
This is the **published version** of the bachelor thesis:

Noguera Segura, Arnau. Genomic clues to our history : natural selection signatures in the human genome. 2023. 1 pag. (833 Grau en Genètica)

This version is available at <https://ddd.uab.cat/record/277619>

under the terms of the  license

GENOMIC CLUES TO OUR HISTORY: NATURAL SELECTION SIGNATURES IN THE HUMAN GENOME

UAB

Universitat Autònoma
de Barcelona

Arnau Noguera Segura

2023

Bachelor's Degree in Genetics

Data analysis

Introduction

As human populations became exposed to new selective pressures, **positive natural selection** favored adaptive genetic variants, resulting in “**selective sweeps**”.

These events leave signatures in the genetic variation pattern of our genomes:

- **Long haplotypes:** linkage disequilibrium.
- **Site frequency spectrum:** population differences, high-frequency derived alleles, rare SNPs.
- **Protein changes:** nonsynonymous substitutions.

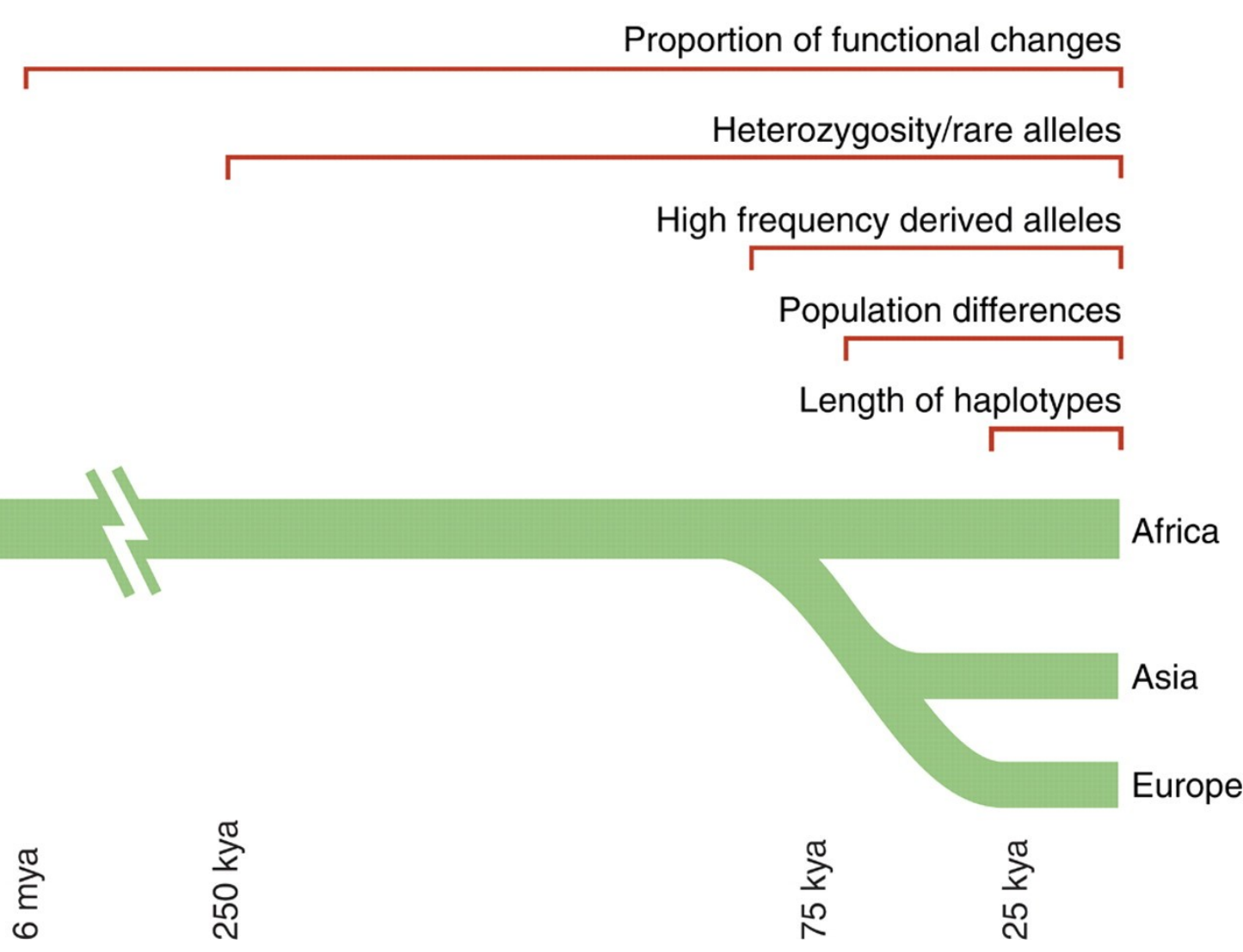


Figure 2. Types of positive selection signatures and their evolutionary time spans. [4]

Diverse **statistics** have been developed for the identification of positive selection signals on the human genome.

Several studies have implemented positive selection statistics for genome-wide detection of **candidate adaptive loci** via genome scans

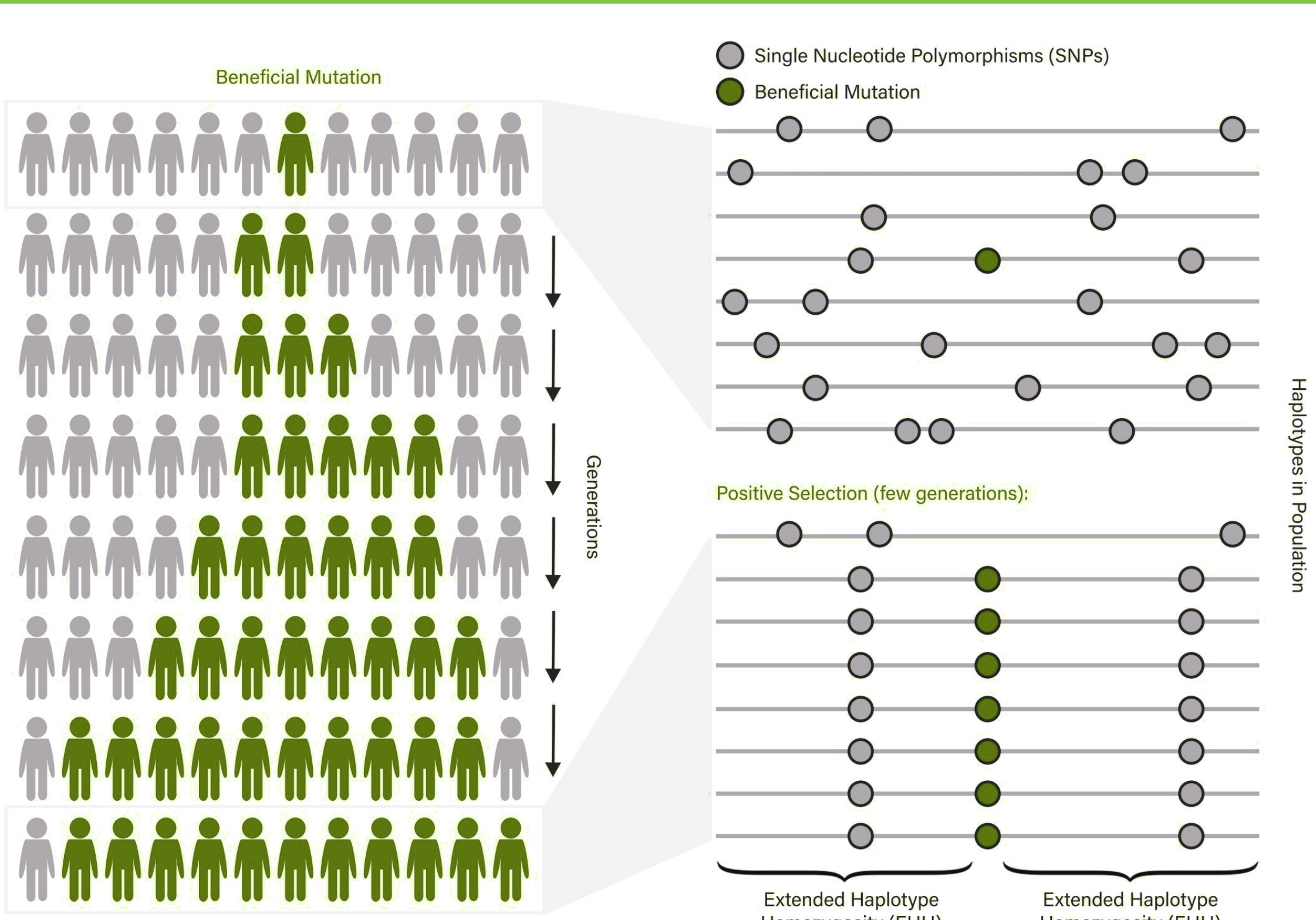


Figure 1. Short-term haplotypic effects of a hard selective sweep. [1]

Objectives

1. To build a **merged repository** of published candidate regions for positive selection and make it publicly available on GitHub.
2. To measure how much of the human genome represent the positively selected regions gathered in the merged repository.
3. To identify regions repeatedly detected by multiple articles as the most robust positively selected candidates.

Methodology

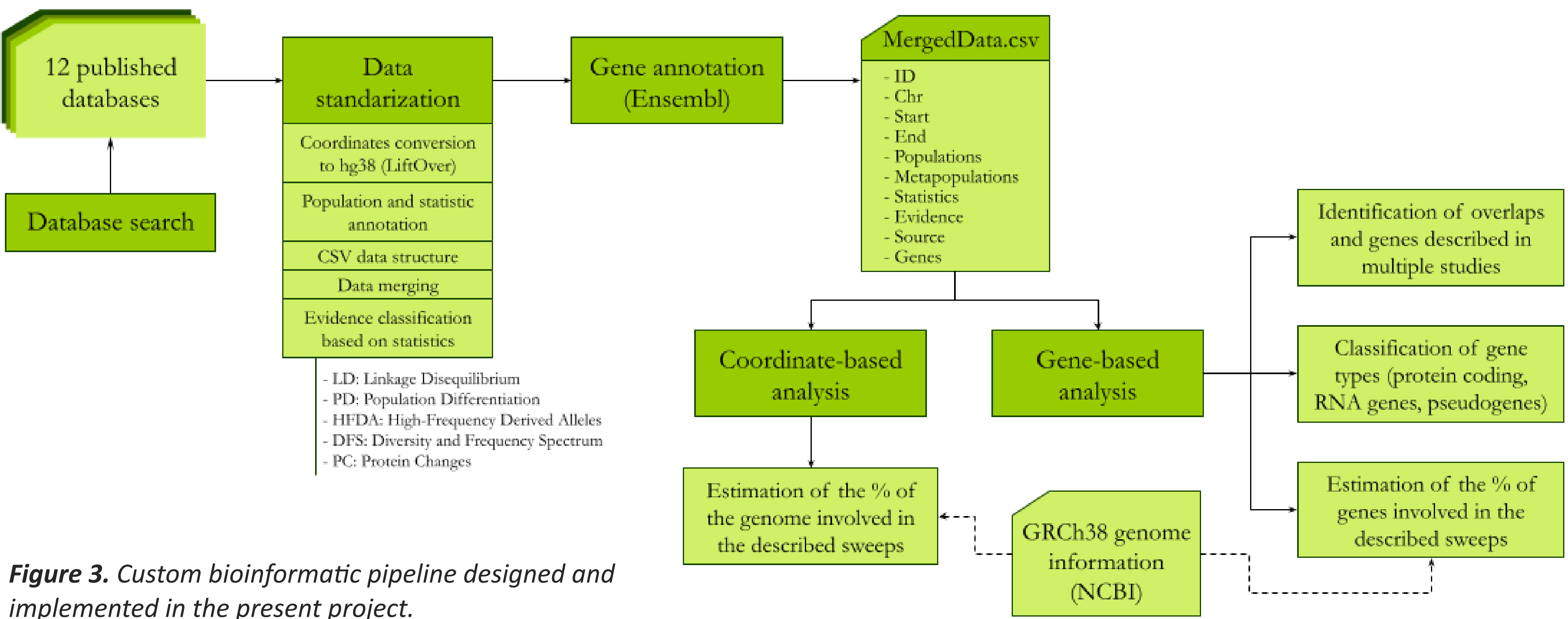


Figure 3. Custom bioinformatic pipeline designed and implemented in the present project.

Results

22,153 candidate regions compiled from 12 articles:

- 1.27 Gbp (40,8%).
- 28,158 genes (44,97%).

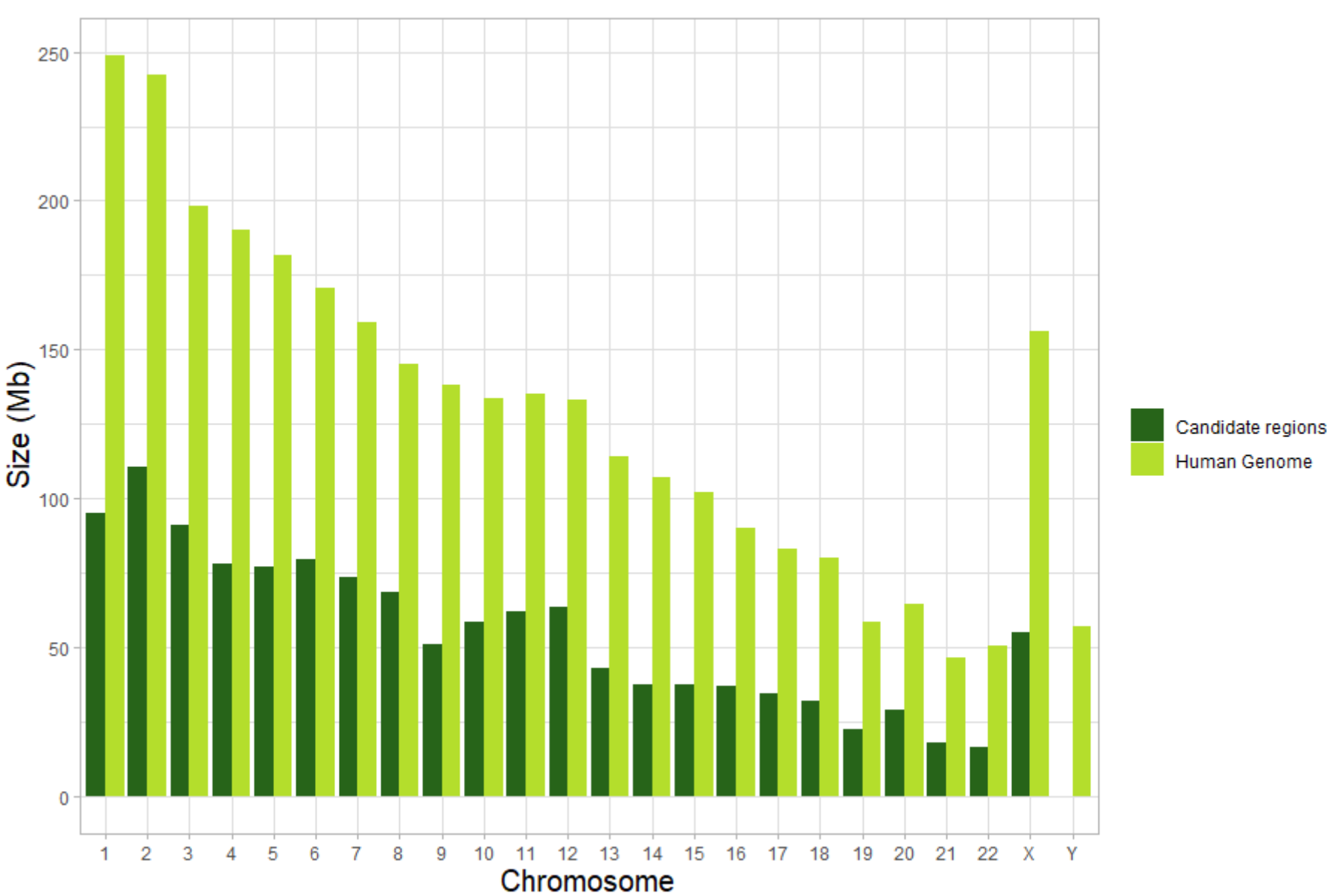


Figure 4. Positively selected regions length size versus chromosome length across chromosomes (in Mbp).

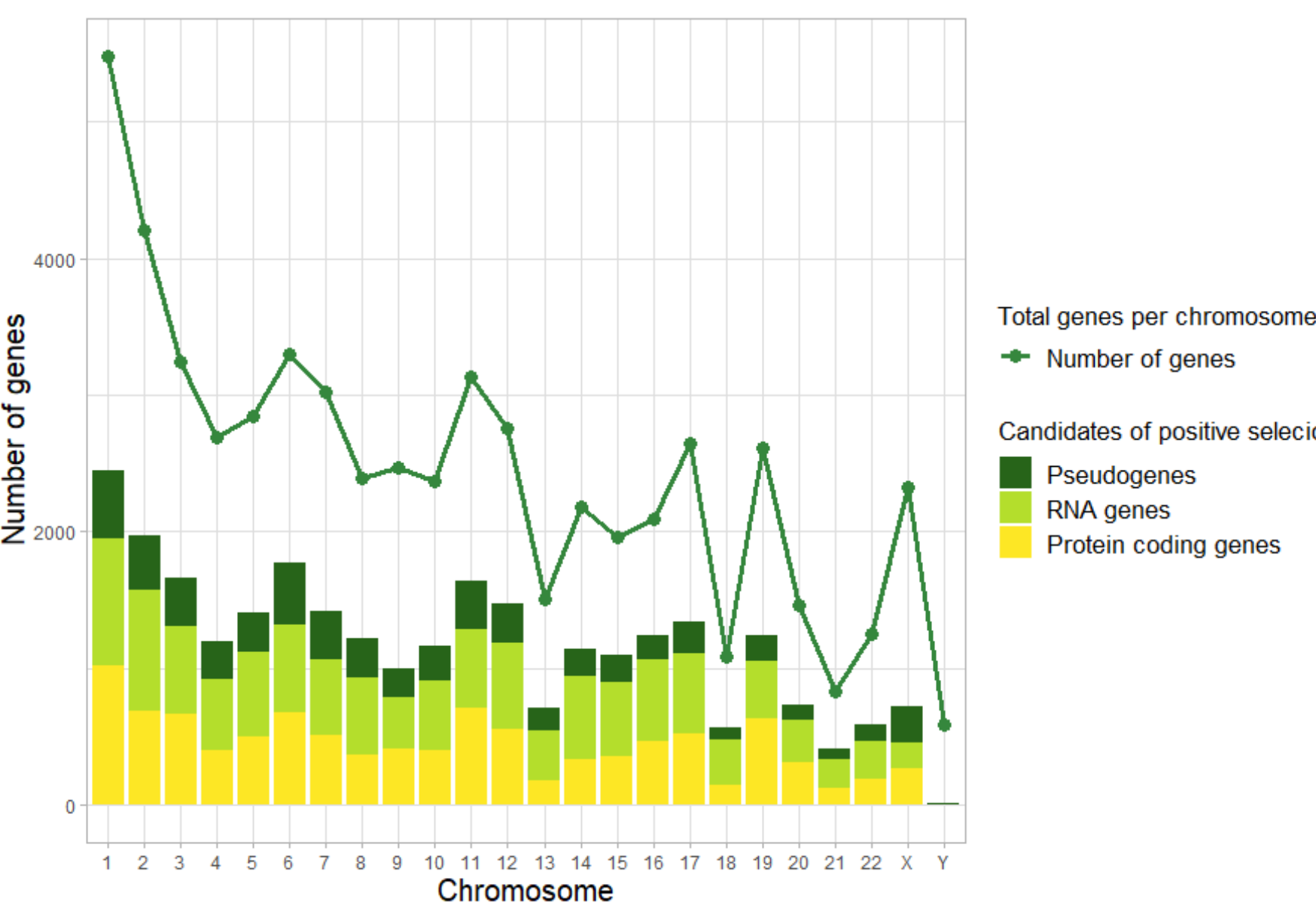


Figure 5. Genes annotated in candidate regions versus total annotated genes per chromosome.

The entirety of the candidate regions is not responsible for the alleged selective sweeps:
1) actual targets, 2) linked regions, and
3) statistical limitations.

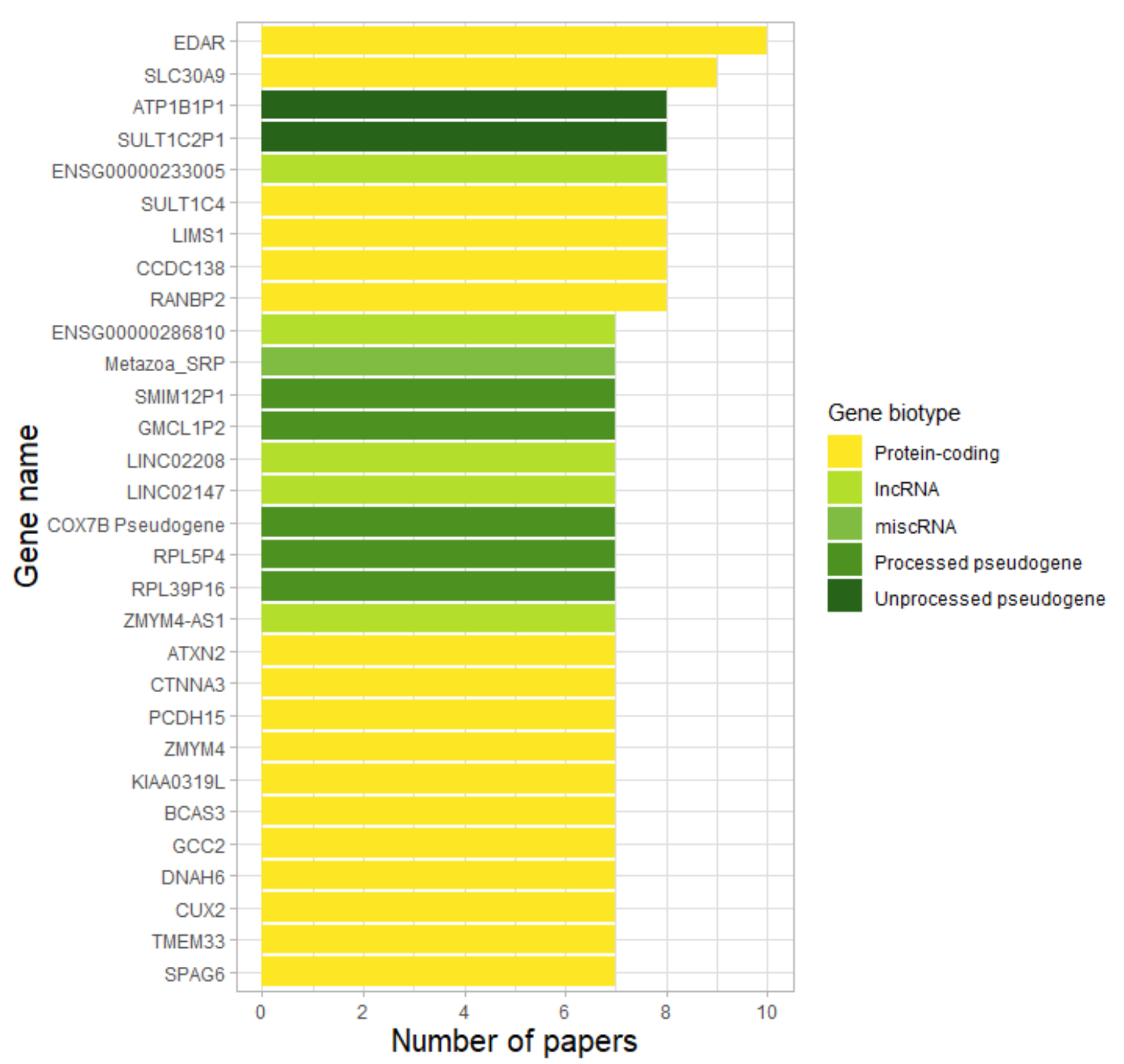


Figure 6. Top 30 candidate loci for positive selection according to the number of times that they were independently detected.

Conclusions

1. A merged repository of published human candidate regions for positive selection was built and made publicly available in GitHub (see QR code).
2. The extracted candidate regions represent a striking half of the human genome, reaffirming the need for developing methods to isolate individual causal mutations.
3. Several genes were repeatedly detected across studies, suggesting their involvement in human adaptation.



References

- [1] Hernandez, M., & Perry, G. H. (2021). Scanning the human genome for "signatures" of positive selection: Transformative opportunities and ethical obligations. *Evolutionary Anthropology*, 30, 113–121. <https://doi.org/10.1002/evan.21893>
- [2] Murga-Moreno, J., Coronado-Zamora, M., Bodelón, A., Barbadilla, A., & Casillas, S. (2019). PopHumanScan: the online catalog of human genome adaptation. *Nucleic Acids Research*, 47, D1080–D1089. <https://doi.org/10.1093/nar/gky959>
- [3] Oleksyk, T. K., Smith, M. W., & O'Brien, S. J. (2010). Genome-wide scans for footprints of natural selection. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, 365, 185–205. <https://doi.org/10.1098/rstb.2009.0219>
- [4] Sabeti, P. C., Schaffner, S. F., Fry, B., Lohmueller, J., Varrilly, P., Shamovsky, O., Palma, A., Mikkelsen, T. S., Altshuler, D., & Lander, E. S. (2006). Positive natural selection in the human lineage. *Science*, 312, 1614–1620. <https://doi.org/10.1126/science.1124309>