INSYGT : a multi-genes multi-genomes browser that highlights synteny conservation in prokaryotes.

Thomas Lacroix, Valentin Loux, Annie Gendrault-Jacquemard Gendrault, Mark M. Hoebeke, Jean-François Gibrat

INSYGT : a multi-genes multi-genomes browser that highlights synteny conservation in prokaryotes.

Thomas Lacroix[1], Valentin Loux[1], Annie Gendrault[1], Mark Hoebeke[2] and Jean-Francois Gibrat[1]

[1] Mathématique, Informatique et Génome, INRA, 78352 Jouy-en-Josas Cedex, France.
thomas.lacroix, valentin.loux, annie.gendrault and jean-francois.gibrat@jouy.inra.fr

[2] Laboratoire Statistique et Génome, Tour Évry II, 2ème étage, 523 Place des Terrasses, 91000 Évry, France.
Mark.Hoebeke@jouy.inra.fr

Abstract: *We have developed a web application to explore synteny conservation in prokaryotes : INSYGT (INteractive SYnteny Graphics and Tools). Its graphic interfaces features AJAX web development techniques. ORIGAMI, a relational database on gene context conservation, is used as a back-end. INSYGT is aimed to be integrated as a tool for the genome annotation system AGMIAL [1].*

Keywords: Synteny conservation, AJAX web interface.

1 Background

AGMIAL has been developed as a modular and extensible genome annotation system which assists expert manual annotators in their annotation strategy. Its design has been based on three cornerstones : (1) interactive web interfaces that allow experimental biologists to manage their annotation projects and contribute to the improvement of the genomic encyclopedia, (2) a suite of relevant bioinformatics tools (detection of genes, domains, subcellular localization, homology searches, fold recognition, etc) and (3) an underlying relational database. Genomic context information tools are currently being integrated to further help improving the annotation process. In that regard, we have been developing INSYGT to display homologies and synteny conservation in a multi-genes multi genomes fashion. It is a rich internet application featuring AJAX technologies which increases the web page's interactivity and speed. Most of the functionalities are handled on the client side so its usability is similar to desktop application. It uses ORIGAMI as a back-end, a relational database that stores genomic and synteny data. A cross comparison of all available microbial genomes is periodically carried out and stored in ORIGAMI; Orthologs are defined by the bidirectional best hit method and common gene evolution events (gene fusion/fission, local rearrangements) are detected by a local alignment method based on dynamic programming. Gene duplication events are not yet supported.

2 Description of the tool

The user interface is currently structured into three tabs :

- the 'Search' tab guides the user into querying the back-end. The user chooses the reference genome along with the set of genes to visualize results for. The choice for the initial gene set can be based on genomic location, functional category, biological process or any other genomic attributes.

We see this feature as a complement to genome browsers that display gene set based strictly on genomic location. An annotator could for example includes all genes from a metabolic pathway in its gene set and then be able to browse other species for similarities and synteny conservation to hint for genomic evolution.

- the 'Results Browsing' tab allows the visualisation of homologous genes across different species. Genes are symbolically represented as a box of uniform length containing the gene name, a match bar representing the alignment and a strand indicator. The reference genome stay visible on top at all time and each homologous genes is kept horizontally aligned in synchronisation with its match in the reference genome. The genes are clustered and background-colored according to the synteny group they belong. The functional annotation of a gene and its alignment information with its homologous counterpart in the reference genome are shown on demand. Different options are available : (1) user-defined set of species which appears at the top of the results list, (2) user-defined criteria for sorting the results, (3) the gene's EC number/functional annotation can be highlighted, etc The user has the possibility to navigate through the reference genome. Private data such as annotation projects are kept confidential and can be viewed in the interface along with the public data.

- the 'Genomic Organization' tab shows arrangements of any two sets of synteny block, with similar genes graphically linked between the two sets. Two main options are available : (1) vertical adjustment of the two synteny blocks so that any two genes can be aligned and (2) the possibility to reverse the orientation of the genomic arrangement (5' to 3' or 3' to 5') for each block.

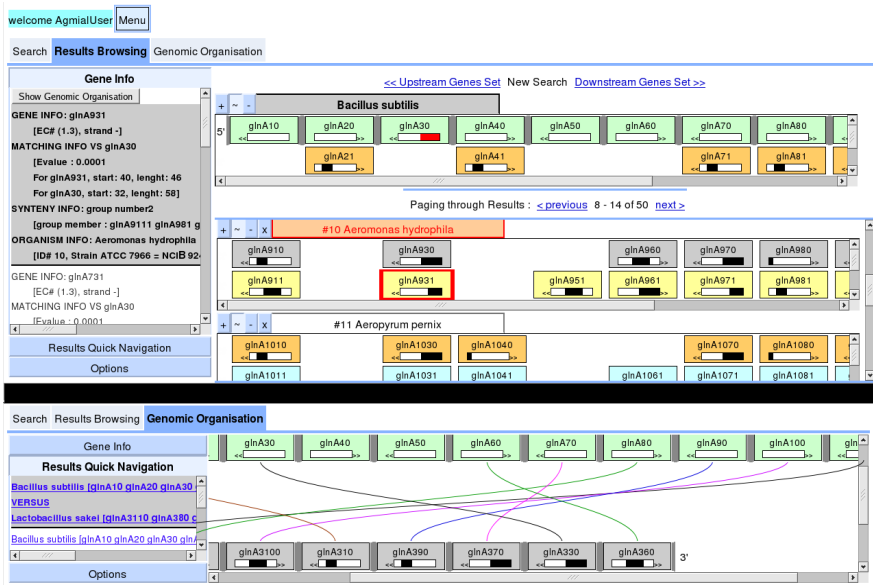

Figure 1. the user interface

Acknowledgements : This work is supported by the Microbiogenomics project (Appel d'offres ANR : ARA Masse de données ; ANR-05-MMSA-0009)

References

[1] Bryson K, Loux V, Bossy R, Nicolas P, Chaillou S, van de Guchte M, Penaud S, Maguin E, Hoebeke M, Bessières P and Gibrat JF., AGMIAL: implementing an annotation strategy for prokaryote genomes as a distributed system. *Nucleic Acids Res.*, 34(12):3533-45, 2006.