Construction and Analysis of the *Moniliophthora roreri* Pangenome

<u>Isabella Gallego Rendón</u>, Diego M. Riaño-Pachón
Computational, Evolutionary and Systems Biology Laboratory
Center for Nuclear Energy in Agriculture, University of São Paulo. Piracicaba, Brazil

GITHUB

LINKEDIN

The cacao pathogen *Moniliophthora roreri* poses a significant threat to global cacao production. This is due to its distinct dispersal ability, its propensity to infect a wide range of cacao cultivars, its adaptability to different ecological niches, and its high pathogenicity. To develop effective control strategies, a deeper understanding of its genetic diversity and functional capabilities is critical. The objective of this study was to perform a pangenomic analysis of 22 publicly available *M. roreri* genomes to gain insights into its genetic composition and potential functional properties.

A total of 456,309 protein-coding genes were identified from the assembled genomes, of which 97.5% were assigned to orthogroups. The pangenome was categorized in hard-core, soft-core, accessory, and exclusive categories. Derivative analysis provided a perspective on the gene pool expansion or contraction patterns in response to the integration of additional genomes. Functional annotation and GO terms enrichment analysis revealed genes associated with various biological processes, these processes were described underscoring the potential genes associated to pathogenic and adaptive mechanisms of *M. roreri*.
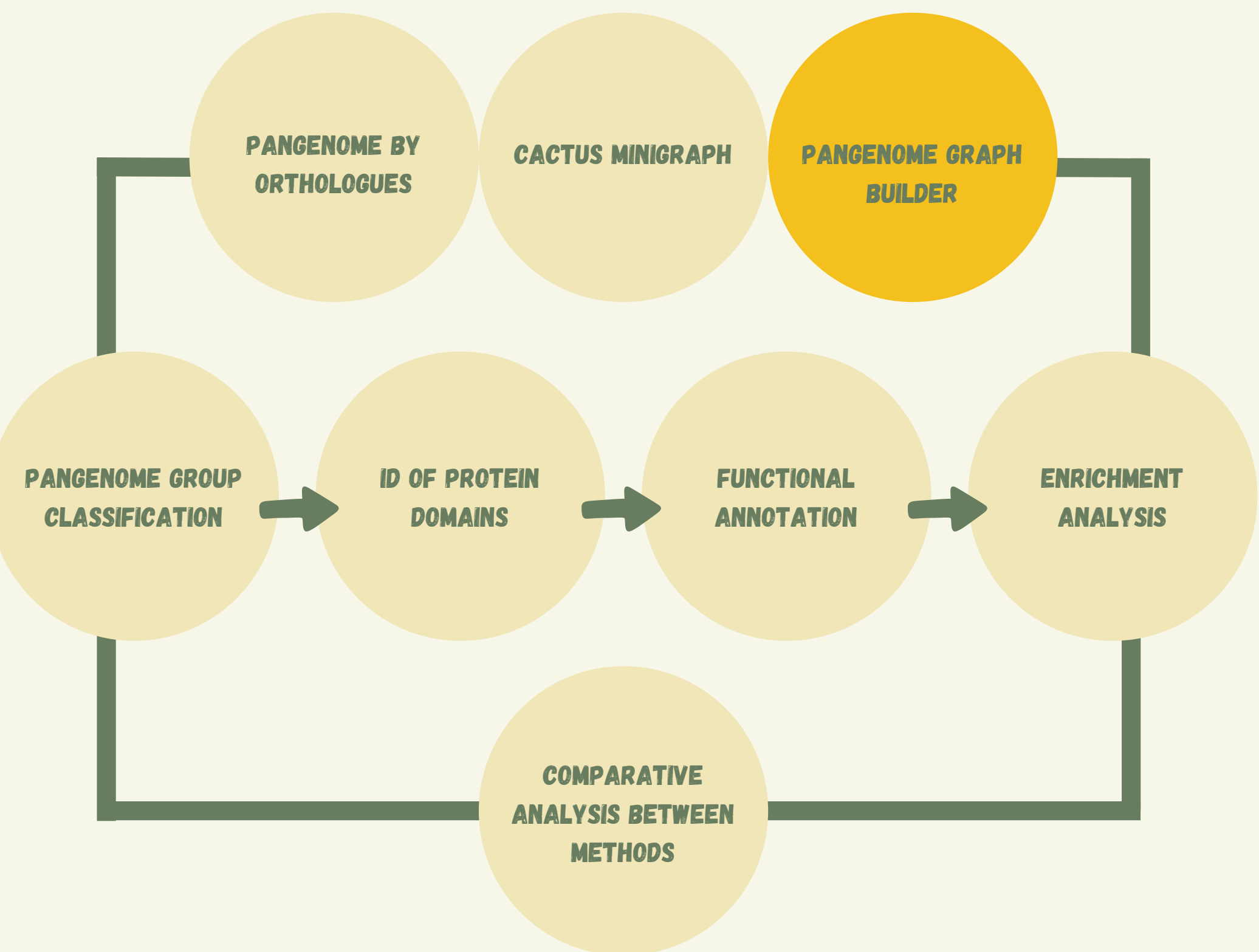
This comprehensive pangenomic study provides a fundamental understanding of the genetic and functional makeup of *M. roreri*. The insights elucidate potential gene clusters that might be of interest for future research in the field of plant-pathogen interactions and targeting interventions for cacao disease control.

Cacao Global Distribution

METHODS

PANGENOME BY ORTHOLOGUES
CACTUS MINIGRAPH
PANGENOME GRAPH BUILDER

PANGENOME GROUP CLASSIFICATION → ID OF PROTEIN DOMAINS → FUNCTIONAL ANNOTATION → ENRICHMENT ANALYSIS

COMPARATIVE ANALYSIS BETWEEN METHODS

SIDE QUESTS

GENOME ANNOTATION

QUALITY ASSESSMENT

DEG ANALYSIS

ML Strategies

DATASET (Minio et al., 2024)

M. roreri: A Fungal Threat to Cacao

Cacao (*Theobroma cacao*), a key crop in the global chocolate industry, is cultivated across 61 tropical countries, with Colombia ranking as the fourth-largest producer in Latin America. Despite its economic significance, cacao production faces challenges due to pests and diseases, notably frosty pod rot, caused by the fungus *Moniliophthora roreri*. This pathogen is highly adaptable, with a hemibiotrophic life cycle, and thrives in humid, tropical environments. It can devastate crops, causing yield losses of up to 80%. *M. roreri*'s genetic diversity and ability to infect multiple cacao varieties underscore the need for advanced genomic research to develop effective control strategies. Recent genomic studies on *M. roreri* have explored its genetic composition and evolutionary potential. Two genomes have been sequenced, revealing significant genetic variation and aiding in understanding its pathogenicity. A pan-genome approach, involving 22 genomes, highlights the extensive genetic diversity within the fungus, providing insights into its adaptability and helping develop solutions to combat its impact on cacao cultivation. This research is crucial for mitigating the spread of the disease, which threatens not only local economies but also the global cacao supply chain.

PANGENOME STATS

ORTHOLOGUE BASED

Moniliophthora roreri genomic content statistics evaluated in 5 isolates

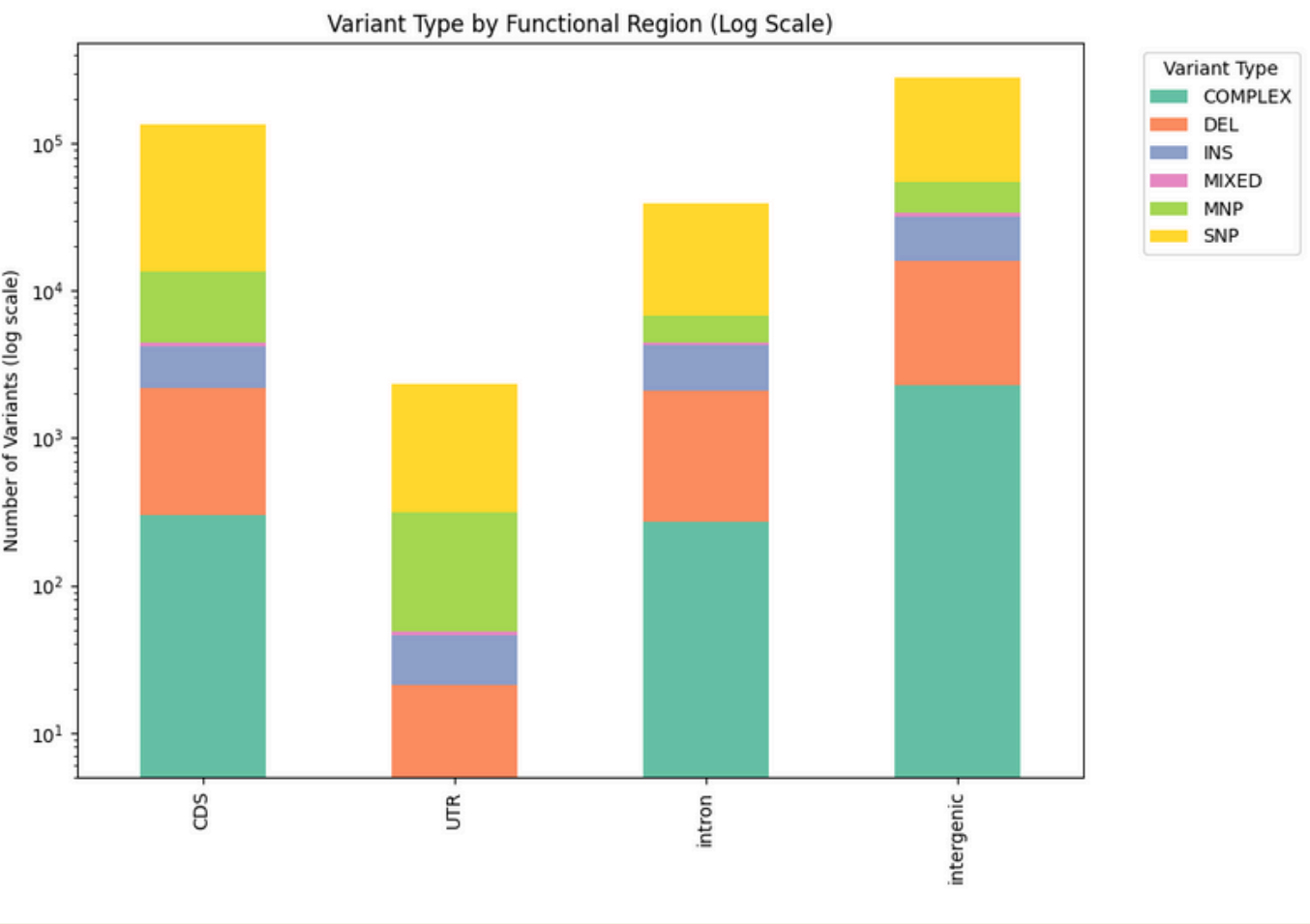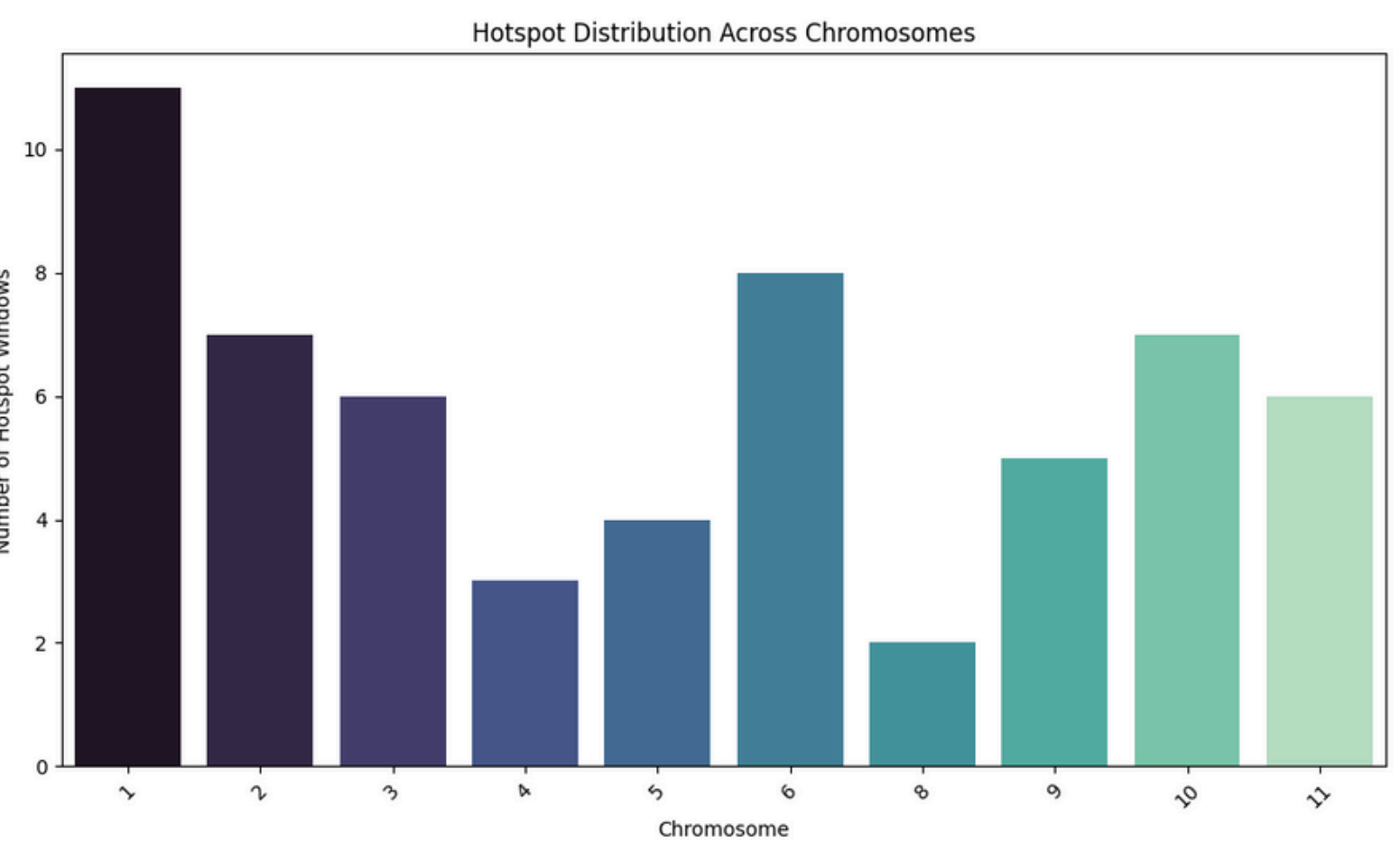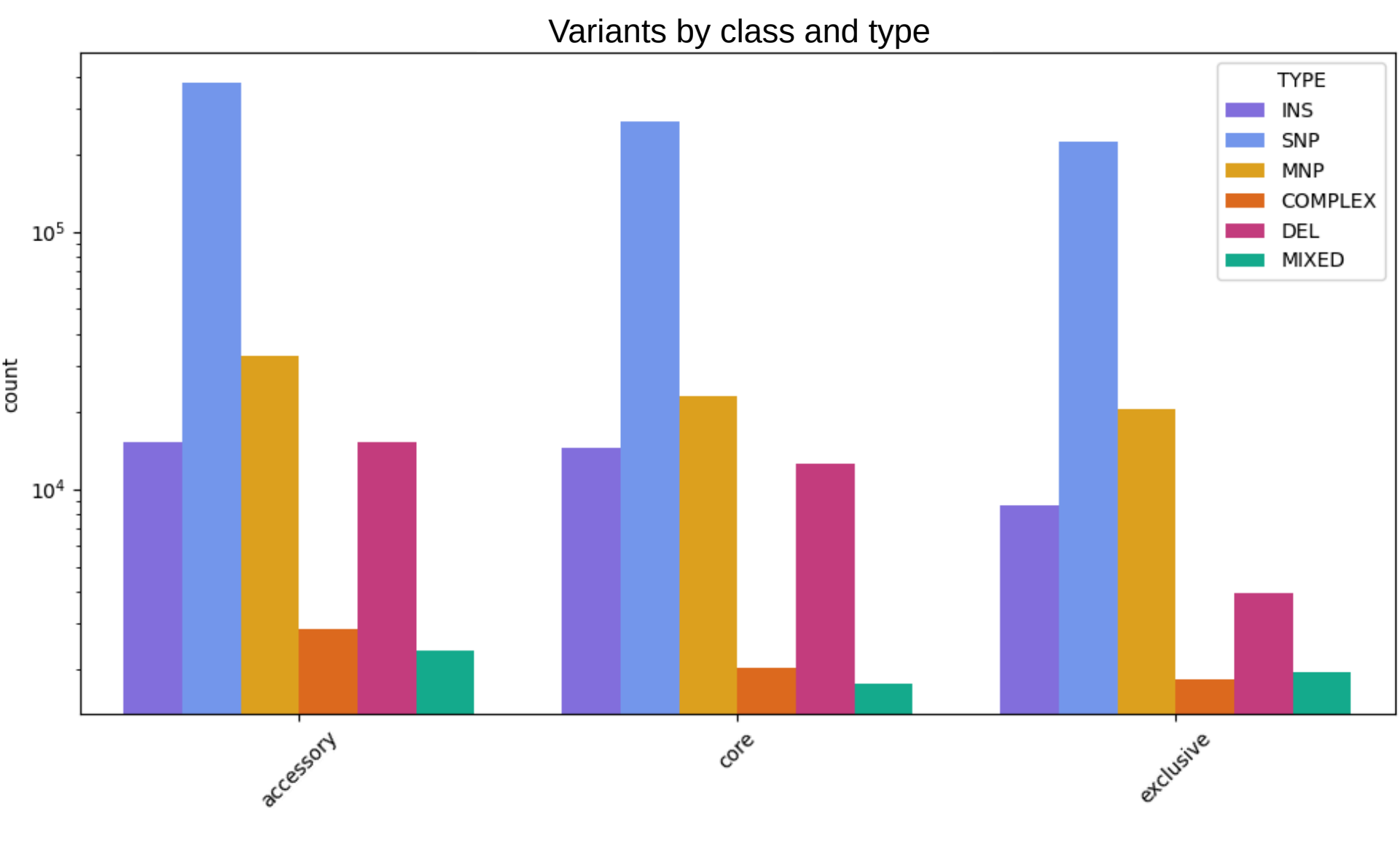| Statistic	Value
Number of genes present in pangenome	104016
Number of orthogroups	19831
Number of genes present in orthogroups	103487
Number of core groups	16186
Number of genes present in core groups	89747
Number of groups present in 90% of the species	16186
Number of genes present in soft-core groups	89747
Number of accessory groups	3570
Number of genes present in accessory groups	13004
Number of exclusive groups	75
Number of genes present in exclusive groups	736

Hard-core 69 %
Soft-core 75 %
Accessory 24.8 %
Exclusive 0.2 %
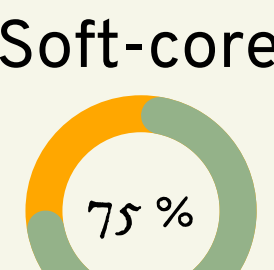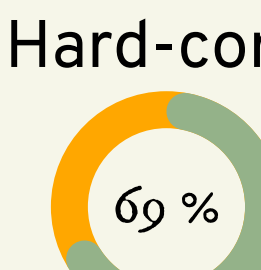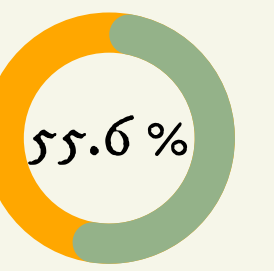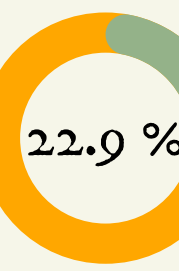
GRAPH BASED

Moniliophthora roreri segment statistics evaluated in 5 isolates

Statistic	Value
Number of segments	1450990
Total lenght (bp)	68883584
Number of core segments	332643
Number of accessory segments	806115
Number of exclusive segments	312232

Core 22.9 %
Accessory 55.6 %
Exclusive 21.5 %

ENRICHMENT ANALYSIS

Exclusive
endomembrane system organization
nicotinamide nucleotide biosynthetic pro...
DNA metabolic process
macromolecule biosynthetic process
nicotinamide nucleotide metabolic proces...
alpha-amino acid catabolic process
non-proteinogenic amino acid catabolic p...
phosphorus metabolic process
cellular macromolecule localization
pyridine-containing compound biosyntheti...
carboxylic acid catabolic process
nucleobase-containing compound metabolic...
cellular catabolic process
RNA processing
mRNA processing
regulation of biological process
nitrogen compound transport

Accessory
intracellular monoatomic cation homeosta...
nucleobase transport
ceramide metabolic process
monoatomic ion transmembrane transport
regulation of actin filament organization
antibiotic biosynthetic process
DNA strand elongation
intracellular protein transport
postreplication repair
L-amino acid catabolic process
carbon catabolite repression
regulation of DNA replication
monoatomic anion transmembrane transport
DNA strand elongation involved in DNA re...
viral genome replication
unsaturated fatty acid biosynthetic proc...
establishment of protein localization to...
nucleotide biosynthetic process

Core
sodium ion transport
asparagine metabolic process
negative regulation of carbohydrate meta...
cellular response to osmotic stress
meiosis I cell cycle process
assembly of actomyosin apparatus involve...
eisosome assembly
glycyl-tRNA aminoacylation
developmental process
regulation of cell cycle G1/S phase tran...
phosphatidylinositol-mediated signaling
tRNA wobble uridine modification
IMP metabolic process
dicarboxylic acid biosynthetic process
cytoplasmic microtubule organization
ethanol catabolic process
mRNA splicing, via spliceosome
regulation of NADP metabolic process
negative regulation of DNA metabolic pro...
aggraphagy

GO:0006259 DNA Metabolic Process

GO:0019079 Viral Genome Replication

GO:0006672 Ceramide Metabolic Process

GO:0170035 L-amino acid Catabolic Process

GO:0017000 Antibiotic Biosynthetic Process

GO:2000241 DNA Recombination

Class
Cluster

Class
accessory
core
exclusive

Cluster
Cytoskeletal Organization
DNA/RNA Processing
Metabolism & Biosynthesis
Regulation & Development
Transport & Localization

Semantic Similarity
0.5
0

VARIANT ANALYSIS

Variants by class and type
count
TYPE: INS, SNP, MNP, COMPLEX, DEL, MIXED

Hotspot Distribution Across Chromosomes
Number of Hotspot Windows
Chromosome

Variant Type by Functional Region (Log Scale)
Number of Variants (log scale)
Variant Type: COMPLEX, DEL, INS, MIXED, MNP, SNP

WHAT'S NEXT?

- The pangenome will be expanded using the whole dataset (Pacbio + Illumina).
- A new pangenome graph will be constructed using PGGB in order to compare tools.
- The potential functional implications of genomic variations will be explored, and selective pressures present in the gene groups categorized as accessory, exclusive, and core will be evaluated.
- An effector prediction and analysis step will be implemented.
- Predictive ML and DL models will be trained using the pangenome classes and enriched GO terms as features in order to detect possible DEGs of interest.
- Future research will focus on further analyzing the accessory and exclusive genome of *Moniliophthora roreri* to understand its role in host range and adaptability, helping identify the pathogen's ability to infect alternative hosts.

FAPESP · CNPq · Research Centre for Greenhouse Gas Innovation · USP · cena · BCE · CENAPAD · INCT BIOTANOL