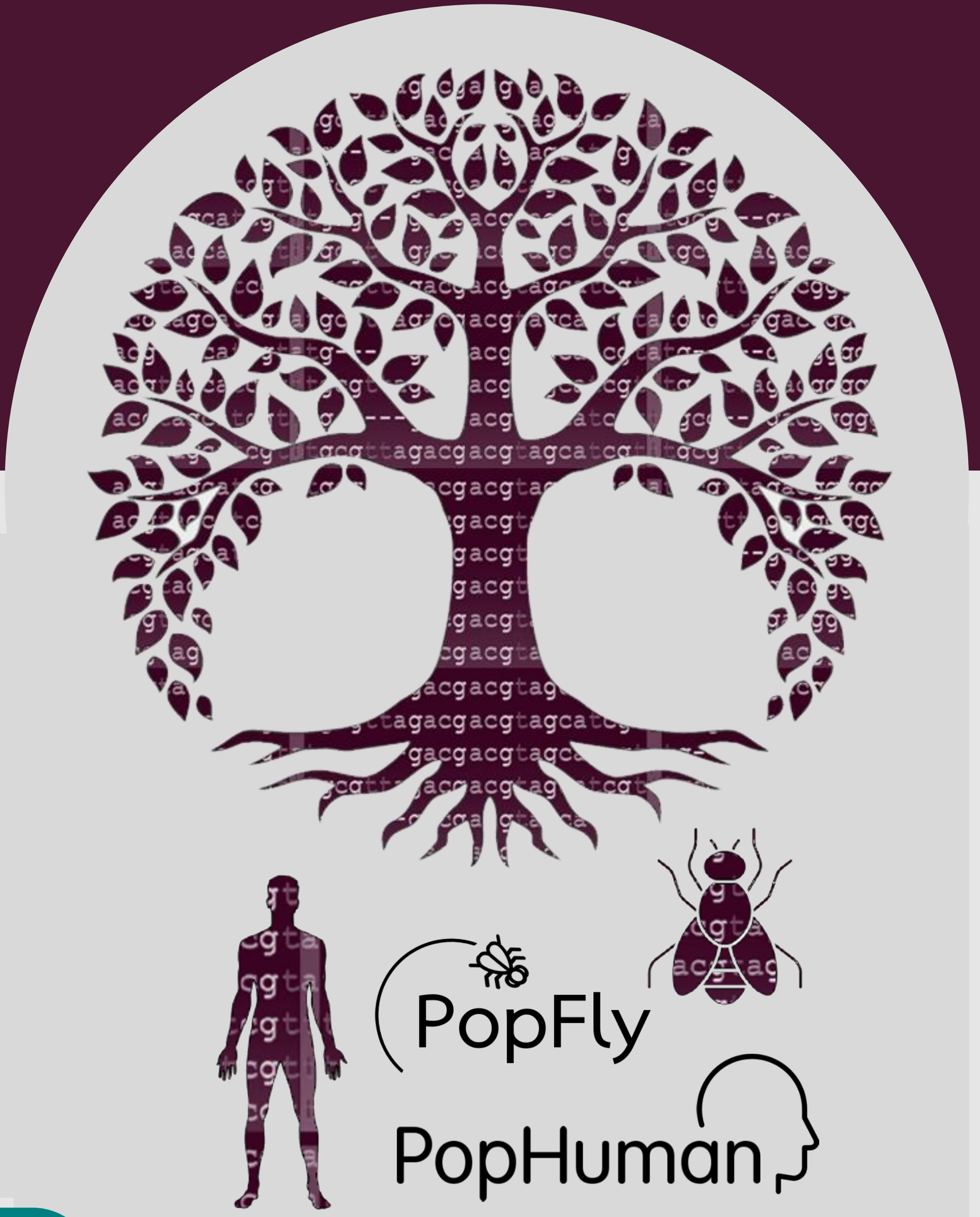
PopLife: A Population Genomics Browser across the Tree of Life

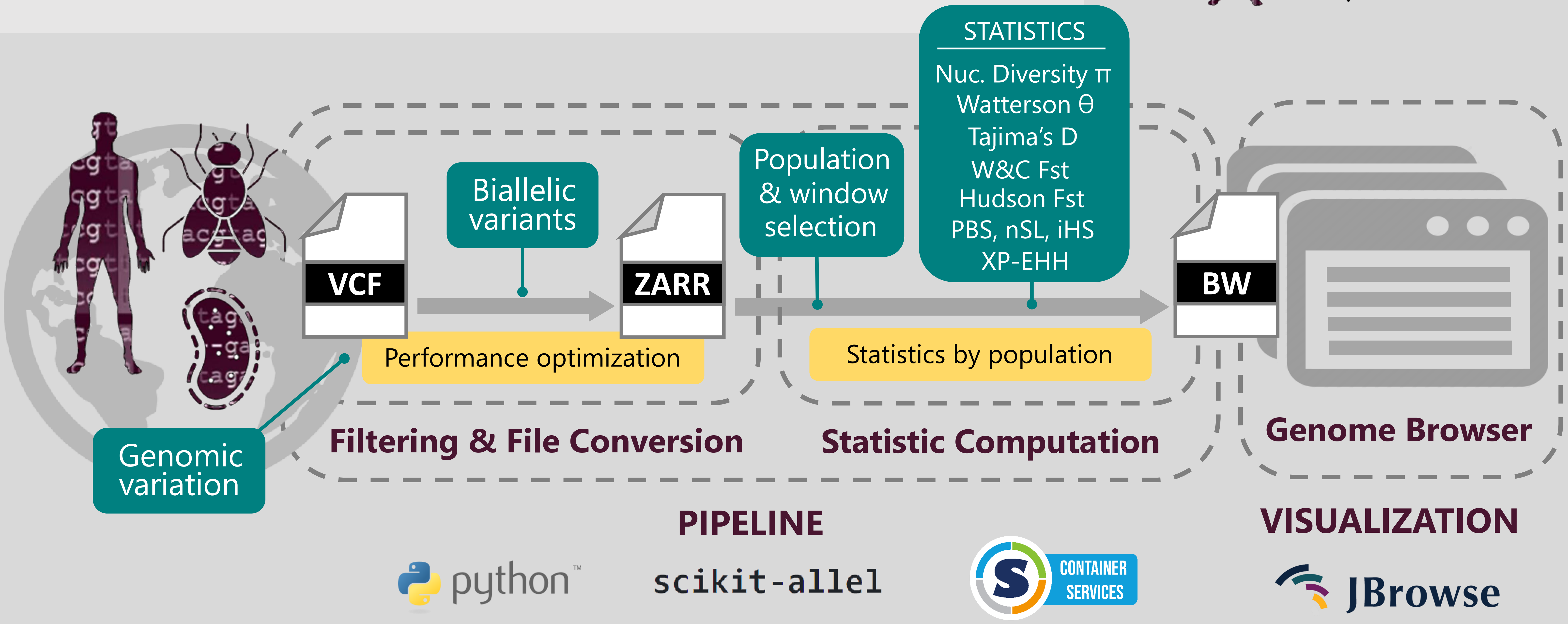
C. Amor-Jimenez, A. Arangua, A. Mompart, S. Casillas, A. Barbadilla



INTRODUCTION

Many genome browsers are currently available, offering comprehensive data on a vast number of genomes. However, they rarely address population genomics information. To fill this gap, we are developing ***PopLife*, a novel genome browser to encompass population genomics data across all species in the Tree of Life**. Leveraging the wealth of data generated by next-generation sequencing technologies, PopLife draws inspiration from our previous successful models PopFly and PopHuman. This browser serves as a valuable tool to describe **population-level variation, investigating genetic adaptation and identifying signals of selection**. Our project aims to facilitate the computation of population genomics statistics at a genome-wide scale through an accessible pipeline, making this information publicly available via a user-friendly interface.

METHODOLOGY



Genomic variation

Filtering & File Conversion

VCF → ZARR

Biallelic variants

Performance optimization

Statistic Computation

Population & window selection

STATISTICS
Nuc. Diversity π
Watterson ϴ
Tajima's D
W&C Fst
Hudson Fst
PBS, nSL, iHS
XP-EHH

Statistics by population

BW

Genome Browser

PIPELINE

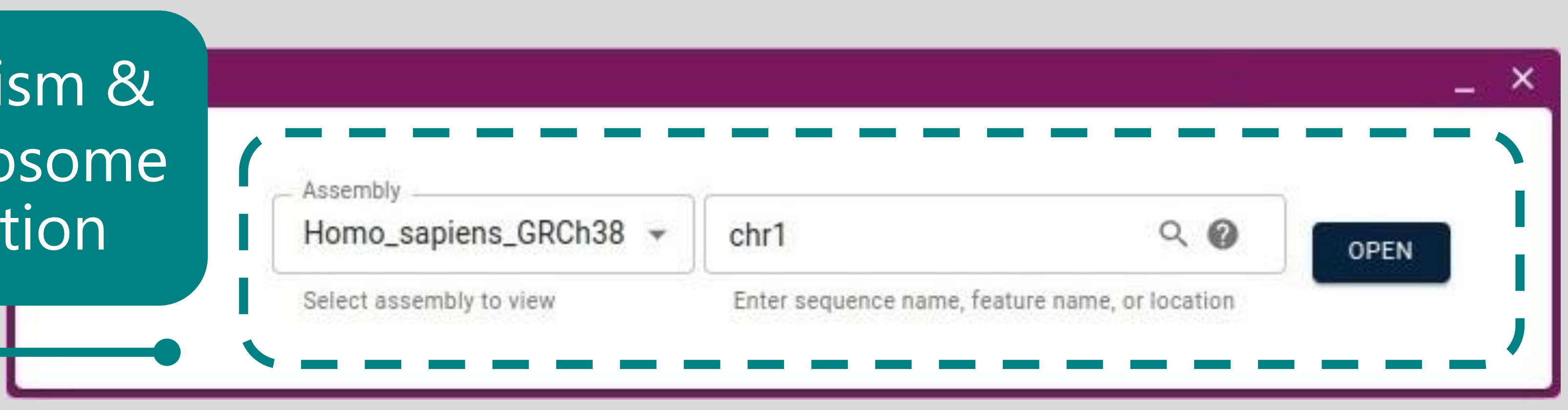
python™ scikit-allel CONTAINER SERVICES

VISUALIZATION

JBrowse

RESULTS

Organism & chromosome selection



Assembly: Homo_sapiens_GRCh38

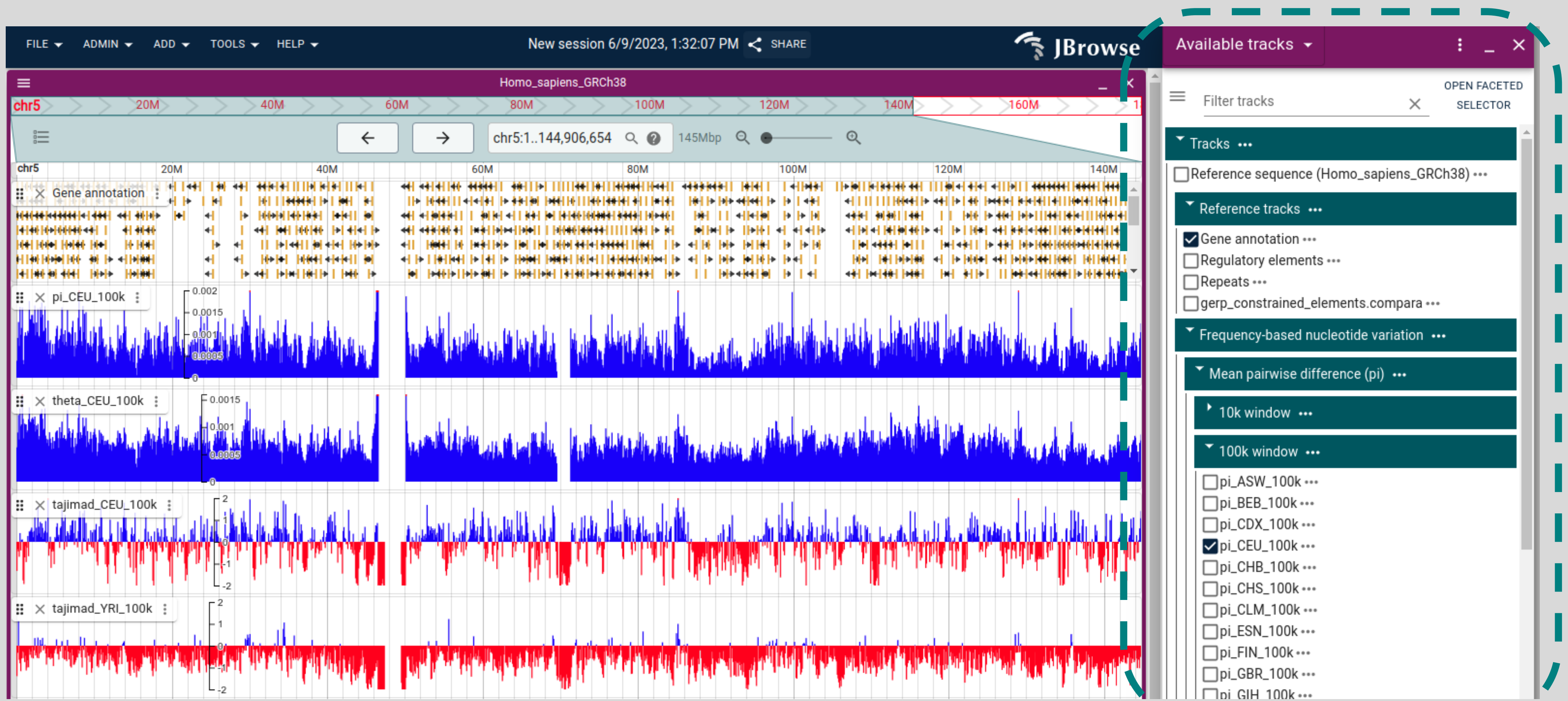
Select assembly to view

chr1

Enter sequence name, feature name, or location

Statistic track selection

- **Modular & Scalable**
- **Containerized**
- **Version independent**
- **Portable**



New session 6/9/2023, 1:32:07 PM

chr5:1..144,906,654 145Mbp

- Tracks
  - ☐ Reference sequence (Homo_sapiens_GRCh38)
  - Reference tracks
    - ☑ Gene annotation
    - ☐ Regulatory elements
    - ☐ Repeats
    - ☐ gerp_constrained_elements.compara
  - Frequency-based nucleotide variation
    - Mean pairwise difference (pi)
      - 10k window
      - 100k window
        - ☐ pi_ASW_100k
        - ☐ pi_BEB_100k
        - ☐ pi_CDX_100k
        - ☑ pi_CEU_100k
        - ☐ pi_CHB_100k
        - ☐ pi_CHS_100k
        - ☐ pi_CLM_100k
        - ☐ pi_ESN_100k
        - ☐ pi_FIN_100k
        - ☐ pi_GBR_100k
        - ☐ pi_GIH_100k

Population genomics statistics are displayed as tracks using **JBrowse 2**, an open-source genome browser with a user-friendly interface and able to handle vast genomic datasets. General annotation tracks have also been incorporated to facilitate comprehensive analysis, including gene annotation, regulatory elements, repetitive regions, and comparative genomics data. Upon selecting the organism and chromosome of interest, tracks can be conveniently picked using the right-side menu. Furthermore, by utilizing multiple linear genomic views, it is possible to compare tracks from different organisms, thereby enabling enhanced comparative analysis.

CONCLUSIONS

- Portable and reliable tool with high usability and dependency management.
- Improved performance in statistic computation at a genome-scale level.
- Highly modular, allowing different configurations in statistics, populations and chromosome selection.

ONGOING WORK

Full pipeline implementation in nextflow

Encompassing a broader range of species

Soon available at **poplife.uab.cat**

References

ibb Institut de Biotecnologia i de Biomedicina
UAB
PIC port d'informació científica
RES RED ESPAÑOLA DE SUPERCOMPUTACIÓN
MINISTERIO DE CIENCIA E INNOVACIÓN