# How to analyze raw sequencing data from microbiota? Using ASaiM!

Berenice Batut[1], Clemence Defois[2], Kevin Gravouil[2], Jean-Francois Brugere[2], Eric Peyretaillade[2],
Pierre Peyret[2]

[1]*University of Freiburg, Germany*
[2]*MEDIS/LMGE/LIMOS, Universit Clermont Auvergne, France*
berenice.batut@gmail.com

The study of microbiota and microbial communities has been facilitated by the evolution of sequencing techniques and the development of metagenomics and metatranscriptomics. These techniques are giving insight into phylogenetic properties and metabolic components of microbial communities. However, meta'omic data exploitation is not trivial: large amount of data, high variability, incompleteness of reference databases, difficulty to find, configure, use and combine the dedicated bioinformatics tools, etc. Hence, to extract useful information, a sequenced microbiota sample has to be processed by sophisticated workflows with numerous successive bioinformatics steps. Besides, bioinformatics tools are often manually executed and/or patched together with custom scripts. These practices raise doubts about a science gold standard: reproducibility. Alternative approaches to improve accessibility, modularity and reproducibility can be found in Open-Source workflow systems such as Galaxy. Galaxy is a lightweight environment providing a web-based, intuitive and accessible user interface to command-line tools, while automatically managing computation and transparently managing data provenance and workflow scheduling [68]. In this context, we developed ASaiM (Auvergne Sequence analysis of intestinal Microbiota), an Open-Source opinionated Galaxy-based framework.

ASaiM provides an expertly selected collection of tools to exploit and visualize taxonomic and functional information from raw amplicon, metagenomic or metatranscriptomic sequences. To help the analyses, several (customizable) workflows are included. The main workflow has been tested on two mock metagenomic datasets with controlled communities. More accurate and precise taxonomic analyses and more informative metabolic description have been obtained compared to EBI metagenomics' pipeline on the same datasets.

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/).

Based on the Galaxy framework, ASaiM offers sophisticated analyses to scientists without command-line knowledge, while emphasizing reproducibility, customization and effortless scale up to larger infrastructures. ASaiM is implemented as Galaxy Docker flavour and can be easily extended with additional tools or workflows. ASaiM provides then a powerful framework to easily and quickly exploit microbiota data in a reproducible and transparent environment.