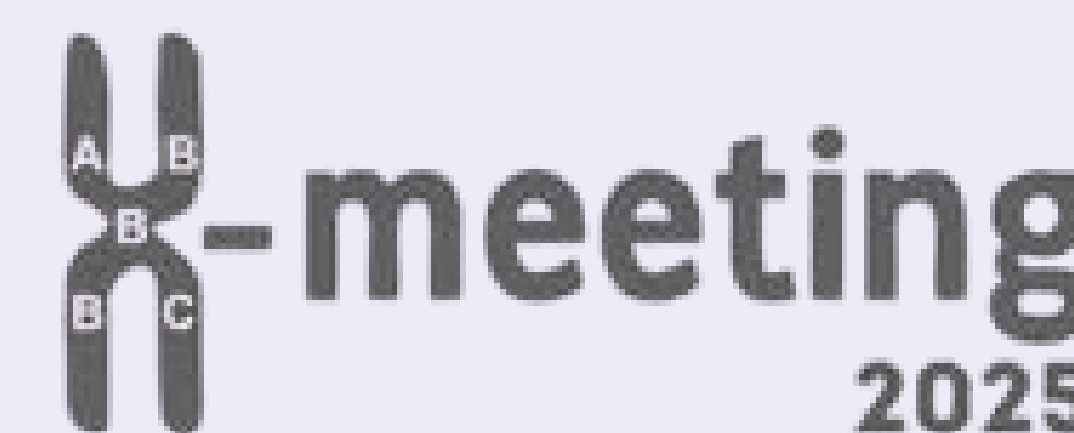


# A Pan-genome of the Saccharum complex



Gustavo Carvalho do Nascimento<sup>1</sup> ; Diego Mauricio Riaño-Pachón<sup>1</sup>

1 Computational, Evolutionary and Systems Biology Laboratory - LabBCES - CENA/USP, Brazil

([gustavocn@usp.br](mailto:gustavocn@usp.br); [diego.riano@cena.usp.br](mailto:diego.riano@cena.usp.br); )

## print("INTRODUCTION")

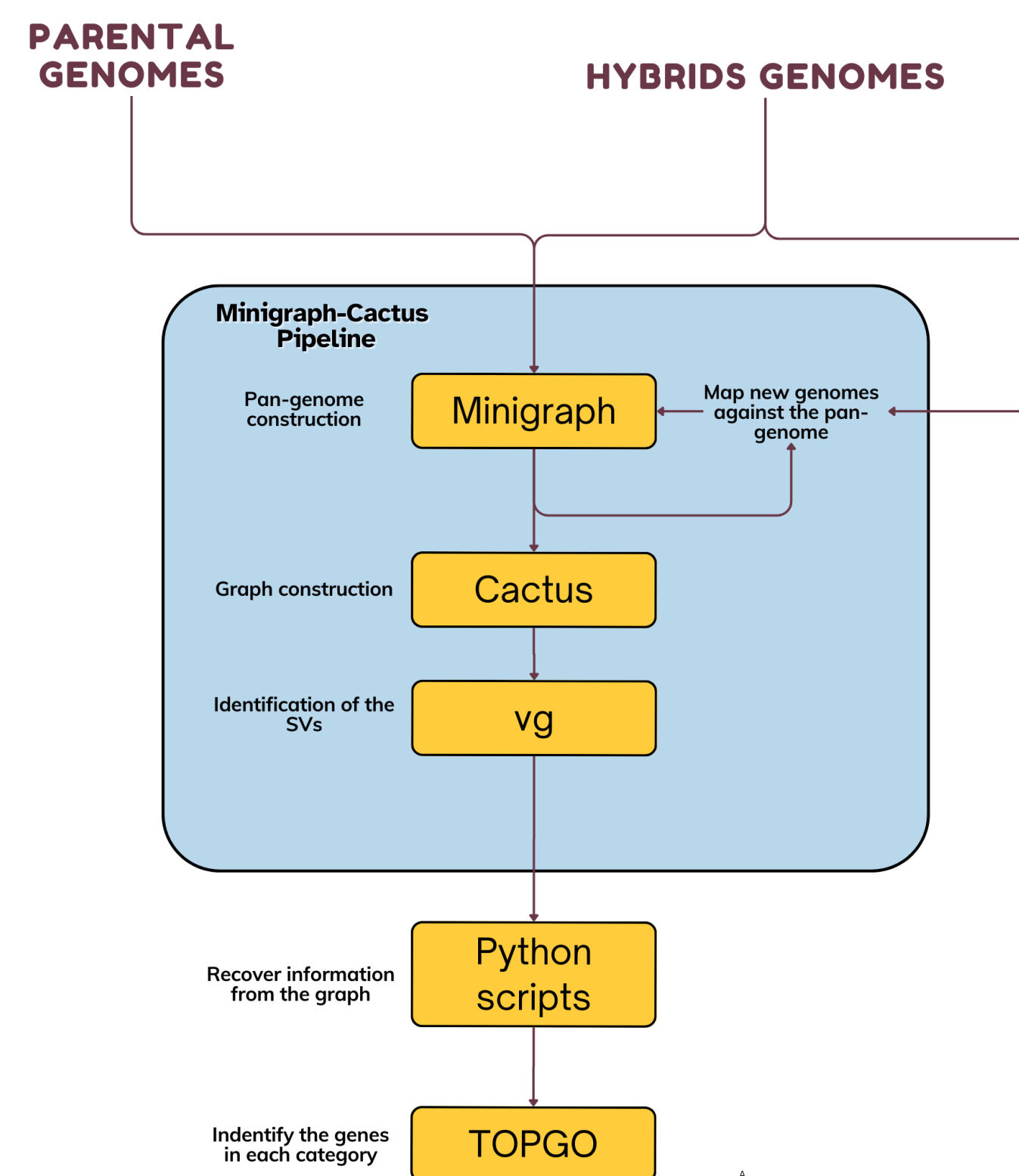
The *Saccharum* complex is a group of species that have contributed to the origin of modern sugarcane. This group includes the genera *Saccharum*, *Erianthus* (specifically, sect. *Ripidium*), *Sclerostachya*, *Narenga*, and *Miscanthus* (sect. *Diantra*). These species are characterized by large genomes and significant chromosomal alterations. Within this complex, chromosome numbers vary considerably. To better understand and represent the genomic diversity in the *Saccharum* complex, we are constructing a pan-genome graph, a way to study the genomic variation and enable a better understanding of the genomic makeup that can be exploited to aid molecular breeding of sugarcane.

## print("METHODS")

We are using genomes from public databases. For the construction of the pangenome graph we are employing the Cactus-Minigraph pipeline, Minigraph for the construction of the SV-only graph, Giraffe for mapping the reads on the graph, Cactus to construct the actual graph that contains variants of all sizes, and vg to find the Structural Variants.

Genome/Variety	A (%)	B (%)	C (%)	D (Gbp)	E (#)	F (Kbp)
LA-Purple SOFFI	99.82	1.21	98.61	6.8	9617	82862
Np-X SSPON	99.86	1.31	98.55	2.7	1033	68640
AP85-441 SSPON	99.75	3.92	95.83	3.1	15303	91359
R570 JGI	99.84	0.47	99.37	5.0	143	79221
KK3	99.81	1.16	98.65	7.0	36193	85005
MLU	97.40	4.82	95.18	2.0	2665	171000
ERUFI	97.27	90.6	9.4	0.9	829	84000

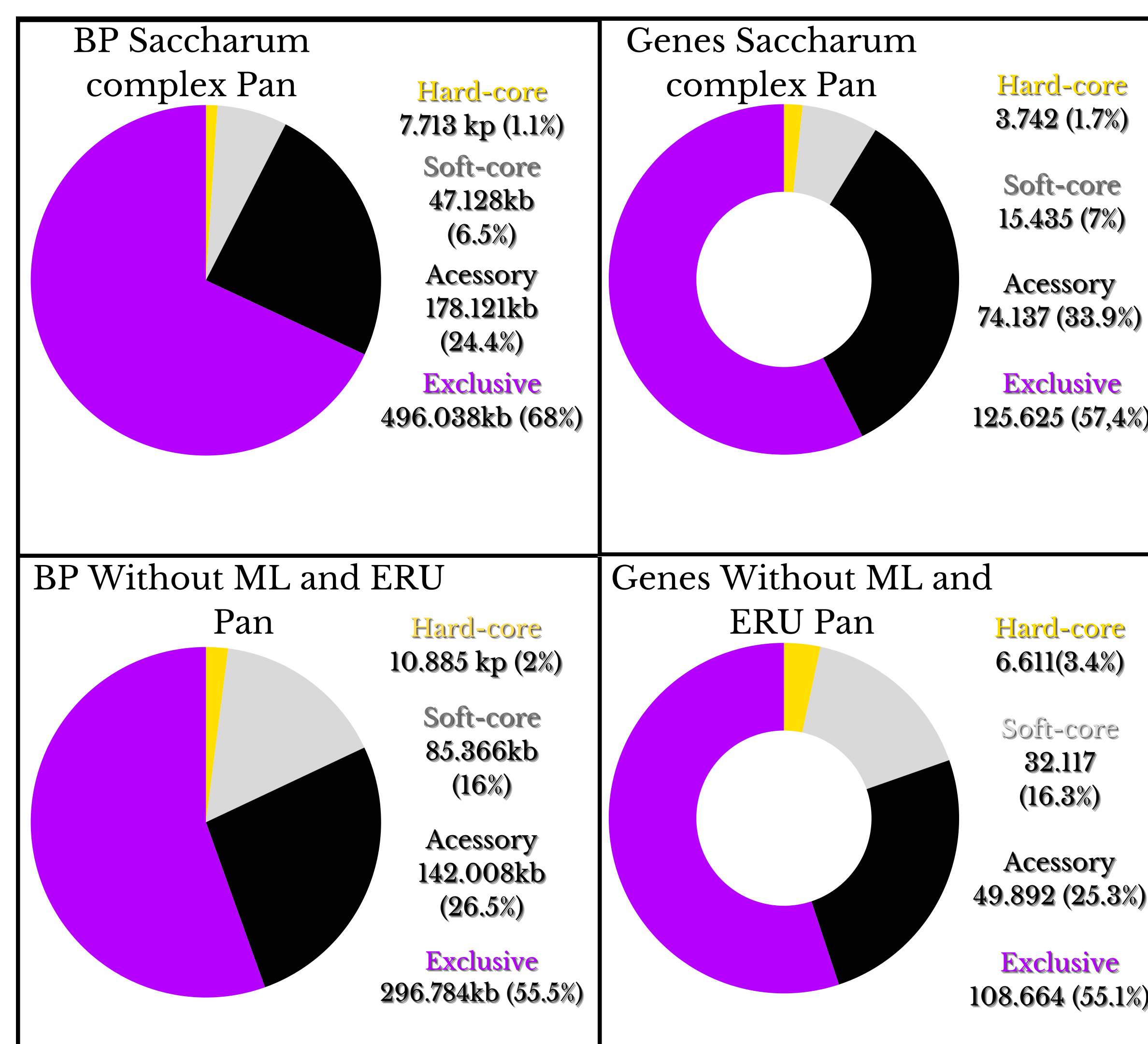
**Table 1.** Statistics of *Saccharum* complex genomes available in this project, including metrics of gene content and contiguity. Gene content metrics were calculated using COMPLEASM with the Poales order database, which has 4896 conserved single-copy genes. A) Complete genes, B) Complete genes in single copy, C) Duplicated complete genes, D) Assembly size, E) Number of contigs, F) N50



**Figure 1.** Illustrative figure of the proposed work plan in this project.

## print("RESULTS")

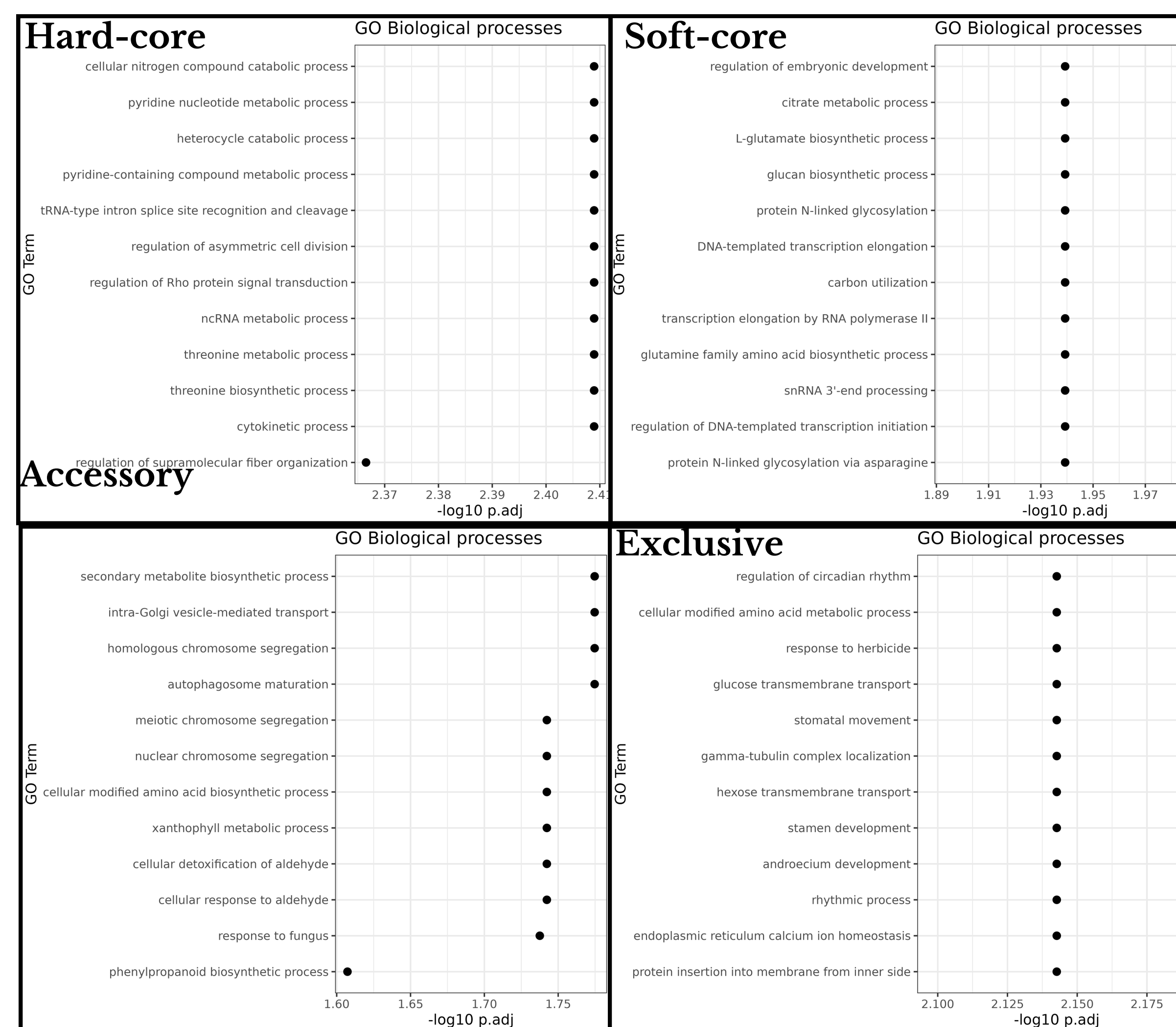
Our current *Saccharum* complex pan-genome graph incorporates seven genomes, collectively representing 729,003,254 base pairs of genetic information. This dataset includes 6,466,956 structural variants (SVs), of which 204,410 are Indels and 6,262,546 are larger structural variations like duplications, inversions, and translocations. We categorized the pan-genome into four classes: **Hard-core** (present in all seven genomes), **Soft-core** (present in >5 genomes), **Accessory** (present in 2–4 genomes), and **Exclusive** (present in only one genome). The proportional size of each category within the total pan-genome was quantified (Figure 2), followed by the identification of genes specific to each category.



**Figure 2.** Figure 2. Pie charts representing the size of each category on the pan-genome and the number of genes. We also compared what would happen if we just used the genera *Saccharum*

Comparing these pan-genomes reveals that the genomic makeup within the *Saccharum* complex is more diverse than previously observed for the *Saccharum* genus alone. This expanded genomic information, which includes elements new to molecular breeding programs, could significantly enhance future efforts.

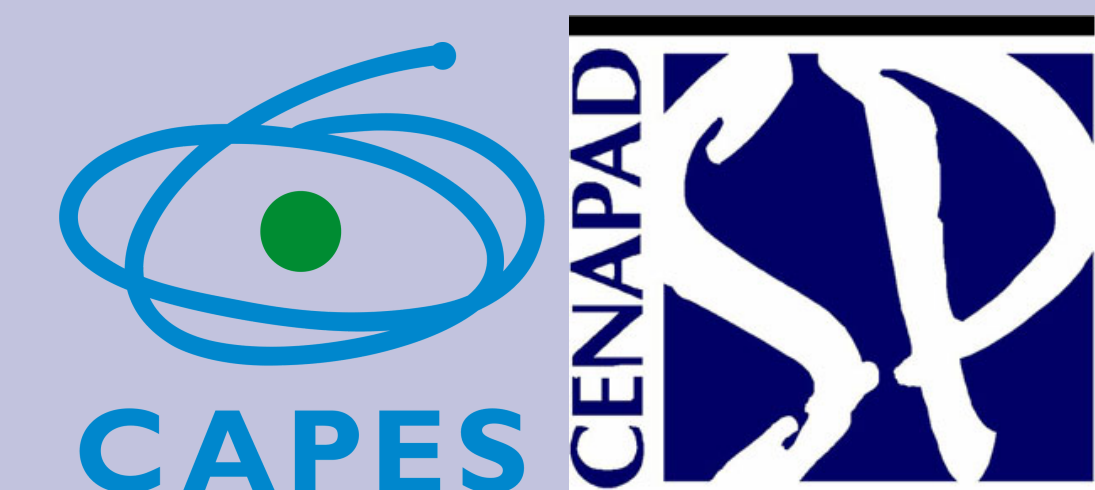
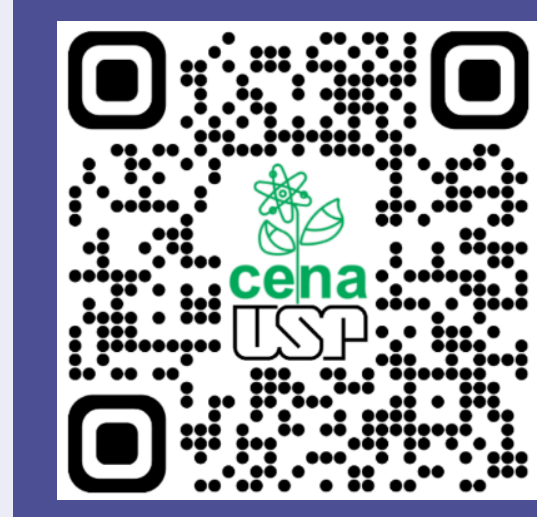
To understand the functional roles of genes in each category, we performed Gene Ontology (GO) enrichment analysis using TOPGO, focusing on Biological Processes. It's important to note that while the number of new genes in the complex may not be high, the overall increase in genomic information is substantial. This includes elements like non-coding RNAs (ncRNAs), transposons, and other genomic features that hold potential for improving future sugarcane hybrids.



**Figure 3.** Gene Ontology (GO) enrichment analysis for genes in each pan-genome category

## print("NEXT STEPS")

We aim to explore which genomic features have been inherited from the ancestral origins of sugarcane and identify valuable genetic information within the broader *Saccharum* complex that can be harnessed to enhance sugarcane crop improvement.



This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Finance Code 001

