

M2 - Internship

Research project title :

RER express - Ruminant Endogenous Retrovirus expression

Co-supervisors :

Jocelyn Turpin (IVPC ; INRAE) jocelyn.turpin@univ-lyon1.fr;

Emmanuelle Lerat (LBBE ; CNRS) emmanuelle.lerat@univ-lyon1.fr;

Vincent Navratil (PRABI ; Univ Lyon 1) vincent.navratil@univ-lyon1.fr

Laboratories :

Université Claude Bernard Lyon1 :

LBBE (Campus DOUA, Villeurbanne)

IVPC (Campus Gerland, Lyon)

Key words :

Bioinformatics, Endogenous Retrovirus, active transposable elements, transcriptomic

Project description :

Retroviridae is a large family of RNA viruses infecting humans and animals and responsible for a variety of diseases such as severe immunodeficiency like AIDS in humans, or cancers. Retroviruses are also major components of the mammalian genomes. Known as endogenous retroviruses (ERV), they result from ancient integrations of exogenous infectious retroviruses into the DNA of germ cells. Approximately 8% of the human genome is made of ERV sequences. They have long been considered as "junk DNA" but it is now well established that they participate to the genome evolution and even bring indispensable functions for the mammalian life and development (Almojil et al., 2021).

Most ERVs originated from now extinct exogenous retroviruses. Still, a few animal species have been described, displaying at the same time ERVs and their still circulating exogenous counterparts (Chiu and VandeWoude, 2021). Among them, the small ruminants' genomes carry multiple ERV copies of β - retroviruses which co-exist with their exogenous counterparts, namely JSRV (Jaagsiekte sheep retrovirus) and ENTV (Enzootic Nasal Tumor Virus) (Arnaud et al., 2007). JSRV and ENTV are oncogenic viruses responsible for respiratory cancers in sheep and goats (Monot et al., 2015).

We have recently characterized and annotated the ERV families across the sheep and goat (wild and domesticated) genomes showing a different dynamic among analyzed species (Verneret et al., 2024). Our results suggest that among them two families may still have transpositional potential. One of them is a family closely related to the circulating oncogenic exogenous retroviruses represented by many full-length copies with conserved ORFs and no syntenic insertions between sheep and goats, whereas the second family is composed of shorter copies with only partial retroviral genes and shared insertions between small ruminants. This suggests different transposition mechanism between the two families.

The main goal of this project is to gather evidences that these two ERV families are still active by analyzing their expression. We will also determine the evolutionary implications on the neighboring gene expression in the different species. This project will be subdivided in three parts corresponding to different approaches:

- **Blind approach:** by using the taxonomic assignment of high-throughput sequencing data (SRA-STAT) which gather over 25 million samples (Katz et al., 2022) including over 15k RNA-Seq

samples of domestic sheep and goats. This approach based on kmer dictionaries, reports the taxonomic composition of reads within a sequencing run, that will be exploited with the help of the [Virome@tlas](https://www.viromeatlas.com/) consortium. SRA-STAT references include kmers for JSRV/ENTV and their endogenous counterpart, which will be exploited to identify, without *a priori*, in which tissues these ERVs might be expressed and what factors (organ, age, sexe, genus, species, breed...) are associated to their expression.

- **Targeted approach:** we will analyze RNASeq data available in public databases to measure the global expression of all ERV families within targeted samples that may play a role in the transmission of ERVs: genital organs, conceptus, semen etc., and compare the results among the different species.
- **Locus specific expression of ERV:** ERV are polymorphic and repeated elements which are by consequence difficult to map (Lanciano and Cristofari, 2020). A benchmark of the available tools for short and long reads sequencing for the allele specific expression will be done, to identify the active copies.

The student will develop expertise in handling high-throughput sequencing data and performing statistical analysis using R. He/She will become familiar with key computational methods commonly used in transcriptomics and gain a solid understanding of endogenous retroviruses (ERVs) and the analysis of gene expression data.

Expected profile and skills of the candidate :

The candidate should have a strong interest in genomics and transcriptomics and be eager to undergo comprehensive training in bioinformatics and statistical data analysis.

Bibliography :

- Almojil, D., Bourgeois, Y., Falis, M., Hariyani, I., Wilcox, J., and Boissinot, S. (2021). The Structural, Functional and Evolutionary Impact of Transposable Elements in Eukaryotes. *Genes* 12, 918. doi: 10.3390/genes12060918
- Arnaud, F., Caporale, M., Varela, M., Biek, R., Chessa, B., Alberti, A., et al. (2007). A paradigm for virus-host coevolution: sequential counter-adaptations between endogenous and exogenous retroviruses. *PLoS pathogens* 3, e170. doi: 10.1371/journal.ppat.0030170
- Chiu, E. S., and VandeWoude, S. (2021). Endogenous Retroviruses Drive Resistance and Promotion of Exogenous Retroviral Homologs. *Annual Review of Animal Biosciences* 9, 225–248. doi: 10.1146/annurev-animal-050620-101416
- Katz, K., Shutov, O., Lapoint, R., Kimelman, M., Brister, J. R., and O'Sullivan, C. (2022). The Sequence Read Archive: a decade more of explosive growth. *Nucleic Acids Research* 50, D387–D390. doi: 10.1093/nar/gkab1053
- Lanciano, S., and Cristofari, G. (2020). Measuring and interpreting transposable element expression. *Nature reviews. Genetics* 21, 721–736. doi: 10.1038/s41576-020-0251-y
- Monot, M., Archer, F., Gomes, M., Mornex, J.-F., and Leroux, C. (2015). Advances in the study of transmissible respiratory tumours in small ruminants. *Veterinary microbiology* 181, 170–177. doi: 10.1016/j.vetmic.2015.08.008
- Verneret, M., Leroux, C., Faraut, T., Navratil, V., Lerat, E., and Turpin, J. (2024). A genome-wide study of ruminants reveals two endogenous retrovirus families still active in goats. *bioRxiv*, 2024.06.21.600049. doi: 10.1101/2024.06.21.600049