryotes, les espèces qui en sont privées ayant un plus grand risque d'extinction. Pourtant, son évolution et sa régulation sont encore mal comprises.

La distribution des taux de recombinaison le long des génomes a longtemps été seulement caractérisée par des cartes génétiques à faible résolution. Récemment, le séquençage haut débit a permis l'obtention de cartes de recombinaison à haute résolution (via l'analyse du déséquilibre de liaison). Ces nouvelles approches ont permis de mettre en évidence, chez les mammifères, l'existence de points chauds de recombinaison, régulés par une protéine liant l'ADN - PRDM9 [6]. Le gène clé PRDM9 a une dynamique évolutive remarquable de type "Reine rouge", du fait de l'érosion mécanique de ses motifs due à un biais de transmission au niveau des points chauds [7]. Remarquablement, ce gène a été perdu chez les oiseaux et les canidés [9], bien qu'il intervienne dans un processus aussi fondamental que la méiose et soit indispensable chez la souris.

Ce projet de recherche vise à étendre notre connaissance des paysages génomiques de recombinaison et de leur déterminisme au-delà des mammifères et des amniotes. Nous souhaitons notamment déterminer si l'existence de points chauds instables est une spécificité des mammifères, ou une situation plus commune - et dans ce cas essayer de comprendre les raisons évolutives d'une dynamique aussi paradoxale. Pour cela nous allons dans un premier temps caractériser les variations du taux de recombinaison le long du génome de deux espèces de tuniciers (*Ciona intestinalis*, *C. robusta*), groupe proche des vertébrés, possédant un homologue de PRDM9. L'examen et la comparaison des cartes de recombinaison de ces deux espèces - position fine des points de cassure, existence de points chauds, niveau de stabilité entre populations, biais de transmission associés - permettront de confirmer/infirmier l'importance du gène majeur PRDM9 et l'origine évolutive de sa fonction. Ces analyses permettront également d'enrichir plus généralement notre compréhension de l'impact de la recombinaison sur l'évolution des génomes.

#### **Méthodes**

Des génomes de référence et données de re-séquençage populationnel sont déjà disponibles pour ces deux espèces cibles au travers de collaborations existantes. L'analyse des données impliquera de comprendre et prendre en main les méthodes et logiciels classiques du domaine (par exemple LDhat), tout en prenant en compte les limitations et biais inhérents aux données haut débit, telles que les erreurs

de séquençage, d'assemblage, et de génotypage. Sur le plan technique ce projet permettra une formation approfondie en bioinformatique, programmation, calcul intensif et analyse de données haut débit.

### Résultats attendus

- cartes génétiques populationnelles dans deux espèces de tuniciers
- analyse de l'évolution moléculaire de PRDM9 et identification de ses variants alléliques
- analyse du lien entre cartes de recombinaison et allèles PRDM9
- synthèse sur l'origine et l'évolution du déterminisme de la recombinaison chez les vertébrés/ animaux et ses conséquences sur l'évolution des génomes

### Profil et compétences recherchées

Le candidat idéal aura un intérêt pour la génomique des populations et l'évolution moléculaire, un souhait de se former à la bioinformatique et à l'analyse de données haut débit, et des capacités pour le travail d'équipe et la collaboration.

### Collaborations

Ce projet s'inscrit dans une collaboration de long terme entre Nicolas Galtier (ISEM Montpellier), Pierre-Alexandre Gagnaire (ISEM Montpellier), Nicolas Lartillot (LBBE, Lyon) et Laurent Duret (LBBE, Lyon).

---

## Characterization of recombination hotspots in animal genomes

### Project summary

Recombination is a fundamental mechanism contributing to genome and trait evolution. By dissipating linkage, recombination leads to more efficient natural selection and modulates genetic polymorphism, genome composition and introgression patterns. Recombination is ubiquitous in eukaryotes, and species devoid of recombination have a higher risk of extinction. Yet, the evolution and the regulation of recombination are still poorly understood.

The distribution of recombination rates across genomes has long been characterized using low-resolution genetic maps. Recently, high-throughput sequencing has opened the way to high-resolution recombination maps (through the analysis of linkage disequilibrium). In turn, these new approaches have led to the discovery, in mammals, of **recombination hot spots**, regulated by a DNA-binding protein, called **PRDM9**. This key regulatory gene has a remarkable Red Queen-like evolutionary dynamics, caused by the erosion of its binding motifs by a biased gene conversion. Strikingly, PRDM9 has been lost in birds and the dog, in spite of being indispensable in the mouse, owing to its implication in such a fundamental mechanism as meiosis.

The aim of this research project is to extend our knowledge of genome-wide recombination landscapes and of their regulation, beyond mammals and amniotes. Specifically, we aim to determine whether short-lived hot spots are specific to mammals or are a more common feature of recombination across animals -- in which case we would like to understand the meaning of this intriguing evolutionary dynamics. To this aim, the first step of the project will be to characterize the variation in recombination rate across the genomes of two species of tunicates (*Ciona intestinalis* and *C. robusta*), which are the closest relatives of vertebrates and have a PRDM9 homologue. The analysis and comparison of recombination maps in these two species (fine-scale distribution of recombination break points, existence and stability of hot spots, transmission biases) will make it possible to assess the importance of PRDM9 in this group and the evolutionary origins of its function. These analyses will also enrich our understanding of the impact of recombination on genomic landscape evolution.

## Methods

Reference genomes and population re-sequencing data are already available for these two species, thanks to existing collaborations. The analysis of these empirical data will offer an opportunity to learn to master the computational methods and the software programs classically used in this scientific domain (e.g. LDhat), while learning to appreciate the limitations and biases inherent to high-throughput data). This project will give a thorough training in bioinformatics, scripting, high-performance computing and high-throughput data analysis.

### Expected results:

- population-based genetic maps in two tunicate species
- analysis of the molecular evolution of PRDM9 and its allelic variants
- analysis of the link between PRDM9 alleles and the distribution of recombination
- a global synthesis on the origin and evolution of recombination landscapes in vertebrates / animals, and its consequence on genome evolution

### Expected profile and skills of the candidate

The candidate should have a particular interest in population genomics and molecular evolution; should be willing to acquire an extensive training in bioinformatics and high-throughput data analysis, and should lean toward team and collaborative work.

This work will be conducted in the context of a long-term collaboration between Nicolas Galtier (ISEM Montpellier), Pierre-Alexandre Gagnaire (ISEM Montpellier), Nicolas Lartillot (LBBE, Lyon) and Laurent Duret (LBBE, Lyon).

## Références bibliographiques

1. Duret L, Galtier N. Biased gene conversion and the evolution of mammalian genomic landscapes. *Annu Rev Genomics Hum Genet.* 2009;10:285-311.
2. Glémin S, Galtier N. Genome evolution in outcrossing versus selfing versus asexual species. *Methods Mol Biol.* 2012;855:311-35.
3. Tine M, Kuhl H, Gagnaire PA, Louro B, Desmarais E, Martins RS, Hecht J, Knaust F, Belkhir K, Klages S, Dieterich R, Stueber K, Piferrer F, Guinand B, Bierne N, Volckaert FA, Bargelloni L, Power DM, Bonhomme F, Canario AV, Reinhardt R. European sea bass genome and its variation provide insights into adaptation to euryhalinity and speciation. *Nat Commun.* 2014 Dec 23;5:5770.
4. Leitwein M, Guinand B, Pouzadoux J, Desmarais E, Berrebi P, Gagnaire PA. A Dense Brown Trout (*Salmo trutta*) Linkage Map Reveals Recent Chromosomal Rearrangements in the *Salmo* Genus and the Impact of Selection on Linked Neutral Diversity. *G3 (Bethesda).* 2017 Apr 3;7(4):1365-1376.
5. Myers S, Bottolo L, Freeman C, McVean G, Donnelly P. A fine-scale map of recombination rates and hotspots across the human genome. *Science.* 2005 Oct 14;310(5746):321-4.
6. Baudat F, Buard J, Grey C, Fledel-Alon A, Ober C, Przeworski M, Coop G, de Massy B. PRDM9 is a major determinant of meiotic recombination hotspots in humans and mice. *Science.* 2010 Feb 12;327(5967):836-40.
7. Lesecque Y, Glémin S, Lartillot N, Mouchiroud D, Duret L. The red queen model of recombination hotspots evolution in the light of archaic and modern human genomes. *PLoS Genet.* 2014 Nov 13;10(11):e1004790.

8. Baker Z, Schumer M, Haba Y, Bashkirova L, Holland C, Rosenthal GG, Przeworski M. Repeated losses of PRDM9-directed recombination despite the conservation of PRDM9 across vertebrates. *Elife*. 2017 Jun 6;6. pii: e24133.