



**HAL**  
open science

## Analyze your microbiota sequencing data using a Galaxy-based framework References

Bérénice Batut, Clémence Defois, Kevin Gravouil, Cyrielle Gasc, Thomas Eymard, Éric Peyretailade, Jean-François Brugère, Pierre Peyret

### ► To cite this version:

Bérénice Batut, Clémence Defois, Kevin Gravouil, Cyrielle Gasc, Thomas Eymard, et al.. Analyze your microbiota sequencing data using a Galaxy-based framework References. JOBIM 2016 - Journées Ouvertes Biologie Informatique Mathématiques, Jun 2016, Lyon, France. hal-04763064

**HAL Id: hal-04763064**

**<https://hal.science/hal-04763064v1>**

Submitted on 1 Nov 2024

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Bérénice Batut, Clémence Defois, Kévin Gravouil, Cyrielle Gasc, Thomas Eymard, Eric Peyretailade, Jean-François Brugère and Pierre Peyret  
EA 4678 CIDAM, Université d'Auvergne, Clermont-Ferrand, France

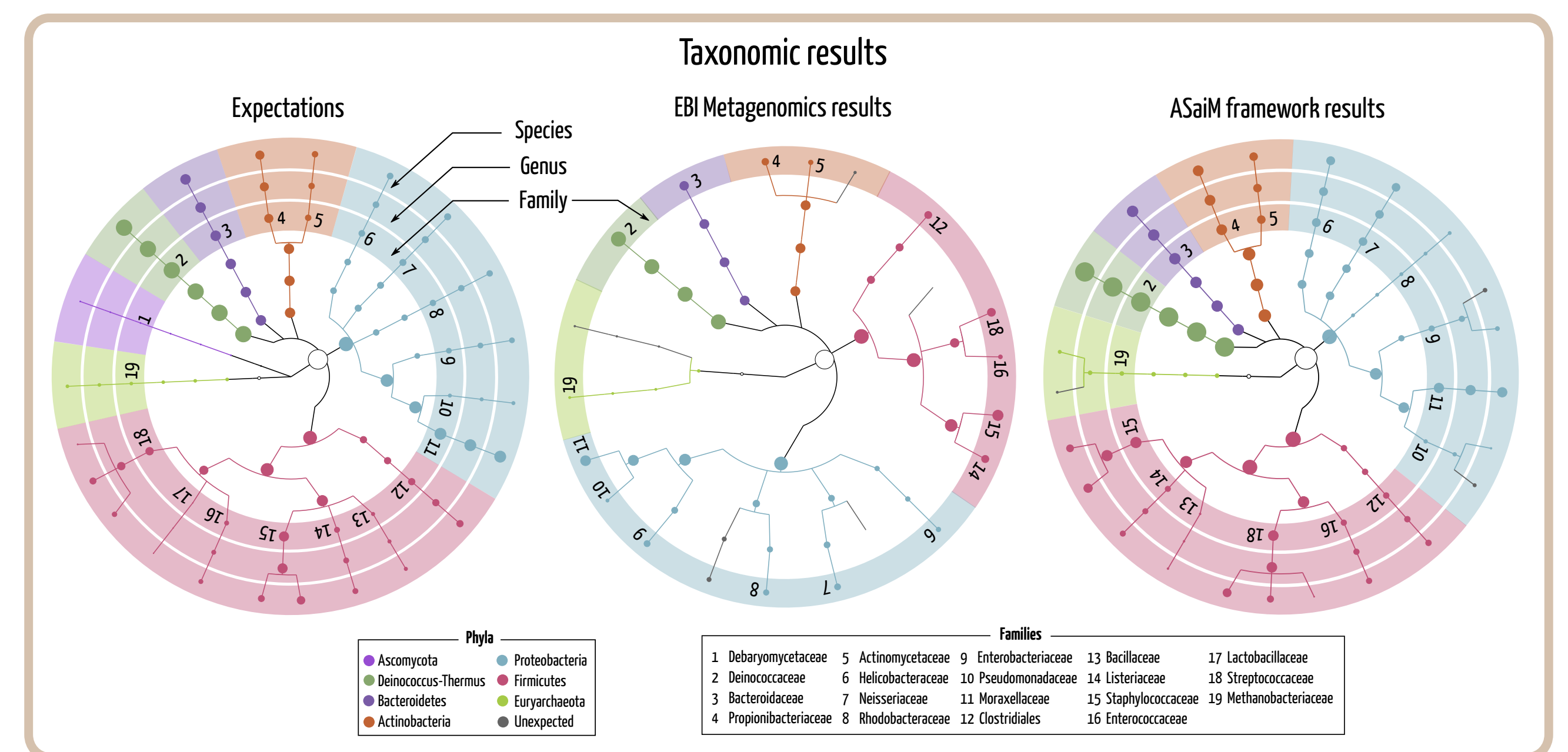
Nowadays, complex communities of microorganisms can be studied in depth with metagenomics and metatranscriptomics, driven by evolution of sequencing techniques. Indeed, these meta'omics techniques offer insight concerning structure and functions of the studied communities. Notwithstanding, analyses of raw microbiota sequencing data are difficult because of data size and so numerous available and needed tools (1,2) for data processing.

To overcome these limitations, we have developed ASaiM (Auvergne Sequence analysis of intestinal Microbiota), an open-source Galaxy-based opinionated framework. It is accessible via a web-based interface to help scientists in their microbiota sequence data analyses. With an expertly selected collection of tools, workflows and databases, the framework is dedicated particularly to taxonomic, metabolic and taxonomically related metabolic information from raw sequences. By its intrinsic modularity, ASaiM allows tuning of workflows, tool parameters and usable databases.

ASaiM framework is a powerful tool to analyze shotgun raw sequence data from complex communities of microorganisms. This open-source and biologist-oriented solution enhances usability, reproducibility and transparency of such studies and their comparative analyses.

## Validation

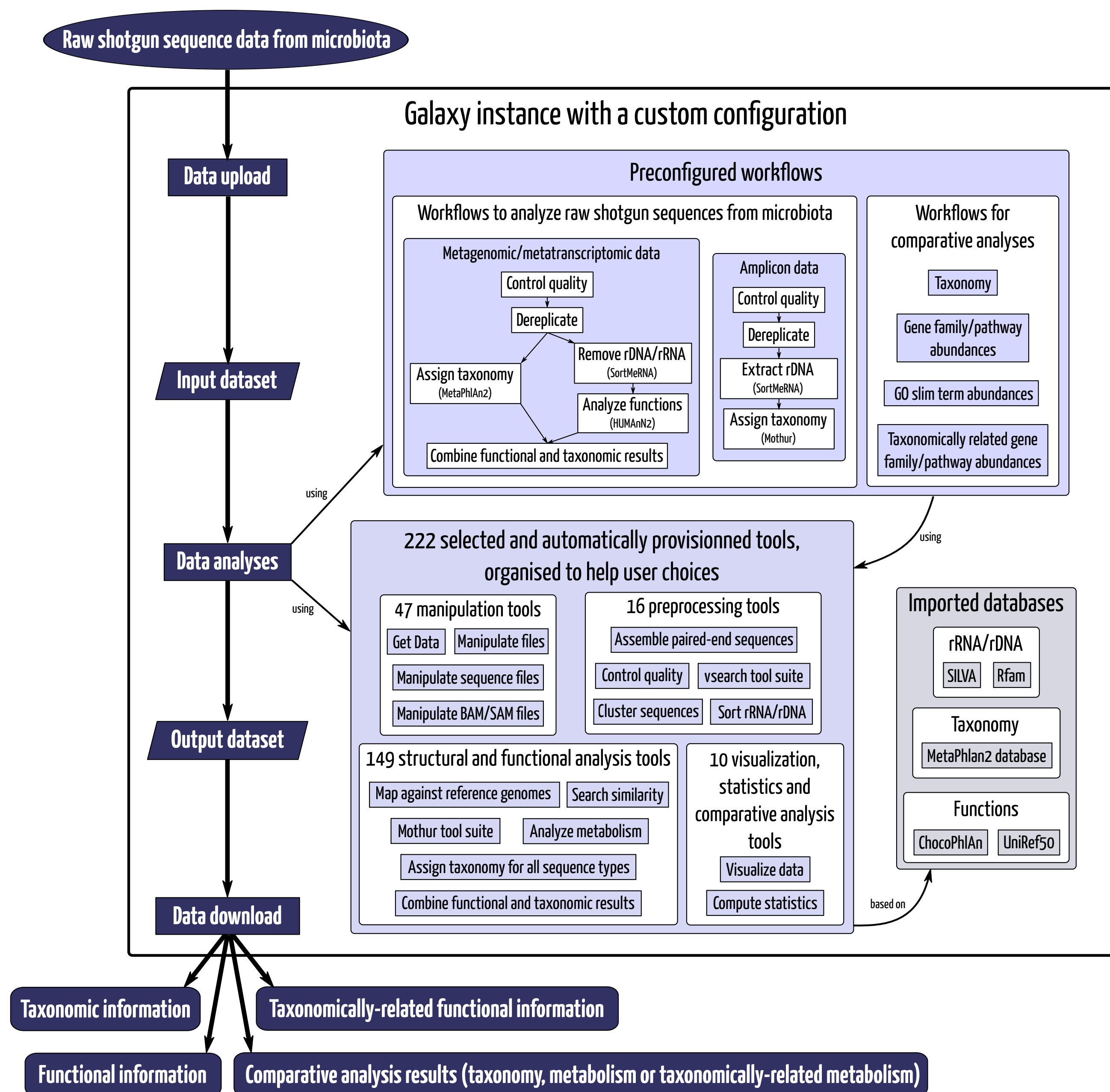
ASaiM framework was tested on two mock metagenomic datasets from HMP metagenomes mock pilot project. These datasets are metagenomics shotgun sequences (>1,200,000 454 GS FLX Titanium single-end sequences) from a controlled community (with 22 known microbial species), available on EBI metagenomic database (SRR072232 and SRR072233). Results obtained with ASaiM framework were intensively analyzed and compared to the ones from EBI metagenomic pipeline (3, version 1.0). Only results for SRR072233 dataset are presented here.



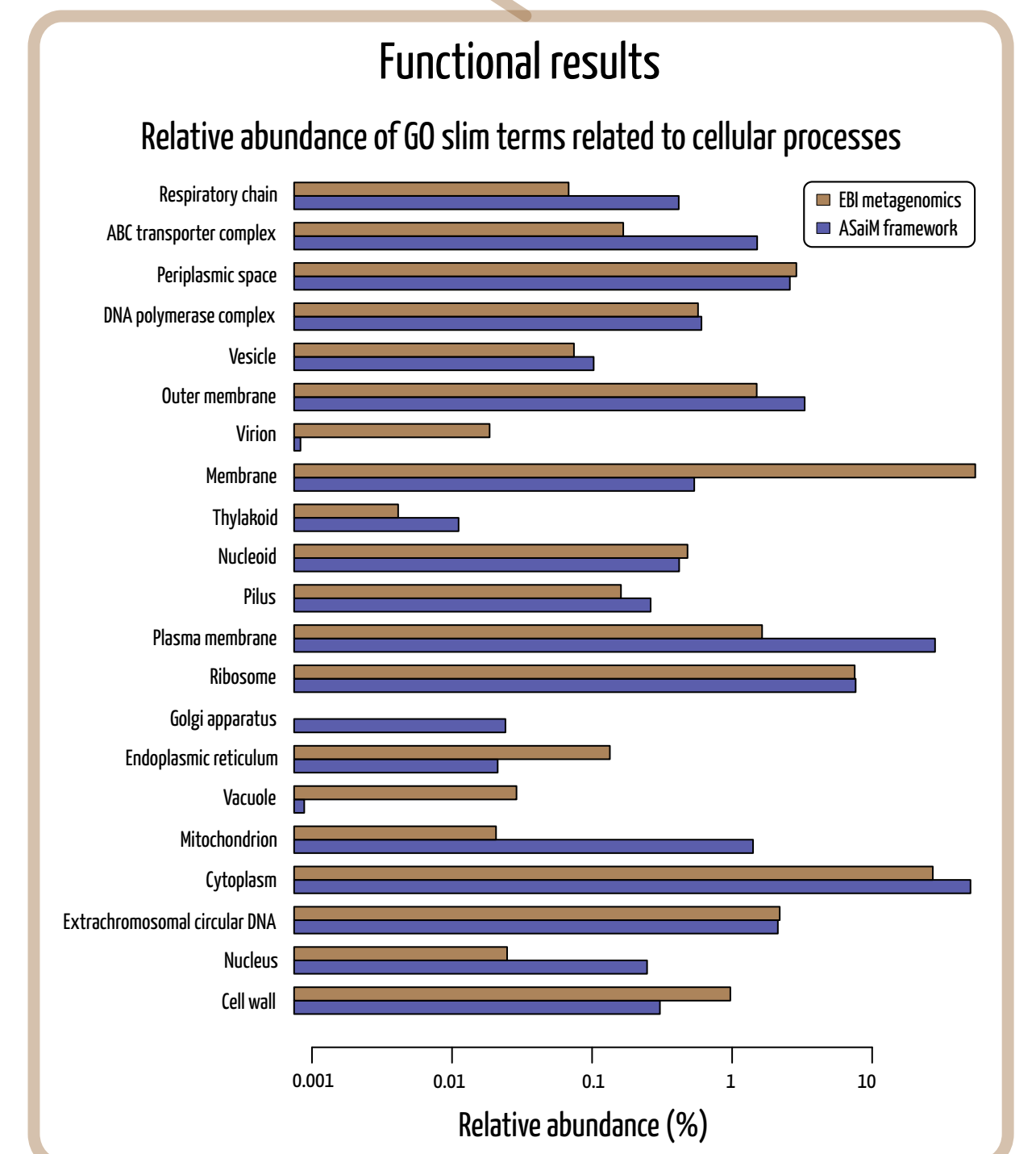
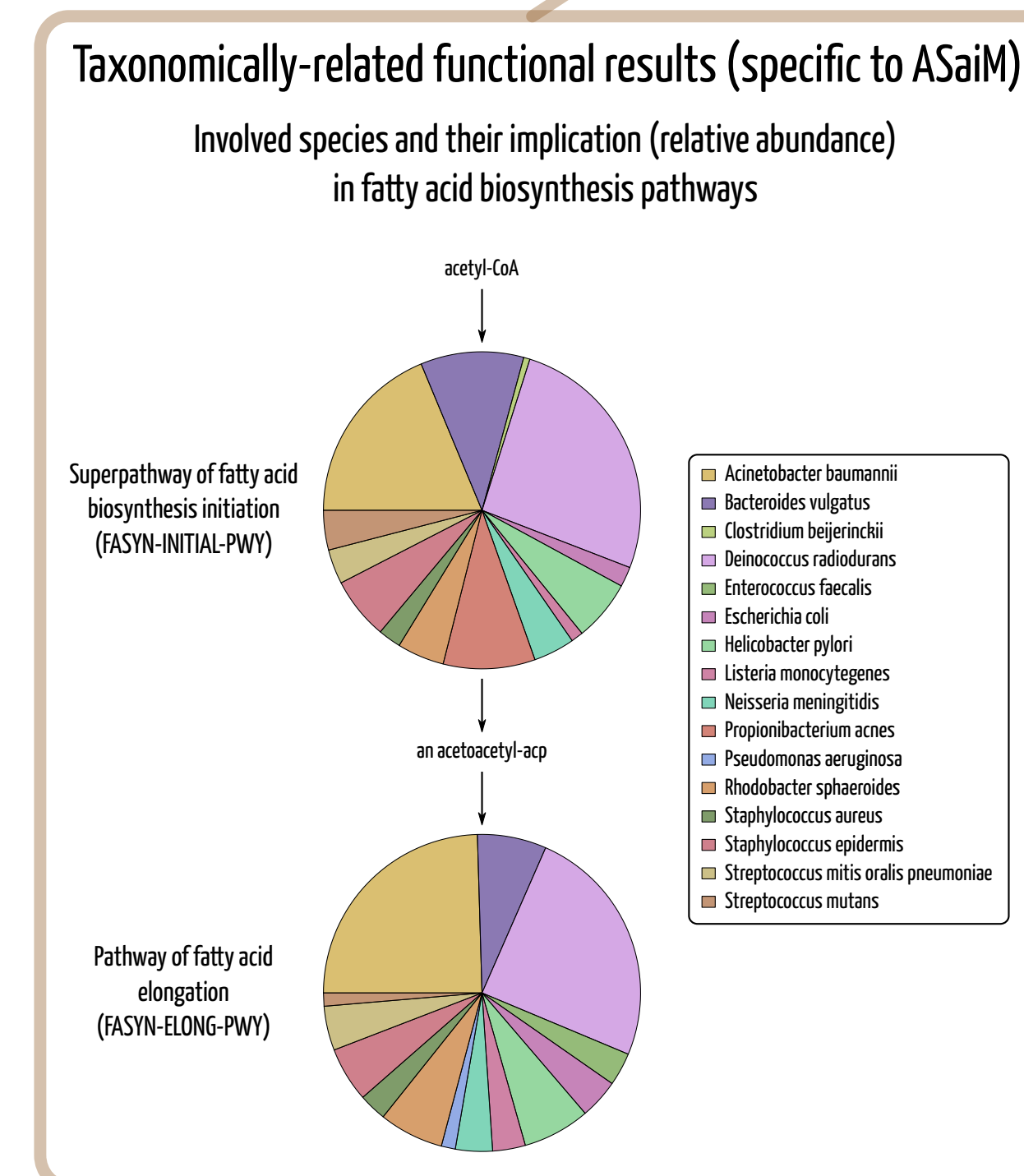
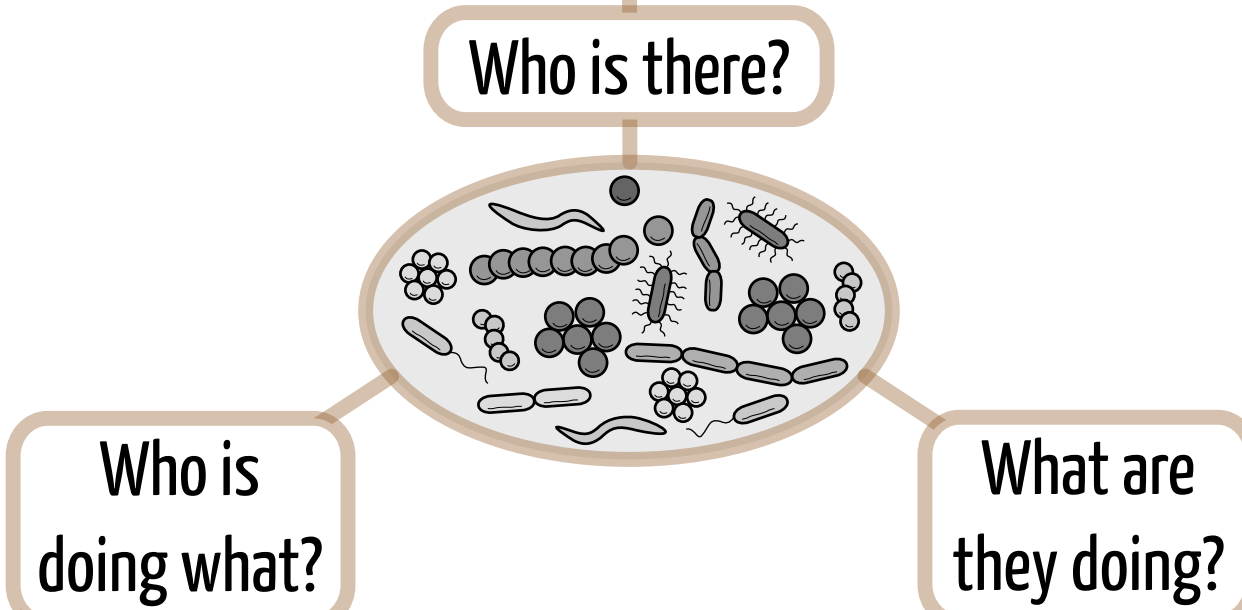
## Implementation

Based on a Galaxy instance with a custom configuration, ASaiM framework integrates >220 tools (Figure 1), specifically selected for microbiota studies. They are hierarchically organized (Figure 1) and documented (<http://asaim.readthedocs.org/>) to orient user choice toward best tools during analyses. These tools can be used separately or orchestrated inside workflows.

To help users with their analyses of raw microbiota sequences, a default but customizable workflow (Figure 1) is proposed in ASaiM framework. It produces accurate and precise taxonomic assignments, multi-level functional results (gene families, pathways, GO slim terms) and taxonomically related metabolism information, in few hours on a commodity computer. Other workflows are also preconfigured for comparative analyses, and also amplicon data analyses.



**Figure 1:** ASaiM framework and components of its Galaxy instance. Using scripts in framework code source, the Galaxy instance is easily configured, launched and provisionned with tools, workflows and databases.



Taxonomic analyses gives a great insight into the community structure with complete, accurate and statistically supported information. A broad overview of metabolic profile is available with gene families, pathways and GO slim terms. The taxonomically-related functional information, specific to ASaiM framework, allows to investigate which species is involved in which metabolic function.

Details about these analyses, more results, analyses and representations are available on [https://github.com/ASaiM/hmp\\_mock\\_tests](https://github.com/ASaiM/hmp_mock_tests).

## Further information

ASaiM framework source code is available under Apache 2 license at <https://github.com/asaim/framework>, as a collection of bash, Python and Ansible scripts. The source code includes framework deployment: automatic configuration of the Galaxy instance, its deployment and its provision with selected tools, workflows and needed databases.

Heavy documentation of the custom of Galaxy instance, the tools, the workflows and the databases is available at <https://asaim.readthedocs.org>

## References

- (1) Mendoza *et al*, Brief Bioinfo, 16 (2015), 745-758
- (2) Segata *et al*, Mol Sys Biol, 9 (2013), 66
- (3) Hunter *et al*, Nucl Acids Res, 42 (2014), D600-D606

## Acknowledgments



## Contacts

Bérénice Batut: [berenice.batut@gmail.com](mailto:berenice.batut@gmail.com)  
Pierre Peyret: [pierre.peyret@udamail.fr](mailto:pierre.peyret@udamail.fr)