Genome analysis of *E. nigrum* and other filamentous fungi reveals molecular mechanisms related to endophytic/pathogenic lifestyles

Almir J. Ferreira[1,2], Liliane S. Oliveira[2], João M. P. Alves[2], Michael Thon[3], Alan M. Durham[4], Léia C. L. Fávaro[1], Arthur Gruber[2]* and Welington L. Araújo[1]*

[1]*Dept. of Microbiology and* [2]*Dept. of Parasitology, Institute of Biomedical Sciences, USP, São Paulo, Brazil;* [3]*Instituto Hispano-Luso de Investigaciones Agrarias (CIALE), University of Salamanca, Spain.* [4]*Dept. of Computer Sciences, Institute of Mathematics and Statistics, USP, São Paulo, Brazil.*

*Correspondence: argruber@usp.br and wlaraujo@usp.br

Proper knowledge of the genomes from the endophytic fungus *Epicoccum nigrum* and plants may contribute to improve agricultural production. The molecular mechanisms determining the pathogenic and endophytic lifestyles of fungi are not fully understood. *E. nigrum* is an endophytic fungus that has been used for plant pathogen biocontrol in different host plants, since it produces a series of secondary metabolites of biotechnological interest, including antimicrobials. We have previously determined that the *E. nigrum* isolate P16 produces secondary metabolites with antimicrobial activity. In this work, we confirm this activity by specific assays and report the genome sequencing and annotation of this isolate. We sequenced the whole genome and the transcriptome of the mycelium using the 454 and SOLiD platforms, respectively. The genome was assembled and gene prediction was performed with the MAKER package, using a dataset of Dothideomycetes proteins and *E. nigrum* transcript sequences as evidence. The genome sequence and gene predictions were submitted to a comprehensive functional annotation pipeline using the EGene2 platform. In order to identify gene clusters associated with secondary metabolites and compare their occurrence across endophytic and pathogenic fungi, we analyzed our genome sequence, together with the genomes from 12 different fungi using the AntiSmash server. Finally, the results were analyzed with Synteny Clusters, a specific tool developed by our group that compares gene clusters associated with secondary metabolite biosynthesis in different organisms. We found 10,320 protein coding genes. Based on structural features, *E. nigrum* presents a genome very similar to closely related filamentous fungi, including some with distinct lifestyles. We identified a total of 38 secondary metabolite gene clusters. The comparative analysis across different fungi revealed three clusters restricted to most of the pathogenic fungi, but absent in *E. nigrum*. These clusters are related to plant diseases and antibiotic activity and may be part of the gene repertoire required for a pathogenic lifestyle. In fact, data from the literature seems to corroborate the importance of some of these genes in pathogenicity. This result suggests that the lifestyle differences observed between endophytic and pathogenic fungi might rely on a relatively low number of genes. This conclusion is in agreement with a phylogenetic analysis using seven protein sequences from endophytic and pathogenic fungi, which revealed a close relationship between *E. nigrum*, an endophyte, and *Didymella exigua*, a phytopathogenic fungus.

Support: FAPESP, CNPq and CAPES.