

## Poster H-2

### GINGA - Graphical Interface for comparative Genome Analysis



#### Authors:

Alexandre Rossi Paschoal (*LabInfo, LNCC*)  
Marcelo Marques Zerillo (*Institute of Biology, USP*)  
Luiz Gonzaga Paula de Almeida (*LabInfo, LNCC*)  
Oberdan de Lima Cunha (*LabInfo, LNCC*)  
Ana Tereza Ribeiro de Vasconcelos (*LabInfo, LNCC*)  
Claudia Barros Monteiro-Vitorello (*LabInfo, LNCC*)

**Short Abstract:** We report a system for comparative analysis of one complete bacterial genome and one of a close species that is being partially sequenced. The system permit the mapping of end sequences, contigs and scaffolds onto the complete genome, identifying conserved blocks of synteny, specific regions, repeats, and rearrangements.

#### Long Abstract:

Recently, the application of the knowledge newly obtained due to large-scale sequencing led to comparative and functional studies of genomes. We report a system for comparative analysis of one complete bacterial genome and one of a close species that is being partially sequenced. The system permits the mapping of end sequences, contigs and scaffolds onto the complete genome, identifying conserved blocks of synteny, specific regions, repeats, and rearrangements. Allows guiding the genome sequencing based on comparisons to the complete genome, having the biologists decide fragments of interest that should be further sequenced.

GINGA is a web-based tool that accesses a database MySQL ([www.mysql.com](http://www.mysql.com)) containing information resulted from assembly of sequencing reads into contigs using phredPhrap ([www.phrap.org](http://www.phrap.org)), organization of contigs into scaffolds using genscaff ([www.lbi.ic.unicamp.br](http://www.lbi.ic.unicamp.br)), mapping using cross\_match of end sequences, contigs and scaffold to the complete genome, BlastX and BlastN (Altschul et al., 1997) results, and annotation data based on SABIA software ([www.sabia.lncc.br](http://www.sabia.lncc.br)). We used BioPerl ([www.bioperl.org](http://www.bioperl.org)) in some of the biological data and module GD to build the graphical viewer.

Three modules compose the complete pipeline. The first one have the information of the complete genome sequenced. The annotation module of SABIA (Almeida et al., 2004) was modified to directly import the information from GenBank ([www.ncbi.nih.gov](http://www.ncbi.nih.gov)) and have it in a compatible format. The second module concerns information about the ongoing sequencing project (assembly and annotation). The third module is the GINGA comparative tool, which is able to read the entire information generated in the other modules, to map one genome onto the other and to show it in a user-friendly interface of a graphical overview. Navigating through the complete genome, you can see the contigs and scaffolds from the partial genome project that are specific and regions of synteny, repeats and rearrangements pointing to their positions at the complete genome. Navigation functionality includes moving left and right, zooming in and out, and looking at details of gene prediction and annotation of both, the complete and the partial genome sequences.

Comparative genome analysis of two close related bacterial pathogens of the same species

infecting different plant hosts: *Leifsonia xyli* subsp. *xyli* (Lxx), a sugar cane pathogen (Monteiro-Vitorello et al., 2004); and *Leifsonia xyli* subsp. *cynodontis* (Lxc), a *Cynodon* grass pathogen was used as our biological model for the development of GINGA. Currently, the genome of Lxc has 9952 reads of sequencing ends, assembled into 1064 contigs and 317 scaffolds. The total number of non-redundant bases is 1.1 Mb, corresponding to approximately 50% of Lxc genome. By mapping these sequences onto the Lxx complete genome, we have identified nearly 731 Kb conserved blocks of synteny and 115 Kb specific sequences of Lxc genome.

#### References:

Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ. (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.*, 25: 3389-402.

Almeida LG, Paixao R, Souza RC, Costa GC, Barrientos FJ, Santos MT, Almeida DF, Vasconcelos AT., et al. (2004) A System for Automated Bacterial Integrated Annotation-SABIA. *Bioinformatics*, 20: 2832-3.

Monteiro-Vitorello, C. B.; Camargo, L. E. A.; Van Sluys, M. A.; Kitajima, K. P.; Truffi, D.; Do Amaral, A. M.; Harakava, R.; De Oliveira, J. C. et al. (2004). The genome sequence of the gram-positive sugarcane pathogen *Leifsonia xyli* subsp. *xyli*. *Molecular Plant-Microbe Interactions*, 17: 827-836.

Support: LabInfo/LNCC/CNPq /MCT/