



HAL
open science

A Galaxy-based framework to analyze raw sequencing data from microbiota

Bérénice Batut, Clémence Defois, Kevin Gravouil, Jean-François Brugère, Éric Peyretailade, Pierre Peyret

► To cite this version:

Bérénice Batut, Clémence Defois, Kevin Gravouil, Jean-François Brugère, Éric Peyretailade, et al.. A Galaxy-based framework to analyze raw sequencing data from microbiota. Galaxy Community Conference 2017, Jun 2017, Montpellier, France. 10.7490/f1000research.1114448.1 . hal-04763022

HAL Id: hal-04763022

<https://hal.science/hal-04763022v1>

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.



B er enice Batut, Cl emence Defois, K evin Gravouil, Jean-Fran ois Brug ere, Eric Peyretailade and Pierre Peyret
UMR 454 MEDIS, Clermont-Ferrand, France
Bioinformatic Group, Freiburg, Germany

New generation of sequencing platforms coupled to numerous bioinformatics tools has led to rapid technological progress in metagenomics, metatranscriptomics and complex microbial community studies. Nevertheless, a combination of different bioinformatics tools remains necessary for efficient extraction of useful information in microbiota studies. Modular, accessible, sharable and user-friendly tools would greatly improve such studies.

We therefore developed ASaiM, an open-source **Galaxy-based framework** dedicated to microbiota data analyses.

ASaiM provides an **expertly selected collection of tools** to explore and visualize taxonomic and functional information from raw amplicon, metagenomic or metatranscriptomic sequences. To help the analyses, **several (customizable) workflows** are included. The available workflows are supported by **tutorials** and **Galaxy interactive tours** to guide the users through the analyses. Furthermore, an effort on **documentation** of ASaiM, its tools and workflows has been made (<http://asaim.readthedocs.io/>). ASaiM is implemented as **Galaxy Docker flavour** and the associated source code is available under Apache 2 license at <http://github.com/ASaiM/framework>.

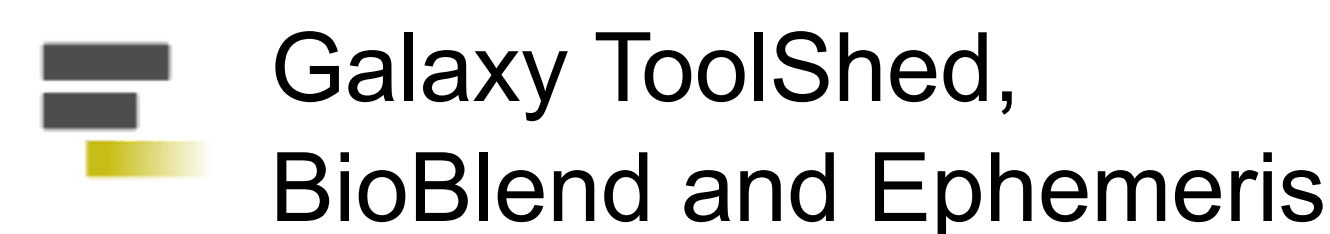
Based on the Galaxy framework, ASaiM offers sophisticated analyses to scientists without command-line knowledge. ASaiM provides a powerful framework to easily and quickly explore microbiota data in a reproducible and transparent environment.

A framework built on the shoulders of giants

To develop a modular, accessible, redistributable, sharable and user-friendly framework for scientists working with microbiota data



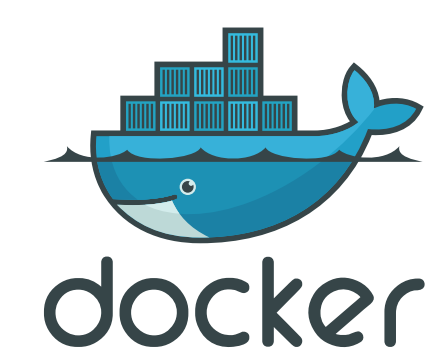
As the base



To install the Galaxy tools, the workflows and databases

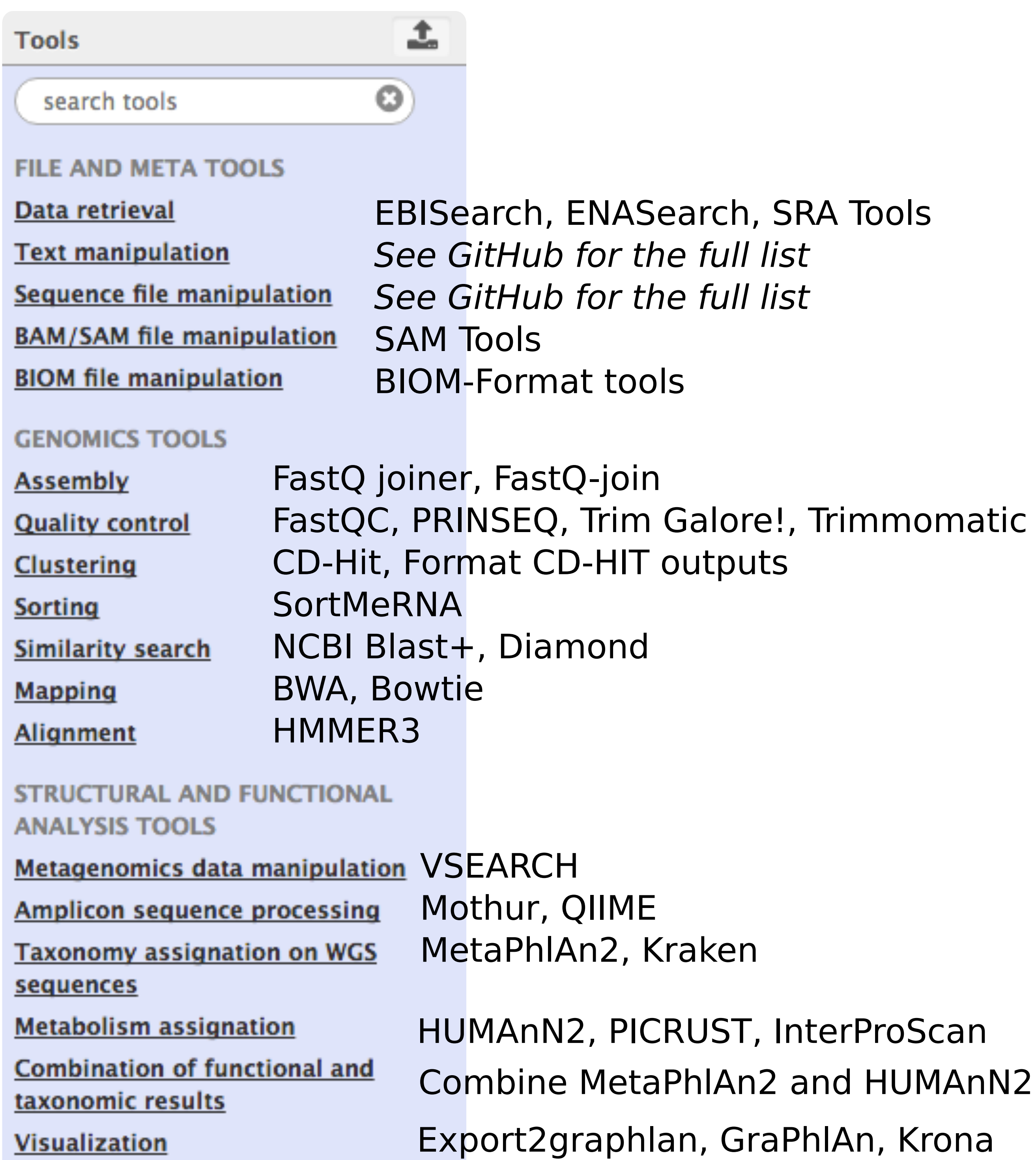


To install the tools and their dependencies



To containerize and ship everything

With a comprehensive set of microbiota related tools



Tools

search tools

FILE AND META TOOLS

- Data retrieval**: EBISearch, ENASearch, SRA Tools
See *GitHub for the full list*
- Text manipulation**: See *GitHub for the full list*
- Sequence file manipulation**: See *GitHub for the full list*
- BAM/SAM file manipulation**: SAM Tools
- BIOM file manipulation**: BIOM-Format tools

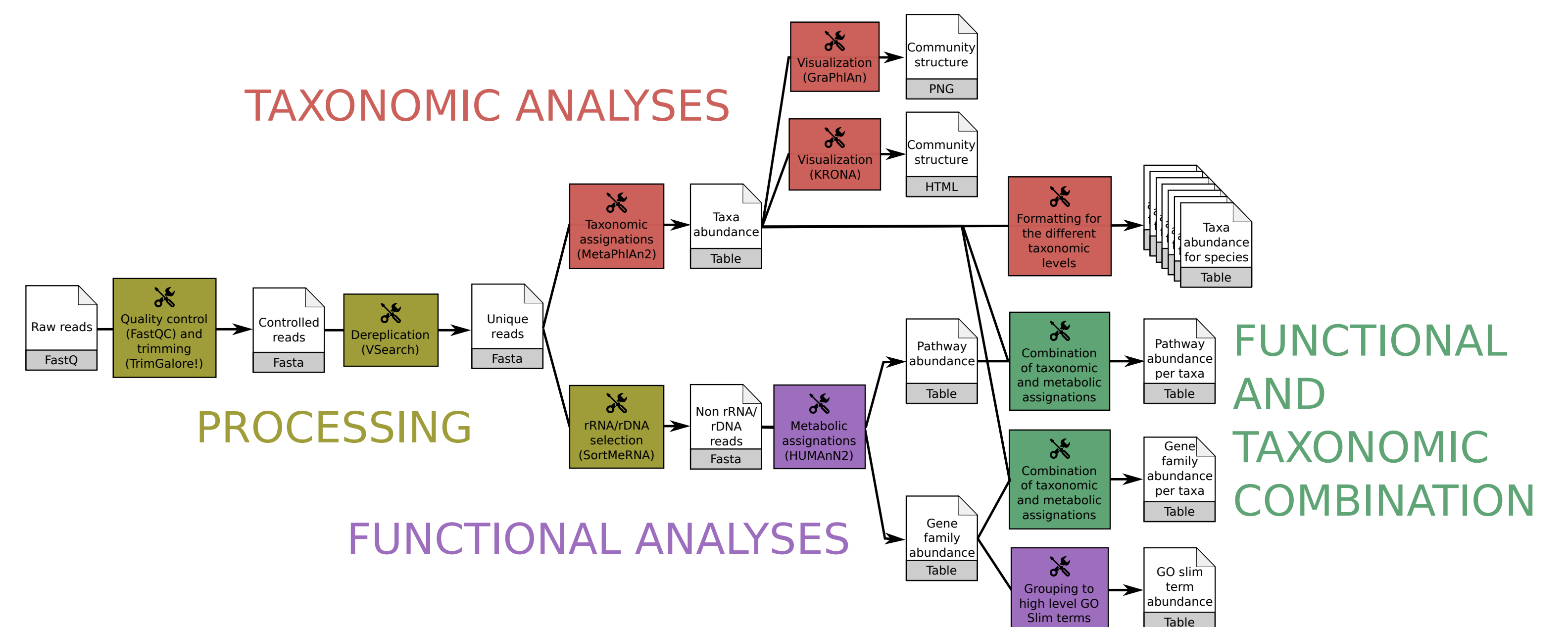
GENOMICS TOOLS

- Assembly**: FastQ joiner, FastQ-join
- Quality control**: FastQC, PRINSEQ, Trim Galore!, Trimmomatic
- Clustering**: CD-Hit, Format CD-HIT outputs
- Sorting**: SortMeRNA
- Similarity search**: NCBI Blast+, Diamond
- Mapping**: BWA, Bowtie
- Alignment**: HMMER3

STRUCTURAL AND FUNCTIONAL ANALYSIS TOOLS

- Metagenomics data manipulation**: VSEARCH
- Amplicon sequence processing**: Mothur, QIIME
- Taxonomy assignation on WGS sequences**: MetaPhlan2, Kraken
- Metabolism assignation**: HUMAnN2, PICRUST, InterProScan
- Combination of functional and taxonomic results**: Combine MetaPhlan2 and HUMAnN2
- Visualization**: Export2graphlan, GraPhlan, Krona

With predefined and tested workflows



Main workflow to analyze raw shotgun metagenomics data

With documentation and training

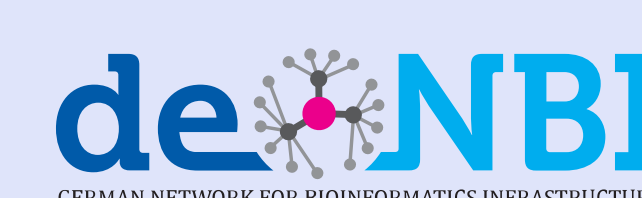
- Online documentation: <http://asaim.readthedocs.io>
how to use and deploy ASaiM
which tools are integrated with some documentation about them
which workflows are integrated and how to use them
- Galaxy Interactive Tours inside ASaiM
- Tutorials
Developed with the help of the Galaxy Training Network (GTN)
Hosted in the GTN website
<http://galaxyproject.github.io/training-material/metagenomics>

Useful information

GitHub: <http://github.com/asaim/framework>
Doc: <http://asaim.readthedocs.io>

Acknowledgements

CPER 2015
Axe T2ANSH
(Action 3, Aliments,
Nutrition Humaine Sant e)



Contact

B er enice Batut
berenice.batut@gmail.com