USMI Galaxy Demonstrator(UGD): a collection of tools to integrate microorganisms information

D.P. Colobraro, P. Romano - IRCCS AOU San Martino IST, Genoa, Italy
{danielepierpaolo.colobraro,paolo.romano}@hsanmartino.it

NETTAB & IB 2015, October 14-16, Bari, Italy

Background

As a result of several heterogeneous repositories, which collect information on microorganisms, some web platforms gather microorganisms information and provide services to mBRCs in limited way. The USMI Galaxy Demonstrator(UGD), presented at Bioinformatics Italian Society (BITS) meeting in June 2015, support both researchers and mBRC staffs to perform bioinformatics pipelines in an automatic way, importing available microbial catalogues, enriching them with enzyme data, ribosomal RNA sequences and taxon IDs according to MIRRI' aims.

Scope

In this new version of UGD, we want to extend the integrative capabilities providing tools that are able to avoid manual, potentially long, searches on the web and to identify and select microorganisms of interest using metabolite, ligand, enzyme and protein names via new tools: *From alignment of proteins to microbial strain* and *Compound and enzyme*.

USMI Galaxy Demonstrator (UGD)

The USMI Galaxy Demonstrator is publicly available on-line at http://bioinformatics.hsanmartino.it:8080, Galaxy version 15.07. The developed tools are available in two sections, ***Get microbial data*** (box 1) and ***Retrieval external information*** (box 2), under the general label 'BASIC TOOLS FOR MIRRI'. The new tool ***From alignment of proteins to microbial strain*** integrates alignment of proteins with strain number, taxon ID, link to CABRI and related DBs by using data provided by blastp. ***Compound and enzyme*** retrieves strain number, link to CABRI, Compound/ligand accession number, EC number, name and synonyms by using a biological term. Galaxy allows to set up workflows to rerun, store and share both specific analyses and data. As shown in fig. 1, tools may be set in various ways in order to define own pipelines. Indeed, implementing basic tools as modular elements allows to make up several pipelines.

CABRI, Common access to biological resource and information, Network Services (http://cabri.org) offer access to 28 catalogues from European Biological Resources Centers (BRCs), since 2000.

MIRRI, Microbial Resource Research Infrastructure, is a pan-European distributed research infrastructure in its preparatory phase which aims to connect all European **mBRCs,** microBiological Resource Centres, with the aim of providing improved and extended services to the research and industry communities. **MIRRI** wants to reach the integration of information on microorganisms with further data that can be found and retrieved from a wide range of biological databases like NCBI, EMBL, BRENDA and UNIPROT.

Results

The most recently developed UGD tools, fig. 2, are able to identify which microorganisms that are related to a molecule or a protein of interest to the end user by integrating information both from mBRCs catalogues and from external data sources.

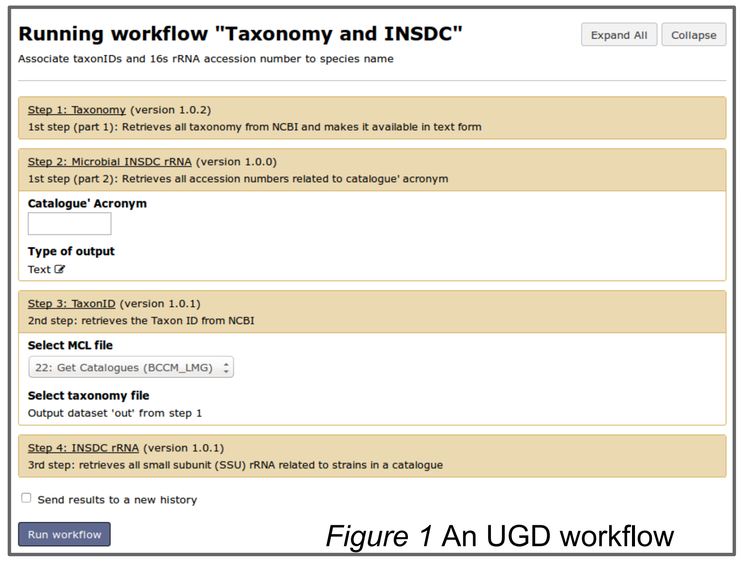

Figure 1 An UGD workflow

Taxonomy *retrieves all taxonomy information*
Microbial INSDC rRNA *retrieves information by using a Catalogue acronym*
Upload file *is a Galaxy' generic tool*
Get Catalogues *is a 'data_source' tool to import catalogues from external-web storage*

box 1

TaxonID *retrieves taxonomy ID for all strains*
ECNumber *gathers information when enzyme names are collected in Catalogue*
Protein FASTA *retrieves protein sequences by using protein accession number*
INSDC rRNA *retrieves rRNA accession number related to Strains in Catalogue*
PMID and DOI *retrieves Pubmed IDs and Digital Object Identifiers (DOIs) of given bibliographic references*
FASTA from INSDC *retrieves rRNA sequences by using accession number*
Uniprot *retrieves protein accession number by using all strains*

box 2

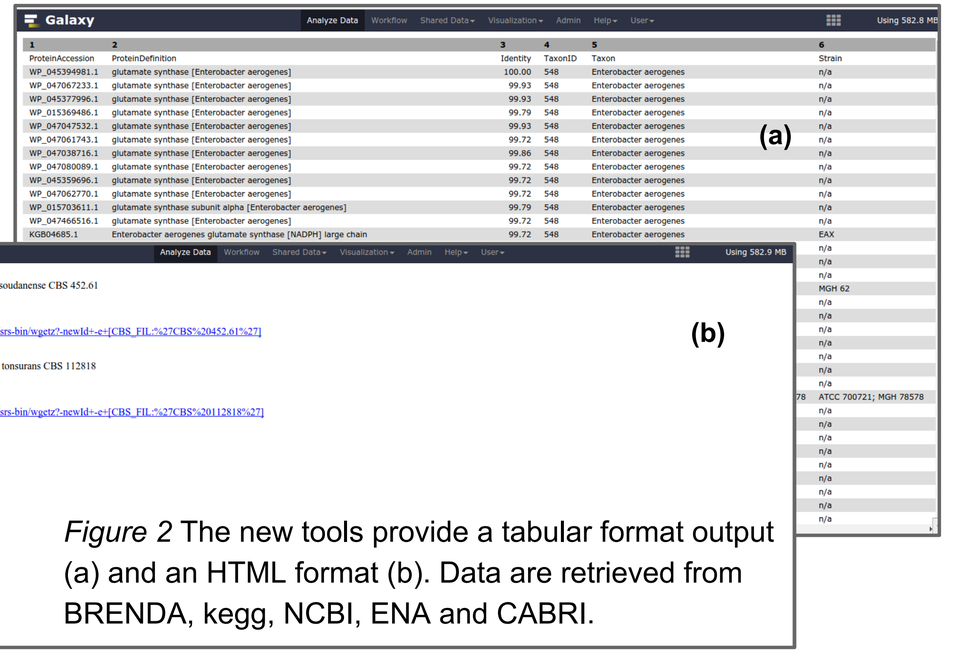

Figure 2 The new tools provide a tabular format output (a) and an HTML format (b). Data are retrieved from BRENDA, kegg, NCBI, ENA and CABRI.

References

- The MIRRI Project: www.mirri.org
- The Galaxy project: http://galaxyproject.org
- Colobraro DP and Romano P. A Galaxy approach to integrate microbial data: the USMI Galaxy demonstrator. Proceedings of BITS 2015, 3-5 June, Milano. Guffanti A et al. (eds) (in press)

Acknowledgements

The research leading to these results has received funding from the European Union's Seventh Framework Programme (FP7/2007-2013) under grant agreement no. 312251.