IMGT® PhD thesis topic co-funded by the Labex MabImprove

Starting date: autumn 2021

Acronym: **ComPanY**

Thesis title: **Com**parative genomic analysis of the adaptive immune response loci of ***Pan*** *troglodytes*

*Subtitle: Application of statistical and machine learning methodologies to genomic data*

The Human Genome Project, which remains the world's largest collaborative genetics project was launched in 1990 and it was declared complete in 2003. Since then, the methodological and technological advances in the field have made it possible to obtain the whole genome of a species in much shorter time and in much deeper detail compared to the first version of the human genome. In 2018, a *Pan troglodytes* (chimpanzee) representative genome with 124x genome coverage was made available via NCBI with GenBank assembly accession: GCA_002880755.3 .

The Hominidae family, whose members are known as great apes, includes eight extant species in four genera: *Pongo* containing the Bornean, Sumatran and Tapanuli orangutan; *Gorilla* containing the eastern and western gorilla; *Pan* containing the common chimpanzee (*Pan troglodytes*) and the bonodo or pygmy chimpanzee (*Pan paniscus*) ; and *Homo,* of which only modern humans (*Homo sapiens*) remain.

Scientists have long been interested in the functional genetic differences that distinguish humans from other ape species including the great apes (1). Comparative genomic methods have allowed researchers to gather information about genetic variation, differential gene expression, and evolutionary dynamics in primates that were indiscernible using previous data and methods (2).

In the current thesis project, we are interested in investigating the genomic organization of loci of immunoglobulin (IG) and T cell receptor (TR) genes involved in the adaptive immune response of the *Pan troglodytes* genome in comparison to the other Hominidae family members already available within IMGT as well as the Old World monkeys namely *Macaca mulatta* and *Macaca fascicularis*.

More specifically, we will address questions such as: do different species house different numbers of genes within their genomes as well as different patterns in the distribution of genes throughout the genome? Are regulatory elements of orthologous sequences conserved to the same degree as coding sequences? To address these questions statistical and machine methodologies will be implemented.

The content and structure of a genome is the product of the molecular and population genetic forces, which act upon that genome. We will try to decipher the forces in molecular evolution that have acted upon the *Pan troglodytes* genome by identifying duplicated genes and/or sets of genes. Comparative studies of different vertebrates for different loci of IG and TR are already available in the scientific literature (3-4), however the study presented in this document, to the best of our knowledge, has not been addressed so far.

Although the thesis work will remain at addressing fundamental research questions, the identified similarities and differences at the genomic level of the adaptive immune system of the chimpanzee and the other genomes could have potential implications for the results of therapeutic trials. Additionally, correlation studies of the expressed IG and/or TR genes in a vaccination situation or in a pathological situation might give insights in the distinct role of each gene.

References


1. High-resolution comparative analysis of great ape genomes. Kronenberg Z N *et al*. Science. 2018 Jun 8;360(6393):eaar6343. doi: 10.1126/science.aar6343.
2. Comparative primate genomics: emerging patterns of genome content and dynamics. Rogers J & Gibbs R. Nat Rev Genet. 2014 May; 15(5): 347–359. doi: 10.1038/nrg3707
3. IMGT® Biocuration and Comparative Study of the T Cell Receptor Beta Locus of Veterinary Species Based on *Homo sapiens* TRB. P Pégorier P et al. Front. Immunol., 05 May 2020. doi.org/10.3389/fimmu.2020.00821
4. Evolution of the T cell receptor (TR) loci in the adaptive immune response: the tale of the TRG locus in mammals. Antonacci R *et al*. Genes 2020, 11, Jun 5;11(6):E624. doi: 10.3390/genes11060624.PMID: 32517024


**PhD profile candidate**: Master's degree or equivalent in biostatistics, bioinformatics or equivalent fields. The candidate will use statistical and machine learning methodologies to address the scientific questions. Enquiries, submissions to be sent to: sofia.kossida@igh.cnrs.fr