6th INTERNATIONAL CONFERENCE OF THE BRAZILIAN ASSOCIATION FOR BIOINFORMATICS AND COMPUTATIONAL BIOLOGY

Topic: **Genomics, Evolution and Phylogeny**
PI: **GEP06**

THE GENOME PROJECTS OF THE CLOSELY RELATED CACAO PATHOGENS MONILIOPHTHORA RORERI AND MONILIOPHTHORA PERNICIOSA

Carazzolle M F[1,2], Costa G G L[1], Herai R H[1,3], Junior O R[1], Nascimento L C[1], Teixeira P J[1], Tiburcio R A[1], Mondego J M C[4], Pereira G A G[1]

[1] *Laboratório de Genômica e Expressão, Instituto de Biologia, UNICAMP*
[2] *Centro Nacional de Processamento de Alto Desempenho em São Paulo (CENAPAD-SP), UNICAMP*
[3] *Laboratório de Bioinformática Aplicada (LBA), Embrapa Informática Agropecuária - EMBRAPA*
[4] *Instituto Agronômico de Campinas (IAC)*

The basidiomycetes *Moniliophthora roreri* and *Moniliophthora perniciosa* are the etiologic agents of the two most devastating diseases in cacao (*Theobroma cacao*): frosty pod rot and the witches' broom, respectively. The species are very closely related and even hybrid cells have been previously obtained. In cacao, both species infect pods, causing necrosis, and *M. perniciosa* is also able to invade other tissues causing changes in plant metabolisms, such as hyperplasia and hypertrophy. In order to understand the molecular basis of these organisms, we sequenced the genome of these two species. We also obtained transcriptomic data (RNA-seq) in several different conditions including the interaction between cacao and both pathogens. This work reports a pipeline utilized to assembly and compare these genomes, and to identify differential gene expression between libraries.

The genomic and transcriptomic sequences for both organisms were obtained using high-throughput sequencing technology (454/Roche and Solexa/Illumina). The Solexa and 454 reads were assembled into longer contigs using *de novo* assembler Velvet and Newbler, respectively. The hybrid assembly was performed through the combination of Solexa contigs and 454 contigs resulting in final hybrid contigs using a pipeline developed in our group. *Ab initio* and comparative gene predictions were obtained with Augustus and Exonerate programs using previous training set (including RNA-seq data) and the closely related organisms, respectively. The final set of gene models were obtained through a union between *ab-initio* and comparative approaches. The identification of gene expressed in both pathogens, for each library, was obtained mapping the RNA-seq reads into the predicted gene models. In order to identify the gene expressed in *T. cacao* , the RNA-seq reads were mapped into transcriptome assembly of T. cacao. The transcriptome assembly was obtained using around 160.000 *T. cacao* ESTs available at NCBI and assembled by CAP3 program. The analysis of differential expressed genes between libraries was performed using DEG-seq package.

The *M. perniciosa* genome project (www.lge.ibi.unicamp.br/vassoura) and *M. roreri* genome project (www.lge.ibi.unicamp.br/roreri) involve several Brazilian and international laboratories. The comparative genomics in the structure level, gene content level, and orthologous gene sequence can provide new insights that can help this community to increase efforts in cacao diseases investigation.