Construction and Analysis of the *Moniliophthora roreri* Pangenome

Isabella Gallego Rendón, 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 into 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.*

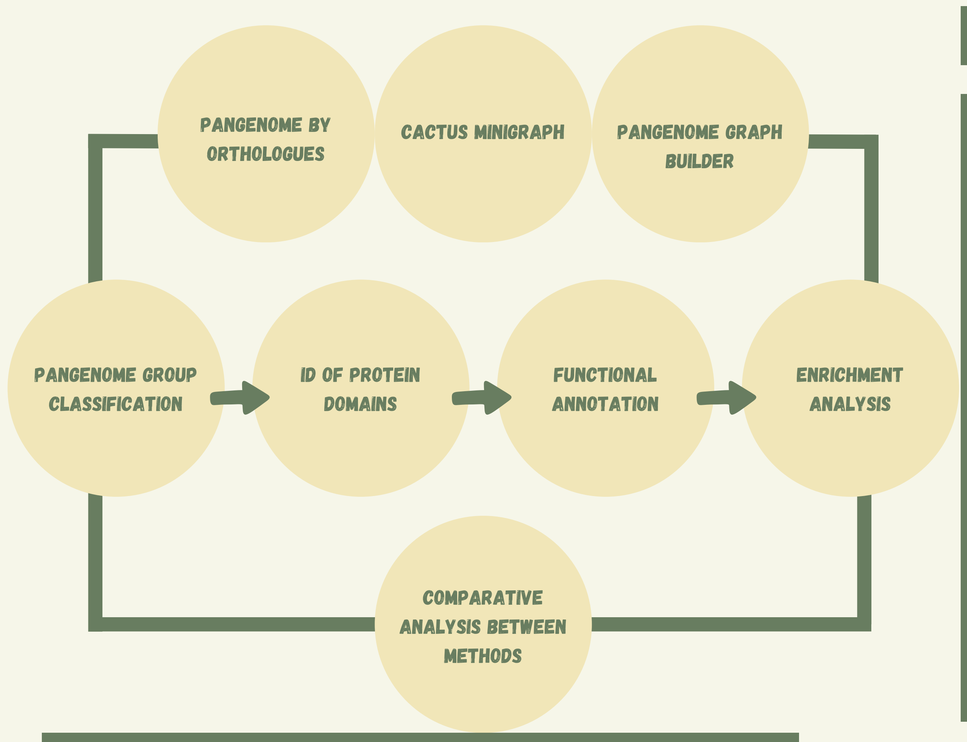
This comprehensive pangenomic study provides a fundamental understanding of the genetic and functional makeup of *M. roreri.* The insights elucidate potential gene clusters that might be of interest for future research in the field of plant-pathogen interactions and targeting interventions for cacao disease control.

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.

METHODS



SIDE QUESTS

- GENOME ANNOTATION
- QUALITY ASSESSMENT
- DEG ANALYSIS

ML Strategies

DEG ANALYSIS USING MACHINE LEARNING

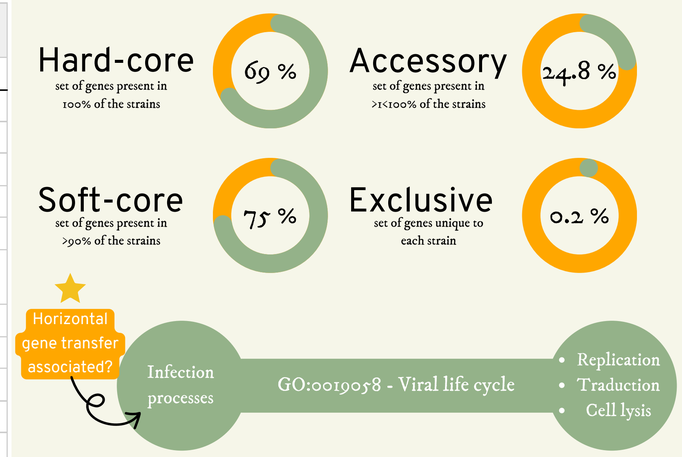
Conditions evaluated:
- 30 dpi (biotrophic phase)
- 60 dpi (necrotrophic phase)

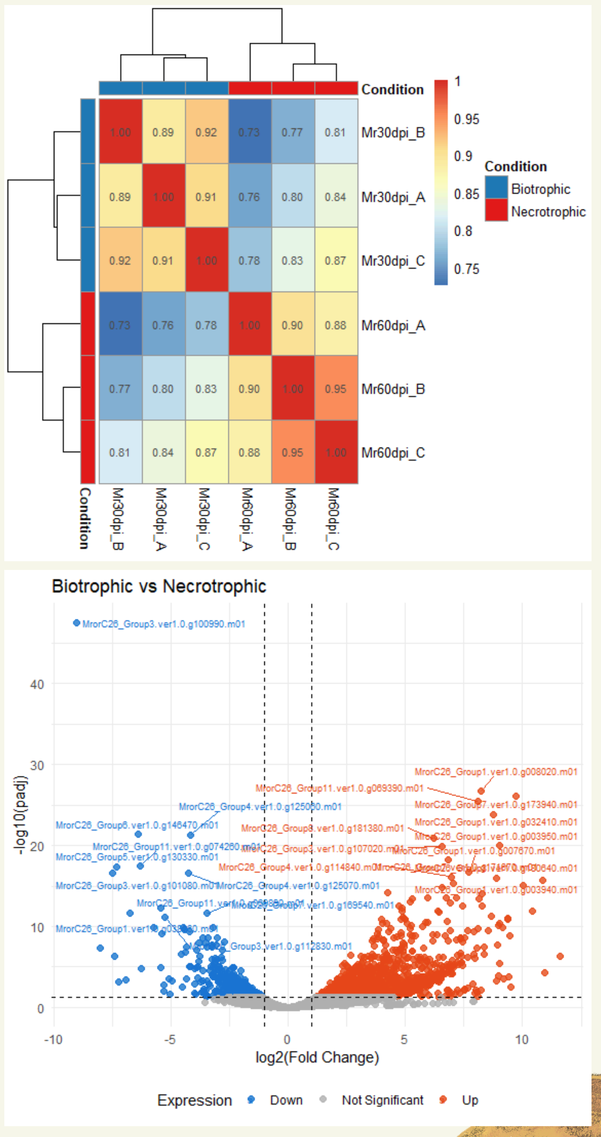
Algorithms tested:
- Random Forest
- Logistic Regression
- SVM
- KNN

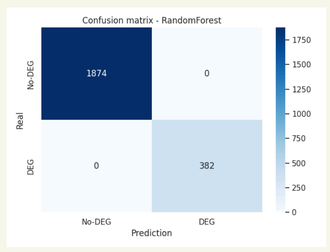
PRELIMINARY RESULTS

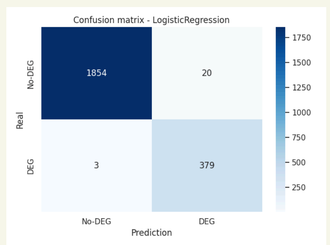
Moniliophthora roreri genomic content statistics evaluated in 22 isolates

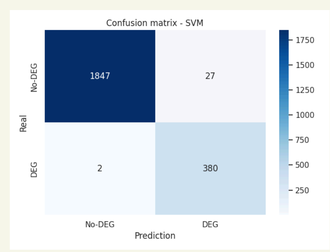
Statistic	Value
Number of genes present in pangenome	456309
Number of orthogroups	28215
Number of genes present in orthogroups	444887
Number of core groups	13908
Number of proteins present in core groups	307081
Number of groups present in 90% of the species (19.8)	15145
Number of proteins present in soft-core groups	333880
Number of accessory groups	12821
Number of proteins present in accessory groups	110185
Number of exclusive groups	249
Number of proteins present in exclusive groups	822

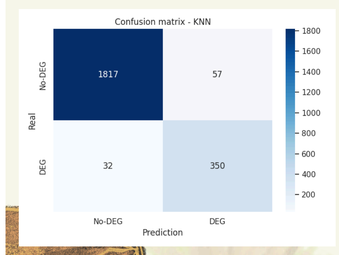


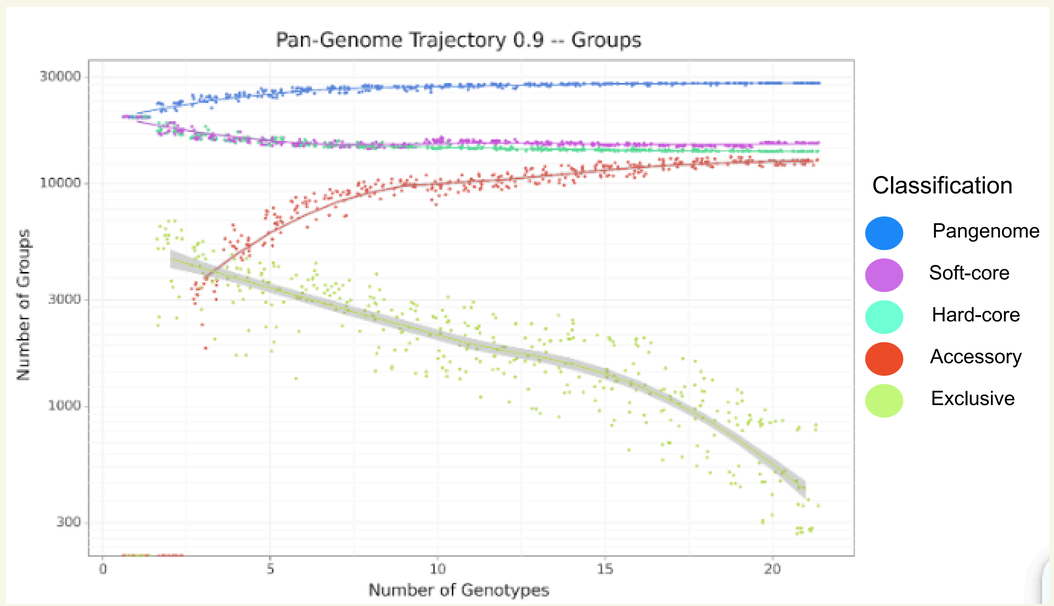












WHAT'S NEXT?

- The pangenome will be expanded using the previously missing genomic data.
- The graph methods will be implemented in order to construct the pangenomes that will be used to compare the building strategies.
- 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.
- The predictive ML model will be expanded using public data regarding a broader range of infection stages (7 & 21 dpi). Also, the model will be adapted in order to predict how DEGs change over the infection stages (temporal model).
- 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 SP INCT BIOETANOL