MASTER BIOINFORMATIQUE

Master Internship offer M2

Internship supervisors and host laboratory

Supervisors: Carina Farah Mugal (CR CNRS, carina.mugal@univ-lyon1.fr), Laurent Duret (DR CNRS, laurent.duret@univ-lyon1.fr)

Host laboratory: Laboratoire de Biométrie et Biologie Evolutive (LBBE), 43 boulevard du 11 novembre 1918, Villeurbanne https://lbbe.univ-lyon1.fr/

Research project title

Identification of non-methylated islands (NMIs) in the genome of a passerine bird and assessment of their role in recombination rate regulation

Project description

The investigation of recombination rate regulation across divergent vertebrates has revealed that in most species **recombination events tend to cluster in narrow regions of the genome referred to as recombination hotspots**. For the localization of such recombination hotspots, two separate mechanisms have been identified. In species that have a functional zinc-finger protein, PRDM9, such as primates and rodents, recombination is typically targeted towards sequence motifs in the genome where PRDM9 binds (Baudat et al. 2010; Parvanov et al. 2010). In species that lack a functional PRDM9, such as canids and birds, CpG islands (CGIs) are suggested to function as primary targets of recombination hotspots (Berglund et al. 2014; Kawakami et al. 2017). **CGIs are a classical epigenetic feature of vertebrate genomes** that are characterized by an elevated frequency of CpG dinucleotides and GC content, low levels of DNA methylation, specific histone modifications, and an enrichment of transcription factor binding sites (Bird et al. 1985; Deaton and Bird 2011; Angeloni and Bogdanovic 2021). Often CGIs are associated with gene promoters (Ponger et al. 2001; Saxonov et al. 2006), which are associated with an open chromatin structure accessible to recombination events. However, not all CGIs are associated with recombination events and the molecular mechanism of CGIs in recombination rate regulation remains at present poorly understood.

Some of this confusion appears though to originate from a purely computational definition of CGIs solely based on their DNA sequence features, i.e. an elevated frequency of CpG dinucleotides and GC content (Gardiner-Garden and Frommer 1987; Saxonov et al. 2006; Cohen et al. 2011), **without validation by epigenetic marks**. Recent advances in sequencing technology and the accessibility of genome-wide DNA methylation data across multiple tissues and species has confirmed the lack of DNA methylation in only a fraction of computationally defined CGIs (Long et al. 2013; Al Adhami et al. 2022). Another interesting observation from these studies was that not all of the non-methylated islands (NMIs) of the genome are found near gene promoters or overlap with CGIs. More specifically, NMIs shared among tissues (‘canonical’ NMIs) generally show high CpG density, frequently overlap with gene promoters and tend to be evolutionary conserved among vertebrates (Long et al. 2013; Long et al. 2016; Al Adhami et al. 2022). Tissue-specific NMIs, in contrast, are frequently found in intergenic regions and show less evolutionary constraint (Long et al. 2013; Long et

al. 2016; Al Adhami et al. 2022). It is therefore possible that not CGIs *per se* but rather NMIs or specifically 'canonical' NMIs could be the molecular target of recombination events.

The internship project will focus on the **identification non-methylated islands (NMIs) in the collared flycatcher** (*Ficedula albicollis*), a representative study system of recombination regulation in birds (Kawakami et al. 2014; Kawakami et al. 2017). A high-quality genome assembly has recently been generated for the species (Chase et al. 2023). In addition, we have performed whole-genome bisulfite sequencing across multiple tissues and individuals of collared flycatcher, which enables us to compute genome-wide DNA methylation landscapes at the base-pair resolution (Boman et al. 2024). These data provide an excellent opportunity to detect NMIs in a bird genome with help of relevant bioinformatic tools and subsequently **investigate their overlap with CGIs and their association with recombination hotspots**. The internship project shall thus advance our understanding of the role of DNA methylation in fine-scale regulation of meiotic recombination in collared flycatcher. In the long run, the project can be continued as PhD project and focus on the characterization of fine-scale as well as broad-scale regulation of recombination rate across multiple bird species, to decipher the molecular mechanisms behind recombination rate evolution in species without PRDM9.

Specific tasks

- Read mapping of whole-genome bisulfite sequencing data onto the newly assembled chromosome-level collared flycatcher reference genome
- Identification of NMIs in the collared flycatcher genome
- Classification of NMIs based on their tissue-specificity and their overlap with CGIs
- Investigation of the role of NMIs in recombination rate regulation

As part of the project work, the student will develop skills in processing high-throughput re-sequencing data, develop a bioinformatics workflow, and get familiar with common genomic data formats (e.g. FASTQ, BAM, BED). The student will also gain experience with a broad variety of computational methodology in the field of bioinformatics, ranging from read mapping (e.g. Bowtie2), analysis of DNA methylation data (e.g. BSmooth) to statistical data analysis with R. Besides, the student will advance their knowledge about molecular evolution.

Expected profile and skills of the candidate

The candidate should have a strong interest in molecular evolution and population genetics; should be willing to acquire an extensive training in bioinformatics and statistical data analysis; should have good communication skills in English and enjoy scientific work.

Relevant publications from the team

Kawakami T, Mugal CF, Suh A, Nater A, Burri R, Smeds L, Ellegren H. 2017. Whole-genome patterns of linkage disequilibrium across flycatcher populations clarify the causes and consequences of fine-scale recombination rate variation in birds. *Mol Ecol* 26:4158-4172. doi: 10.1111/mec.14197

Boman J, Qvarnstrom A, Mugal CF. 2024. Regulatory and evolutionary impact of DNA methylation in two songbird species and their naturally occurring F(1) hybrids. *BMC Biol* 22:124. doi: 10.1186/s12915-024-01920-2

Ponger L, Duret L, Mouchiroud D. 2001. Determinants of CpG islands: expression in early embryo and isochore structure. *Genome Res* 11:1854-1860. doi: 10.1101/gr.174501

References

Al Adhami H, Bardet AF, Dumas M, Cleroux E, Guibert S, Fauque P, Acloque H, Weber M. 2022. A comparative methylome analysis reveals conservation and divergence of DNA methylation patterns and functions in vertebrates. *BMC Biol* 20:70. doi: 10.1186/s12915-022-01270-x

Angeloni A, Bogdanovic O. 2021. Sequence determinants, function, and evolution of CpG islands. *Biochem Soc Trans* 49:1109-1119. doi: 10.1042/BST20200695

Baudat F, Buard J, Grey C, Fledel-Alon A, Ober C, Przeworski M, Coop G, de Massy B. 2010. PRDM9 is a major determinant of meiotic recombination hotspots in humans and mice. *Science* 327:836-840. doi: 10.1126/science.1183439

Berglund J, Quilez J, Arndt PF, Webster MT. 2014. Germline methylation patterns determine the distribution of recombination events in the dog genome. *Genome Biol Evol* 7:522-530. doi: 10.1093/gbe/evu282

Bird A, Taggart M, Frommer M, Miller OJ, Macleod D. 1985. A fraction of the mouse genome that is derived from islands of nonmethylated, CpG-rich DNA. *Cell* 40:91-99. doi: 10.1016/0092-8674(85)90312-5

Boman J, Qvarnstrom A, Mugal CF. 2024. Regulatory and evolutionary impact of DNA methylation in two songbird species and their naturally occurring F(1) hybrids. *BMC Biol* 22:124. doi: 10.1186/s12915-024-01920-2

Chase MA, Scofield DG, Kraft F, Segami JC, Alund M, Qvarnstrom A, Wheatcroft D, Mugal CF. 2023. The combination of HiFi and HiC sequencing technologies enables the investigation of structural variants in speciation of *Ficedula* flycatchers. *unpublished PhD thesis manuscript*.

Cohen NM, Kenigsberg E, Tanay A. 2011. Primate CpG islands are maintained by heterogeneous evolutionary regimes involving minimal selection. *Cell* 145:773-786. doi: 10.1016/j.cell.2011.04.024

Deaton AM, Bird A. 2011. CpG islands and the regulation of transcription. *Genes Dev* 25:1010-1022. doi: 10.1101/gad.2037511

Gardiner-Garden M, Frommer M. 1987. CpG islands in vertebrate genomes. *J Mol Biol* 196:261-282. doi: 10.1016/0022-2836(87)90689-9

Kawakami T, Mugal CF, Suh A, Nater A, Burri R, Smeds L, Ellegren H. 2017. Whole-genome patterns of linkage disequilibrium across flycatcher populations clarify the causes and consequences of fine-scale recombination rate variation in birds. *Mol Ecol* 26:4158-4172. doi: 10.1111/mec.14197

Kawakami T, Smeds L, Backstrom N, Husby A, Qvarnstrom A, Mugal CF, Olason P, Ellegren H. 2014. A high-density linkage map enables a second-generation collared flycatcher genome assembly and reveals the patterns of avian recombination rate variation and chromosomal evolution. *Mol Ecol* 23:4035-4058. doi: 10.1111/mec.12810

Long HK, King HW, Patient RK, Odom DT, Klose RJ. 2016. Protection of CpG islands from DNA methylation is DNA-encoded and evolutionarily conserved. *Nucleic Acids Res* 44:6693-6706. doi: 10.1093/nar/gkw258

Long HK, Sims D, Heger A, Blackledge NP, Kutter C, Wright ML, Grutzner F, Odom DT, Patient R, Ponting CP, et al. 2013. Epigenetic conservation at gene regulatory elements revealed by non-methylated DNA profiling in seven vertebrates. *Elife* 2:e00348. doi: 10.7554/eLife.00348

Parvanov ED, Petkov PM, Paigen K. 2010. Prdm9 controls activation of mammalian recombination hotspots. *Science* 327:835. doi: 10.1126/science.1181495

Ponger L, Duret L, Mouchiroud D. 2001. Determinants of CpG islands: expression in early embryo and isochore structure. *Genome Res* 11:1854-1860. doi: 10.1101/gr.174501

Saxonov S, Berg P, Brutlag DL. 2006. A genome-wide analysis of CpG dinucleotides in the human genome distinguishes two distinct classes of promoters. *Proc Natl Acad Sci U S A* 103:1412-1417. doi: 10.1073/pnas.0510310103