

# Building LeifDB: a database for storing information about genome comparative analysis of the *Leifsonia xyli*

Pedro Bittencourt, Lucas Taniguti, Claudia Monteiro-Vitorello, Saul Leite, Wagner Arbex, Fernanda Almeida

Federal University of Juiz de Fora, ESALQ/USP

## Abstract

In Brazil, sugarcane culture is used for the production of sugar and ethanol. However, there are diseases that affect this culture and cause considerable reduction in productivity. Amongst them is the ratoon stunting disease caused by the *Leifsonia xyli* subsp. *xyli* bacterium, which is one of the most onerous to the sugarcane sector. There are many species that compose the *Leifsonia* genus and only of the them is plant pathogenic, the others are free-living bacteria. This genus comprises seven species that have been isolated from distinct niches, such as plants, soil, distilled water, and an Antarctic pond. *Leifsonia xyli* is the only species causing plant disease and comprises of two subspecies: *L. xyli* subsp. *cynodontis* (Lxc), a pathogen that causes stunting in Bermuda grass (*Cynodon dactylon*) and *L. xyli* subsp. *xyli* (Lxx). The objective of this work is to organize the information about completely sequenced genome of these bacteria, when these genomes are available. Our goal is to establish an environment that organizes information about these bacteria in order to assist sequence comparison and functional annotation. In this sense, we report the construction of LeifDB, a relational database that groups and standardizes information about the genome sequences of *Leifsonia* species. Currently, the LeifDB contains the 11 completely sequenced genomes (i.e., five *Clavibacter* and six *Leifsonia*). *Clavibacter* species are plant pathogens and the closest of the *Leifsonia* species. The database contains the ORF (Open Reads Frame) prediction of these 11 genomes using Prokka software (<http://vicbioinformatics.com/>) and functional categorization based on the COG Database (<ftp://ftp.ncbi.nih.gov/pub/COG/COG2014/data>). We also stored the analysis generated by the OrthoMCL DB (Database of Ortholog Groups of Protein Sequence) with all the predicted proteins of the 11 species. The database is under construction and was created with the data management system MySQL. Next, we intend to propagate functional annotations of the *Xanthomonas* Genome Project to help the comparison with other genomes sequenced by the group in previous analyses. The main objective of this work is to organize a group of genes potentially associated with pathogenicity not present in free-living bacteria. With this work, we aim to provide researchers and other interested parties with a trusted, well-structured source to support their research and work.

Palavras-chaves: doenças da cana de açúcar, sequenciamento genoma,