

Joëlle Amselem¹, Nicolas Lapalu¹, Baptiste Brault¹, Michaël Alaux¹, Fabrice Legeai², Françoise Alfama¹, Laetitia Brigitte¹, Nathalie Choise¹, Aminah Keliet¹, Erik Kimmel¹, Jonathan Kreplak¹, Isabelle Luyten¹, Cyril Pommier¹, Sébastien Reboux¹, Stéphanie Sidibe-Bocs³, Delphine Steinbach¹, Marc-Henri Lebrun⁴ and Hadi Quesneville¹

¹INRA-URGI, Versailles, France; ²UMR BIO-3P, Rennes, France; ³CIRAD Montpellier, France; ⁴INRA-BIOGER, Thiverval grignon, France;

Contact : joelle.amselem@versailles.inra.fr

Abstract

Nowadays with the development of NGS, more and more genomes are sequenced, producing very large amount of data. However, annotations can't keep pace, introducing a lack between genome data and annotation releases. To face this challenge, the URGI (<http://urgi.versailles.inra.fr>) platform aims at providing tools to annotate entirely sequenced genome comprising: pipelines, databases and user-friendly interfaces to browse and query the data. We will focus here on complementary systems developed in the frame of GnpIS Information System QuickSearch (<http://urgi.versailles.inra.fr/gnpis>) and GnpAnnot (<http://www.gnpannot.org/>) projects.

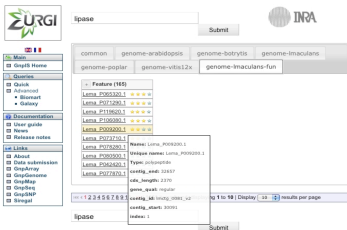


Figure 1: **Quick search** request on whole GnpIS. Results are displayed through lists for each DBs containing at least one occurrence of the keyword search. Concerning Genome DBs links lead to Browse region or Gene Report.

Quick and Advanced search

GnpIS gateway is based on the Apache Lucene™ full-featured text search engine library. GnpIS DBs have been indexed included two well known database (DB) schemas: SeqFeature:Store and Chado (gmod.org). Indexes are generated in several ways to query structural or functional data stored in same or separate DBs. Results are returned according to significance with term criteria to provide links to Gbrowse or Genome Report System (GRS).

BioMart based datamarts were set up to be used as an advance search tool. Results of the complex search criteria could be exported in different formats or directly send to a Galaxy framework for further bioinformatic analysis.

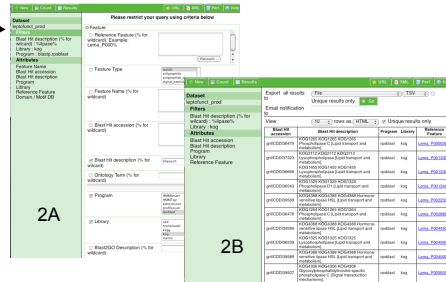
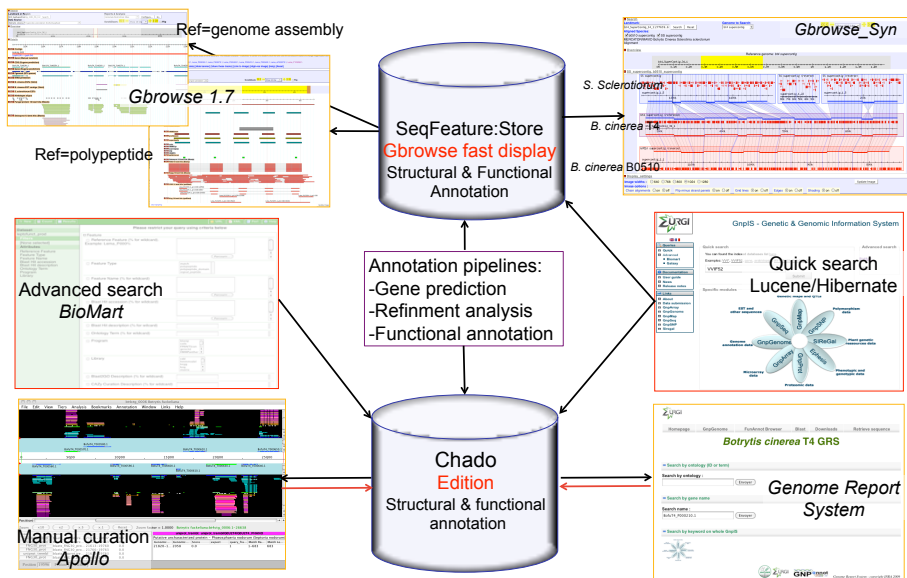


Figure 2: Biomart based **advanced search** request using filters (2A) and results displayed in attributes selected columns (2B).



A distributed annotation system allowing gene structure & function manual curation

This system relies on the well known GMOD databases and interfaces (chado / Gbrowse / Apollo). Curated data are available and shared by the consortium community as soon as they are committed in the database using "pure JDBC" direct communication protocol between Apollo and Chado. Annotation provided by Structural and functional annotation pipeline are respectively inserted into Chado and DB::Bio::SeqFeature::Store DBs schemas from GMOD.

Figure 3: Dataflow and interoperability of the Integrated genome annotation system. Cross ref links on the different interfaces allow the interoperability between them. Black arrows: DB access in read mode. Red arrows: DB access in write mode (GRS access in write mode restricted to the "Edition" report)

The Genome Report System (GRS): an interface to visualize information related to a gene and its genome locus and edit functional annotation

The GRS, written in Java, was developed to produce various and user-friendly Web reports. GRS uses structural and functional genomic data stored in Chado DB to provide users with a comprehensive categories of reports (functional domains, blast hits, orthologous genes, genome same locus. GRS proposes also a Gene Ontology browser and an editing module (GRE) to allow manual functional annotation.

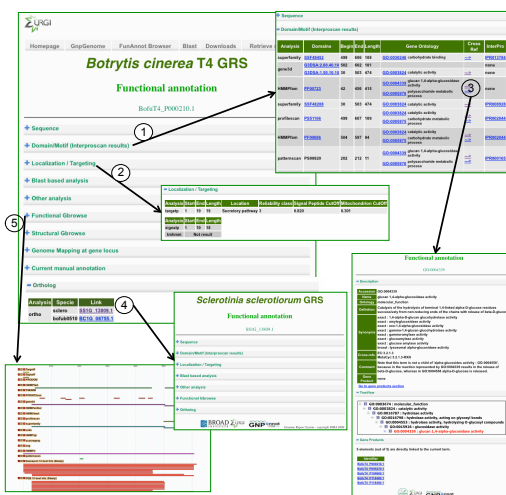


Figure 4: Genome Report System screen shot for one *B. cinerea* gene. Each Category can be opened, each table / link / other pane displayed correspond to a specific service developed.

1. Domain/motifs interspersed result service
2. SignalP/TargetP result service
3. Link to ontology tree and cross reference to all gene sharing the same GO
4. Link to *S. sclerotiorum* (*B. cinerea* B0510) orthologous gene report
5. Link to Domain/motif graphical display (Gbrowse/SeqFeature::Store)

Conclusion and Perspectives

The integrated genome annotation system was successfully set up for fungal genome annotation of *Botrytis cinerea* T4 (grey mould disease) and *Leptosphaeria maculans* (stem canker), urgi.versailles.inra.fr/index.php/Species. GRS has been developed and set up in the frame of the Aphid sequencing and annotation project, www.aphidbase.com/aphidbase. The GRS edition service is still under development and will be soon released, for these genomes.