Genetic diversity of chimpanzees' species

Here we study the genetic diversity of a subset of data about chimpanzees and bonobos from Manuel et al. (2016 Science). The data are in the form of a genotypic data matrix where each entry is either 0, 1 or 2 for the number of alternative allele per site.

To summarize the genetic diversity and show what are the most and less diverse populations, we can compute the Expected and Observed Heterozygocity and estimate the Fixation index / inbreeding coefficient (deviation from HWE) noted F_{IS}.

Figure 1 Distribution range of Chimpanzee Pan troglodytes. Blue: verus, Red: ellioti, Green: troglodytes, Orange: schweinfurthii, Pink: Bonobo - Pan paniscus

In the graph bellow, we can observe the *Mean expected heterozygosity* and the *Expected vs Observed heterozygosity*.

The expected mean heterozygosity is calculated across the SNPs for each species as the probability to take two sequences with different alleles based on the allele frequency of the population.

The observed mean heterozygosity is the actual proportion of heterozygotes sites (i.e. number of 1's) per individual in a population.
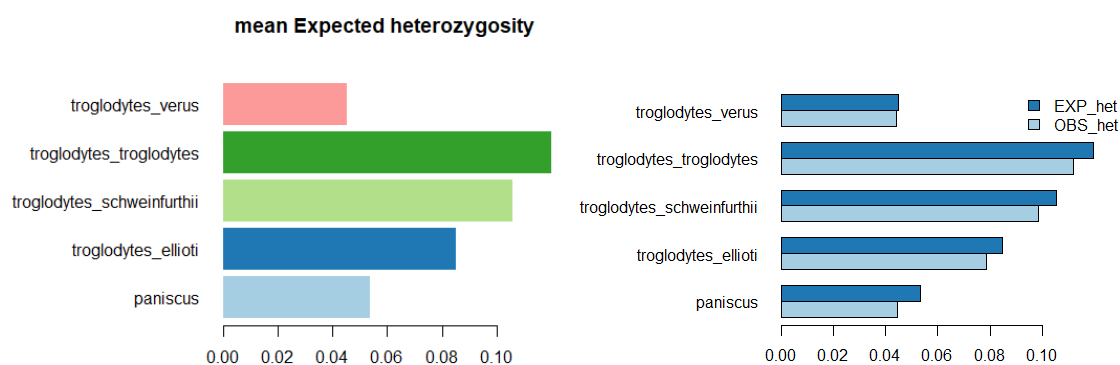

Figure 2 (left) Mean expected Heterozygosity (right) Expected vs observed heterozygosity

The species with the higher Expected and Observed heterozygosity is *P. t. troglodytes*, chimps from Gabon and Congo, meaning they have the highest genetic diversity. The observed heterozygosity is lower than the expected one in all the population because the expectation is made under Hardy-Weinberg assumption which does not reflect the reality.

The two population with the least diversity, *P. t. verus* and *P. paniscus* (bonobos) might be explained by isolation on the west for *P. t. verus* and by sympatric speciation for the

P. paniscus. As mentioned by Manuel et al. the two species, chimpanzee and bonobo, can hybridize in captivity but in the wild it remains unclear. We can imagine that they do not mix on a regular basis thus creating population with a low diversity even if they are close to a bigger population with a bigger diversity (they are surrounded by *troglodytes* and *schweinfurthii.*

Inbreeding coefficient

The estimate for the inbreeding coefficient is based on the excess of homozygotes as

$$F_{IS} = 1 - \frac{H_o}{H_e}$$

Inbreeding reduce the genetic diversity in a population but here we observe that *Pan verus*, even if they have a small genetic diversity, they have a really small inbreeding coefficient which is not clearly explained. For *Pan paniscus* it is clear that they have a high inbreeding coefficient related to their poor genetic diversity and geographical isolation.