

Study of the human African genome landscape through the analysis of complete genomes

1. Introduction and goals

Understanding the genetic diversity in humans within the African continent is pivotal to have a full picture of the demographic history of the human species. Here, we study complete genome sequences of a diverse panel of African individuals in terms of geographical location, linguistic context and lifestyle.

Most African diversity studies have mainly focused on uniparental markers or selected autosomal markers, which introduce an ascertainment bias in the analysis. The analysis of complete genomes has been poorly developed in the study of African human genomics and internal population diversity.

The main goals of this study are:

1. Characterisation of the **internal human diversity in Africa** overcoming ascertainment bias-related problems by using complete genomes.
2. Characterization of the **deepest splits in the human lineage** and the processes (such as migrations and admixtures) that have shaped the current genetic map.



Figure 1. African samples used in this project.

2. Uniparental markers

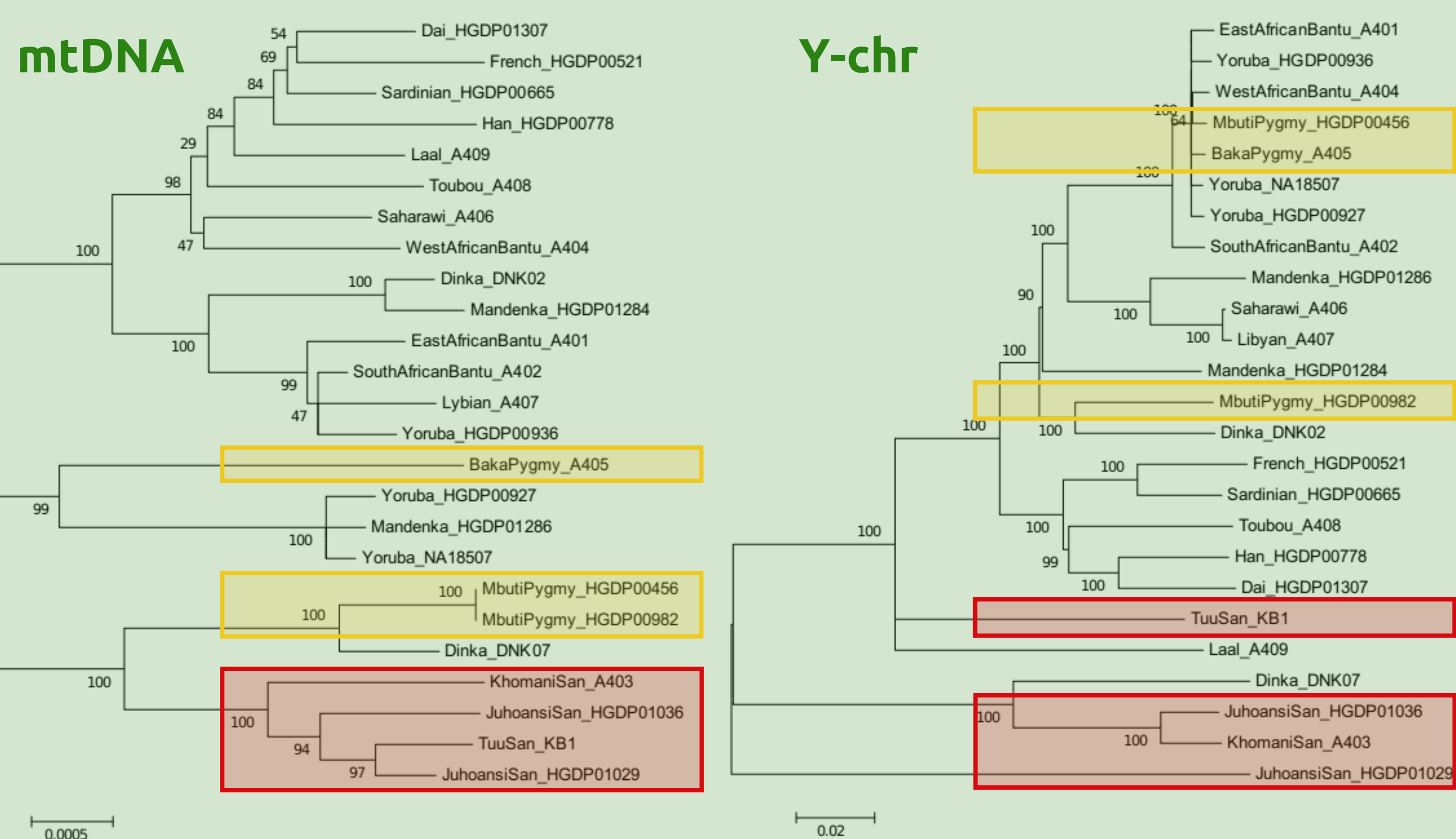


Figure 2 (a & b). Maximum likelihood phylogenetic trees of the **mitochondrial** (left) and **Y-chromosome** (right). The basal branches are mainly composed by hunter-gatherer samples, highlighted in red (Khoisan) and yellow (Pygmies), with discrepancies between Y and mtDNA.

3a. Descriptive analyses

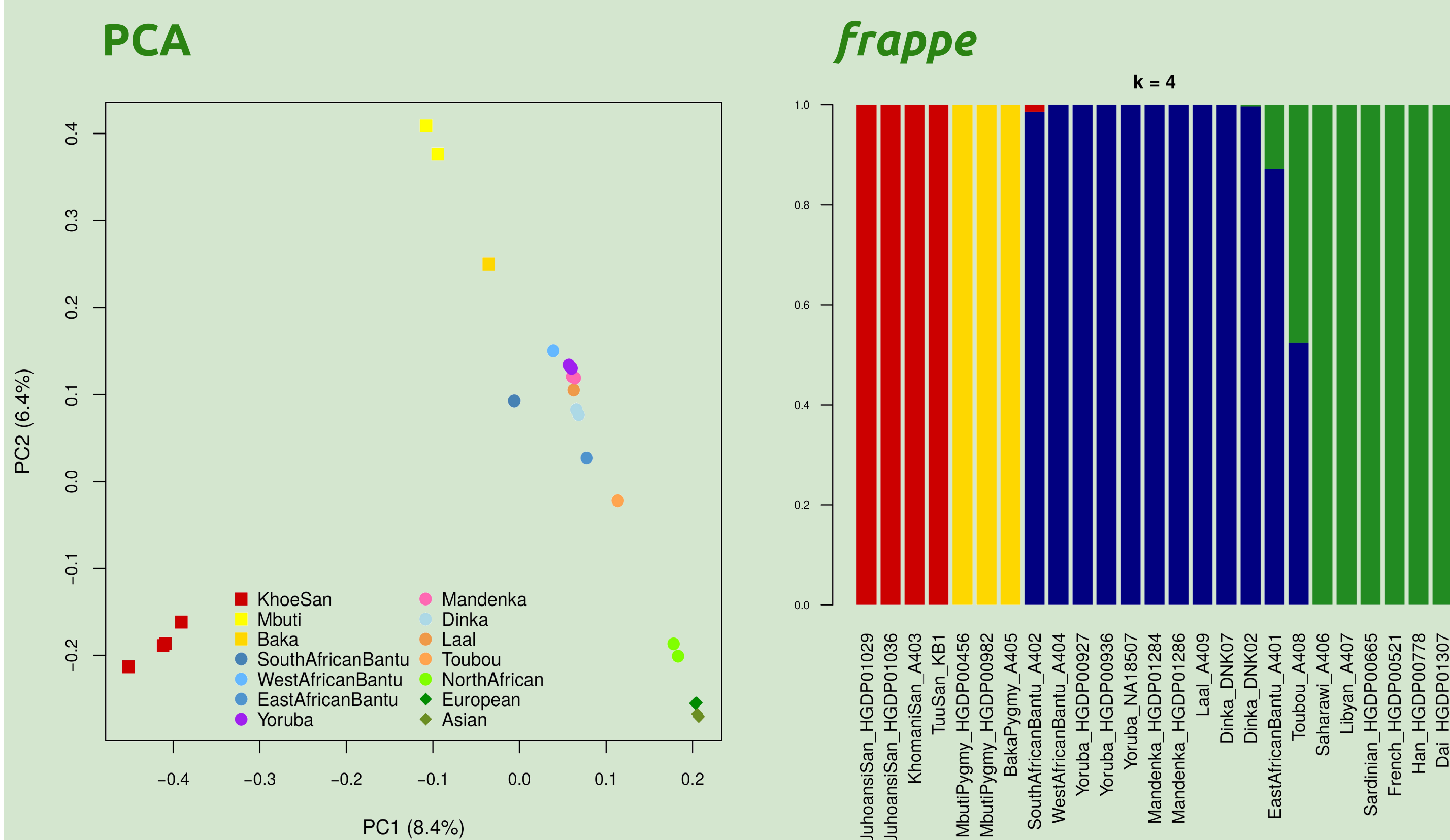


Figure 3 (a & b). **Principal component analysis** (left) and **individual ancestry components** (right). (3a) PC1 separates the Khoisan from the other samples, whereas PC2 differentiates the Pygmies. (3b) *frappe* results for four ancestral components. The four groups correspond to the Khoisan, the Pygmies, the rest of Sub-Saharan individuals and the North African and Eurasian samples.

3b. Descriptive analyses (cont.)

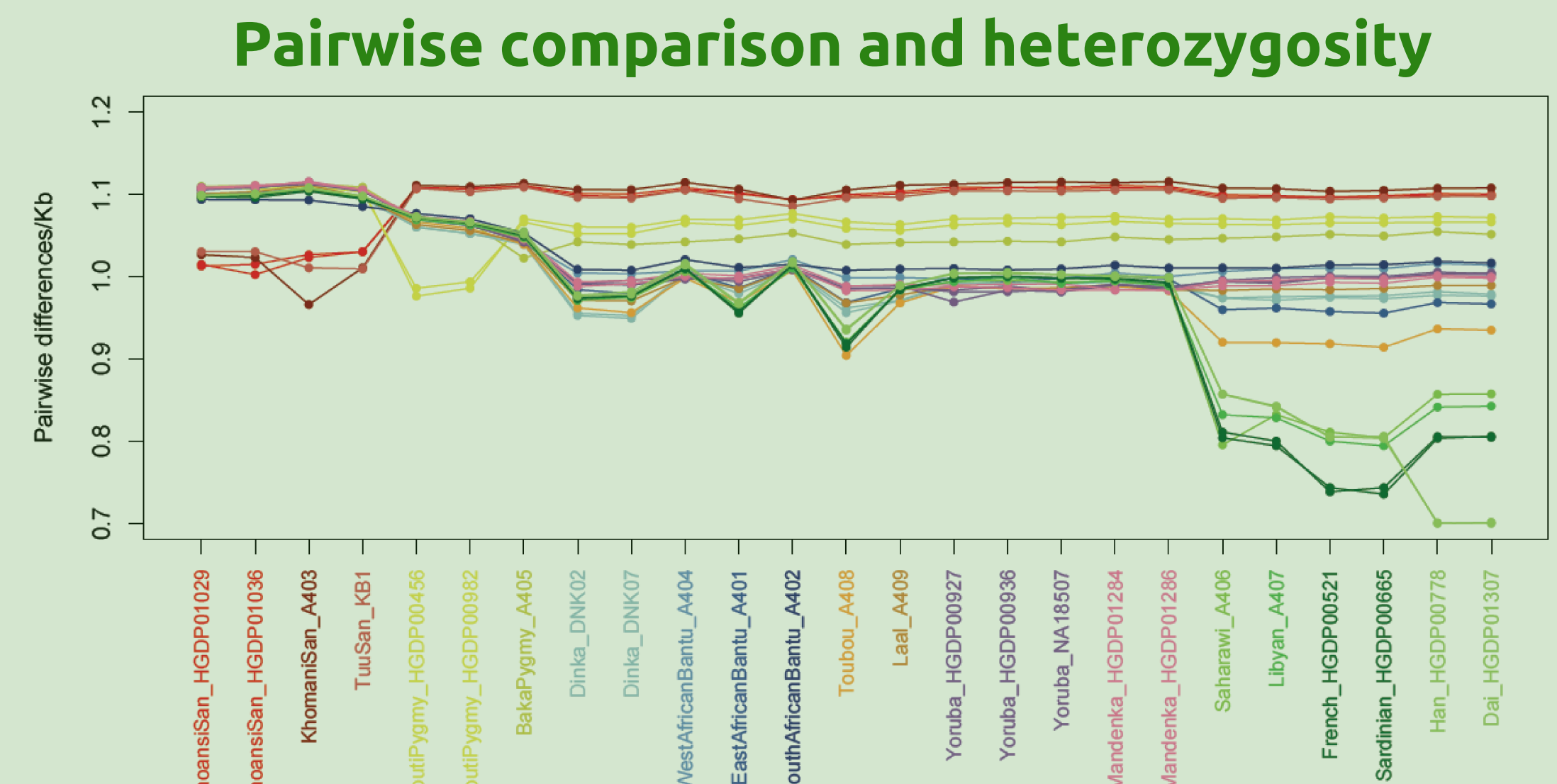
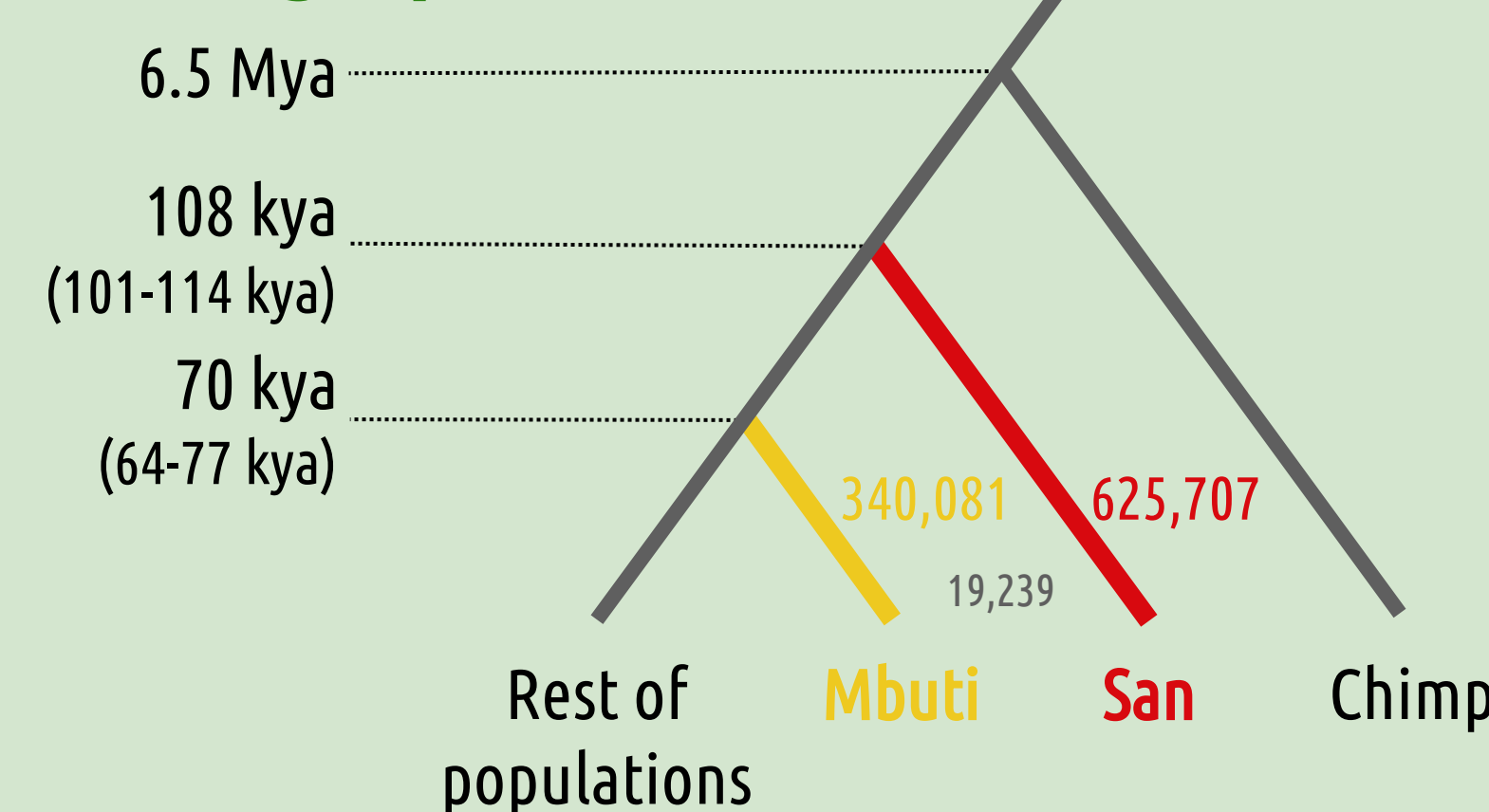


Figure 4. **Heterozygosity and pairwise differences** between the samples under study. The Khoisan (red) and the Pygmies (yellow) are found to be the most differentiated with respect to the other samples.

4. Demographic inferences

Demographic model



PSMC

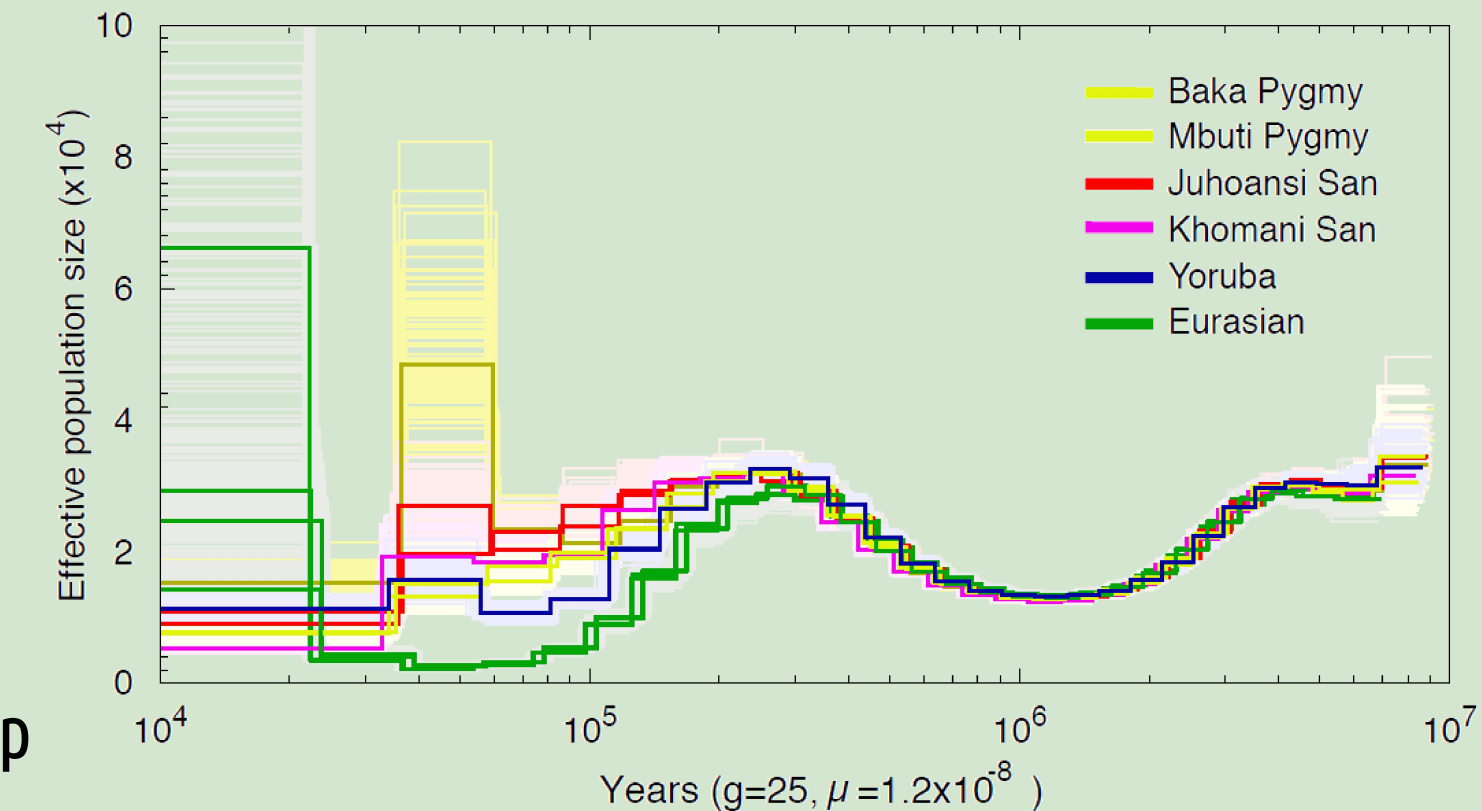


Figure 5 (a & b). **Demographic model** characterising the first splits in the human lineage (left) and **Pairwise Sequentially Markovian Coalescent analysis** (right). (5a) The ancestry status (ancestral/derived) was assigned to all private variants of San and Mbuti samples, and the number of private derived variants were computed for each population. To increase the analysis power, "rest of populations" include all non-hunter-gatherer samples in the study + all Illumina samples from the 1000 Genomes Project. Divergence times were estimated using a Bayesian MCMC coalescent approach (*G-PhoCS*). (5b) PSMC results, where the effective population size of the hunter-gatherers is higher than the other African samples (Yoruba) and non-African (Eurasian).

5. Archaic admixture

D introgression test

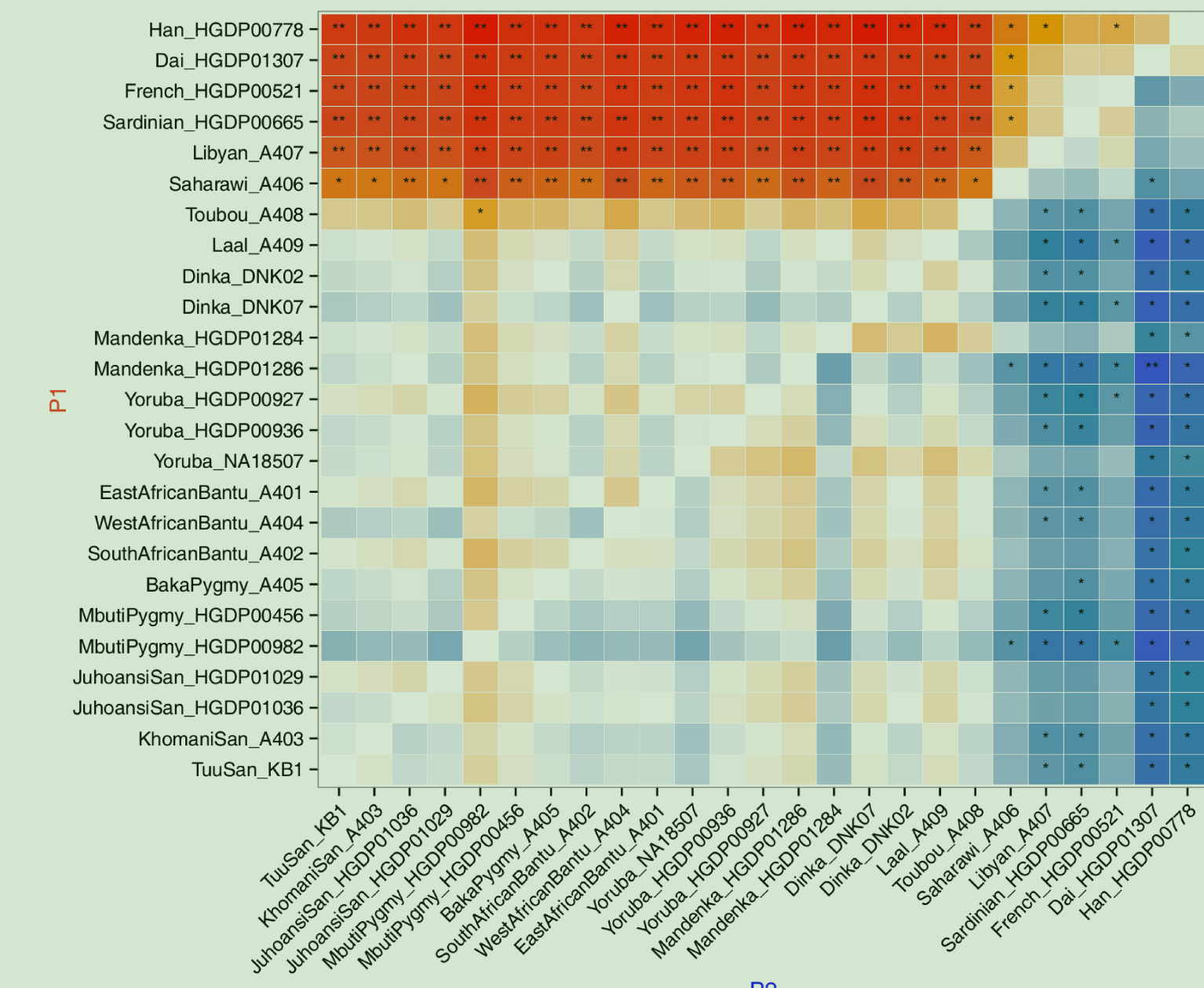


Figure 6. **Archaic admixture from archaic hominins** into studied samples. The results of pairwise *D* test for Neanderthal (above diagonal) and Denisovan (below diagonal) introgression are displayed. No significant results (*) are observed in the Sub-Saharan samples, contrary to the North Africans and Eurasians.

6. Conclusions

1. **Four main genetic groups** in Africa:
 - Heterogeneity in hunter-gatherers:
 - Khoisan
 - Pygmies
 - North Africans, closely related with out-of-Africa populations.
 - Rest of Sub-Saharan populations, which show large genetic homogeneity.
2. **Hunter-gatherers**: the most differentiated and diverse: first split round 100kya (Khoisan), highest number of genetic differences per kb, more private variants, highest N_e .
3. **No signals of archaic introgression** from available genomes (Denisova, Neanderthal) in Sub-Saharans, but present in North Africans.