

# Analysis of comparative genomics reveal evidence of positive selection on pathogenicity-related genes of Witches' broom disease on cocoa trees

Paulo Tokimatu<sup>1</sup>, Marcelo F. Carazzolle<sup>1</sup>, Paulo J. P. L. Teixeira<sup>2</sup>, Daniela P. T. Thomazella<sup>3</sup>, Leandro C. Nascimento<sup>1</sup>, Gonçalo G. A. Pereira<sup>1</sup>, Juliana José<sup>1</sup>

(1) Genomic and Expression Laboratory; Institute of Biology; UNICAMP; Brazil, (2) Department of Biology; University of North Carolina, USA, (3) Department of Plant & Microbial Biology; UC Berkeley; USA

*Financial support provided by CAPES and FAPESP.*

The basidiomycete fungus *Moniliophthora perniciosa* is responsible for the Witches' broom disease, infecting cocoa trees (*Theobroma cacao*) and causing great damage on the cocoa seed industry, specially in Brazil. It is currently known 3 biotypes which can infect different hosts. The C-biotype is the most studied because it infects cocoa trees. The S-biotype infects Solanaceae plants like tomato, causing symptoms similar to the disease caused by the C-biotype. The L-biotype is found in species from the Bignoniaceae group but differently from the two other biotypes, it presents an asymptomatic infection, behaving more like an endophytic fungus. For this work we sequenced the genomes from 18 samples of *M. perniciosa* (8 samples from C-biotype, 7 from S-biotype, 2 from L-biotype and *M. roreri* as the outgroup) to investigate the evolutionary history of pathogenic genes potentially related to the difference in the host range of each biotype. After assembling, predicting and annotating our samples, comparative analysis using features like SNP counting or size of gene families showed that both nuclear and mitochondrial genomes are highly variable but have constant differences among the biotypes. It is likely that depending on their function on the infection, some pathogenic genes may suffer differential selection pressure among host species. This work focuses on the phylogenetic reconstruction, analysis of the evolutionary rate and testing for positive selection models in genes which are known to have a pathogenic role on the infection of the C-biotype on cocoa trees. We found that the genes cerato-platanin (MpCP), pathogenesis-related-1 (MpPR-1) and necrosis and ethylene-inducing proteins (MpNEP) appear in different number of copies for each biotype and that, by comparing their gene trees with a conserved tree among biotypes, are good candidates to deeper analysis. These pathogenic-related gene families showed different trees from the observed in the conserved tree, suggesting that selective processes may have act differentially on the evolution of the genes. At least one group of orthologs among biotypes showed higher divergence between L-biotype and the C- and S-samples, but a group of MpNEPs recent paralog in C-biotype and a MpPR-1 cluster shared among biotypes presented a specially higher evolutionary rate suggesting the action of positive selection. Proteins with evidences of selection might be fundamental in the process of arms-race between the host plant and the pathogen and their role can be further investigated as potential targets for the development of new technologies to contain the disease.