

ASaiM

A Galaxy framework to analyze gut microbiota data



Bérénice Batut – EA CIDAM, Clermont-Ferrand
November 19th, 2015

Context

Gut microbiota

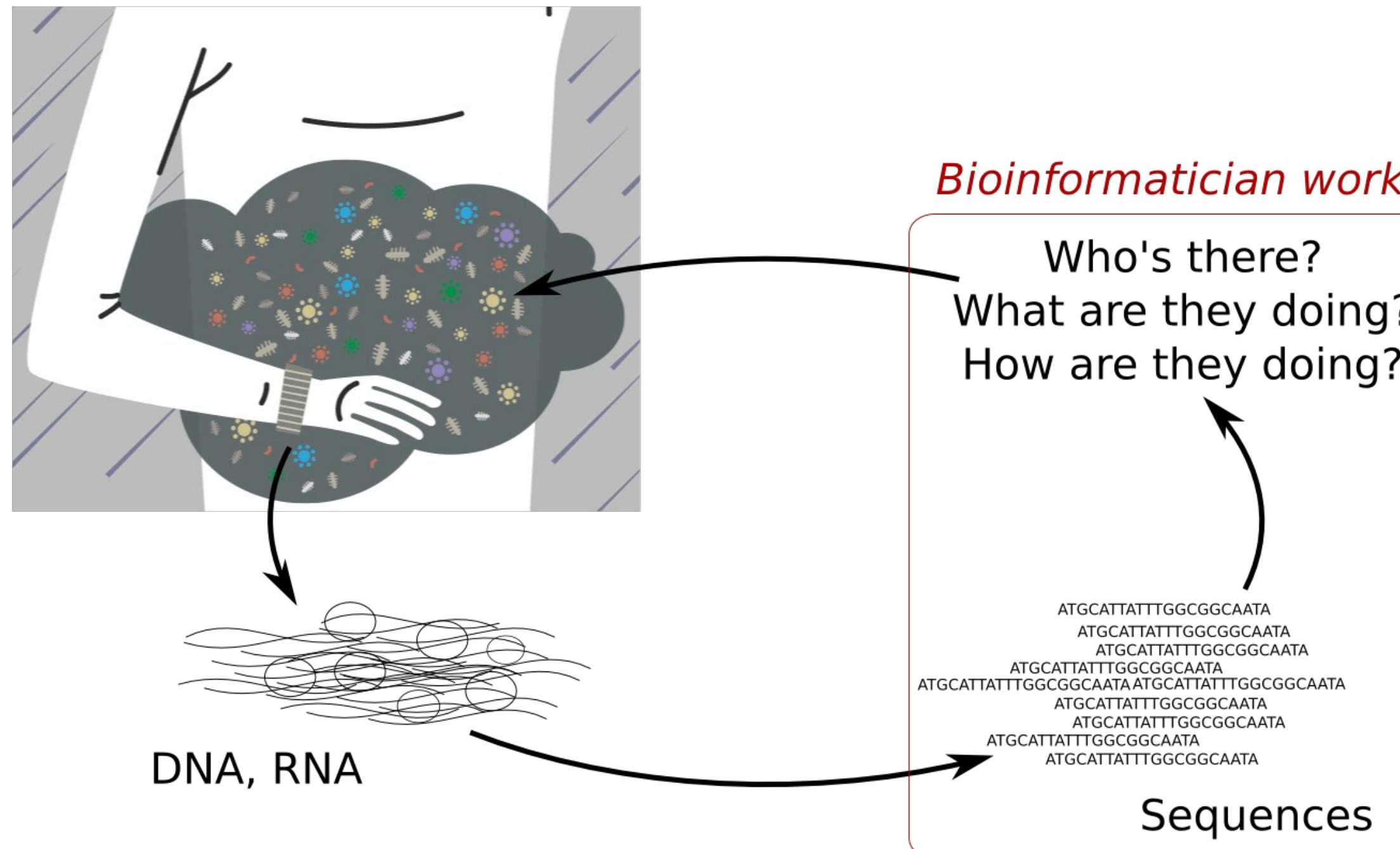


Community of microorganism species that live in the digestive tracts

Importance of gut microbiota

"Forgotten" organ

Metagenomic / Metatranscriptomic



Comparative meta-omic
Study of microbiota in its globality

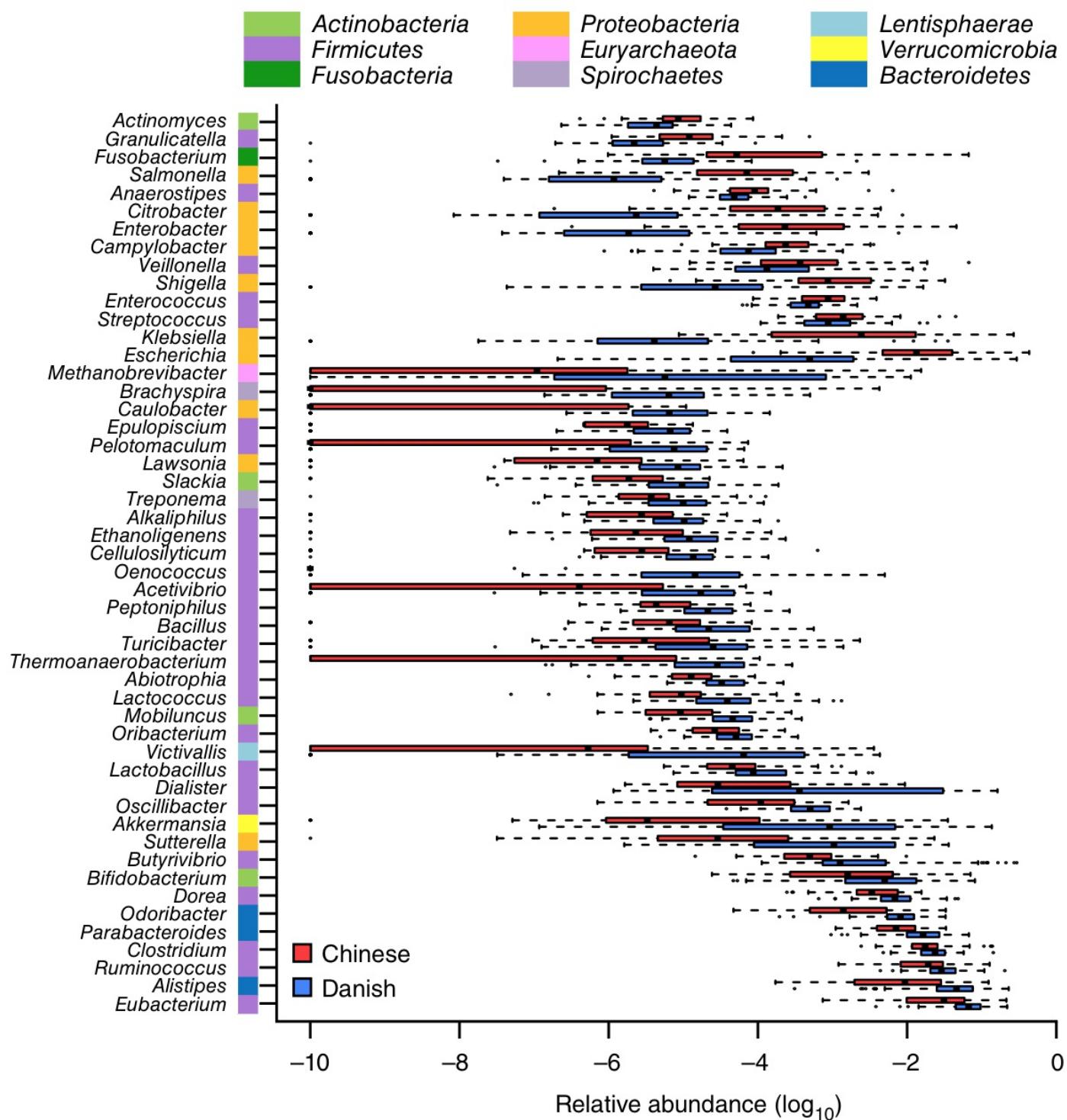


Figure 4C - Li et al, Nat Biotech, 2014

Comparative meta-omic of different projects
For a global view?

Public data repository
Source of information

European Nucleotide Archive

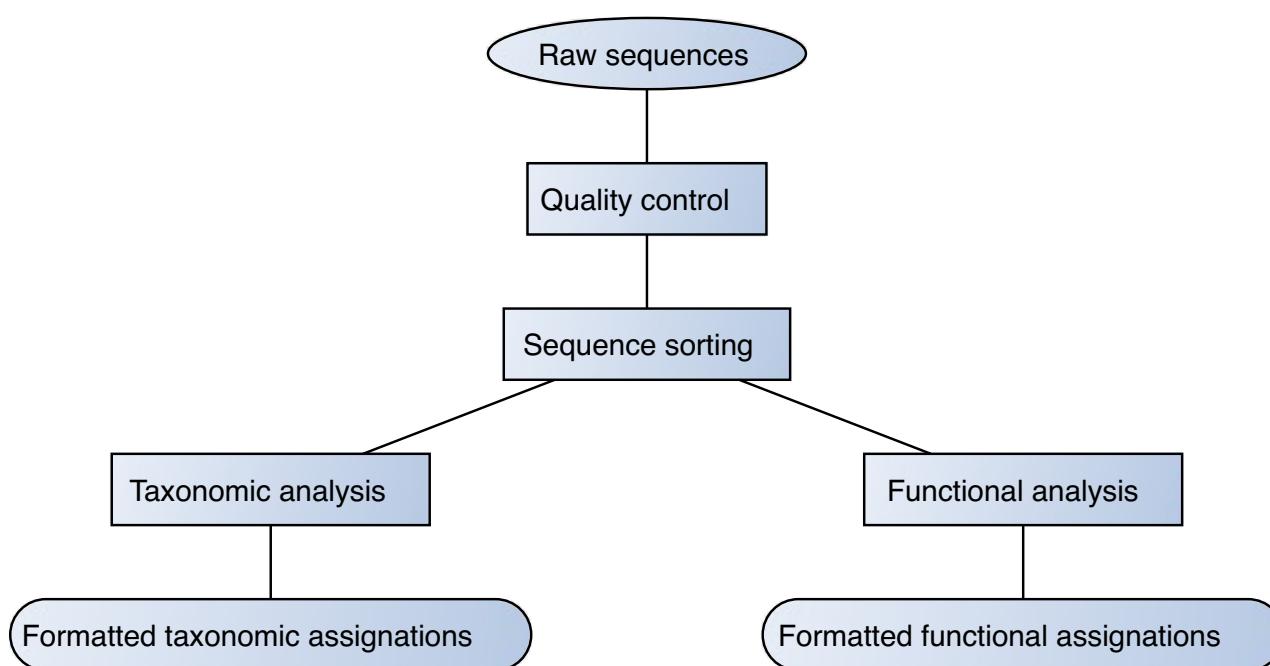
"human gut metagenome" search:

- 236 studies
- 1,545 runs, ...

i Dispersed and not comparable information

Must

- Collect datasets
- Analyze them given a standard workflow



Existing tools

QIIME, MG-RAST, EBI metagenomics, MetAMOS, ...

But none of them follows all the requirements:

Existing tools

QIIME, MG-RAST, EBI metagenomics, MetAMOS, ...

But none of them follows all the requirements:

- Analyze datasets given the standard workflow

Existing tools

QIIME, MG-RAST, EBI metagenomics, MetAMOS, ...

But none of them follows all the requirements:

- Analyze datasets given the standard workflow
- Use gut microbiota specific databases

Existing tools

QIIME, MG-RAST, EBI metagenomics, MetAMOS, ...

But none of them follows all the requirements:

- Analyze datasets given the standard workflow
- Use gut microbiota specific databases
- Combine user-friendly interface and command-line

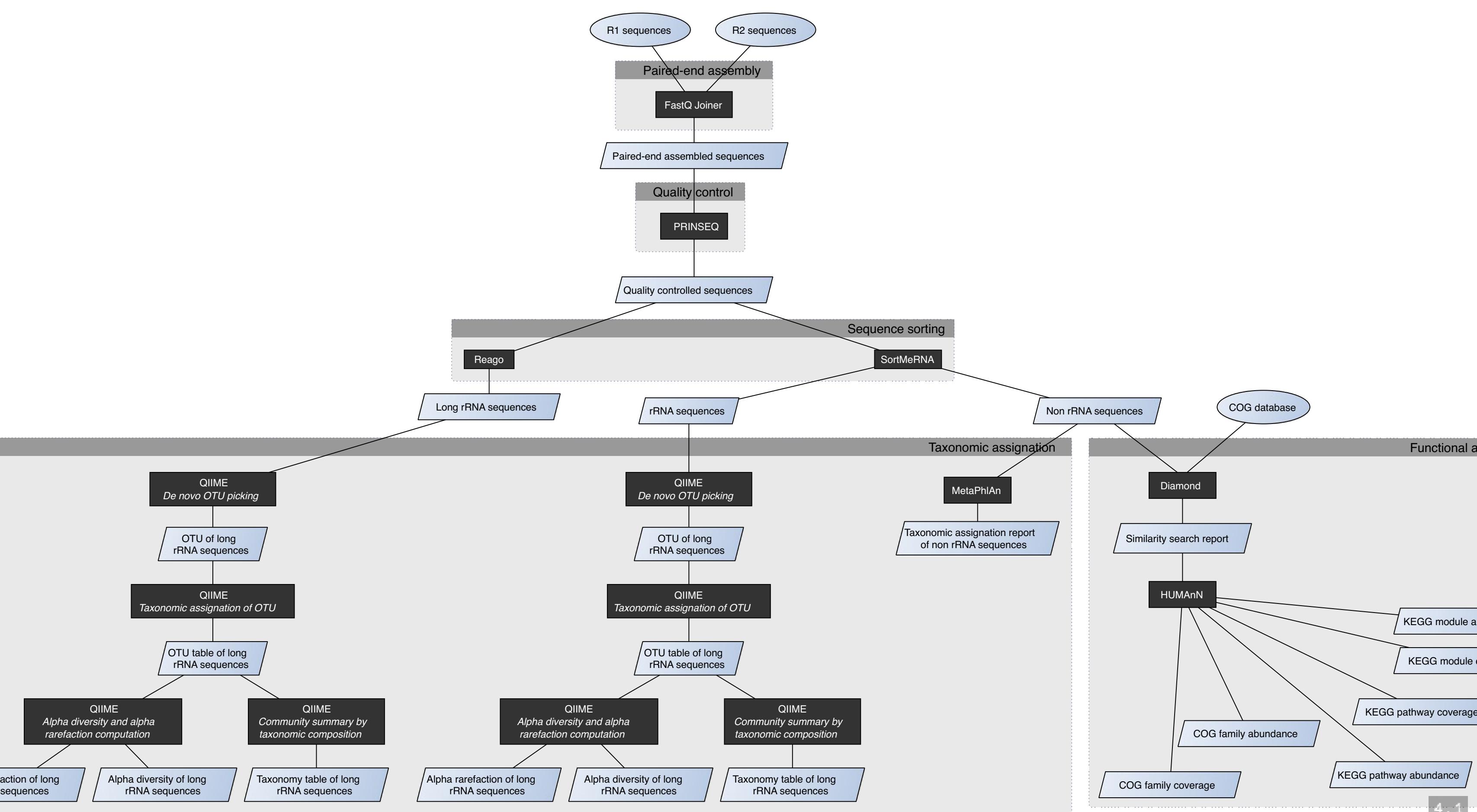
ASaiM

Auvergne Sequence analysis of intestinal Microbiota

An environment to analyze metagenomic and
metatranscriptomic sequences from gut microbiota

Components

- Expert database
- Web interface
- Framework



ASaiM framework

Bioinformatics framework to generate workflows for
analyses of gut microbiota data

Main requirements

- Generation of workflow with numerous tools
- Easy use
- Flexibility and modularity
- Incorporation of wanted/needed tools and databases

Things I tried

Things I tried

- Simple Python scripts

Things I tried

- Simple Python scripts
- Workflow managers such as Luigi, Airflow, ...

Things I tried

- Simple Python scripts
- Workflow managers such as Luigi, Airflow, ...
- Homemade approach
 - Configuration file
 - Workflow description
 - Web interface for generation
 - Python scripts to execute workflow



Galaxy

Fit main requirements

- Generation of workflow with numerous tools
- Easy use
- Flexibility and modularity
- Incorporation of wanted/needed tools and databases

ASaIM Galaxy instance

The screenshot shows a web browser window for a Galaxy instance at 127.0.0.1:8080. The main content area displays the ASaIM logo and a brief description of the environment. The left sidebar contains a list of tools categorized under 'COMMON TOOLS' (Get Data, Manipulate text, Filter and Sort, Join, Subtract and Group, Extract data), 'PRETREATMENTS' (Assemble paired-end sequences, DerePLICATE, Detect chimeric sequences, Cluster sequences, Manipulate RNA, Search similarity), 'TAXONOMIC ASSIGNATION' (Assign non rRNA sequences), and 'FUNCTIONAL ASSIGNATION' (Analyze metabolism). The top navigation bar includes links for Analyze Data, Workflow, Shared Data, Visualization, Help, and User. The right sidebar shows an empty history panel with a message encouraging users to load their own data or get data from an external source.

ASaIM (*Auvergne Sequence analysis of intestinal Microbiota*) is an environment to analyze intestinal microbiota.

This environment is composed of:

- A framework to analyze gut microbiota data
- A database which takes an inventory of gut microbiota data from public data repositories
- A Web interface to query the database

[Read more about ASaIM...](#)

The ASaIM framework is a Galaxy instance to generate workflows for analyzes of gut microbiota data. Follow the [Tutorial](#) to learn how to use ASaIM framework to analyze gut microbiota data.

This framework relies on numerous tools for data pretreatments, taxonomic analyses and functional analyses, and chosen databases (COG, ...) dedicated to process gut microbiota data. Several standard workflows are also proposed. For more information about the tools, databases, workflows and ASaIM framework in general, please read the [ASaIM documentation](#).

Galaxy is an open, web-based platform for data intensive biomedical research. The [Galaxy team](#) is a part of [BX](#) at [Penn State](#), and the [Biology](#) department at [Johns Hopkins University](#). The [Galaxy Project](#) is supported in part by [NHGRI](#), [NSF](#), [The Huck Institutes of the Life Sciences](#), [The Institute for CyberScience at Penn State](#), and [Johns Hopkins](#).

To launch the instance

- Get the code source from [GitHub](#)

```
$ git clone git@github.com:ASaiM/framework.
```

- Install the required dependencies
- Launch the instance

```
$ cd framework  
$ ./src/launch_galaxy.sh
```

- Browse it on <http://127.0.0.1:8080/>

Behind the magic

Shell scripts to configure the instance

1. Get latest revision of Galaxy from GitHub
2. Prepare databases and local tools
3. Configure with
 - Custom configuration files
 - Wanted tools
 - Wanted databases
4. Launch Galaxy

Tools

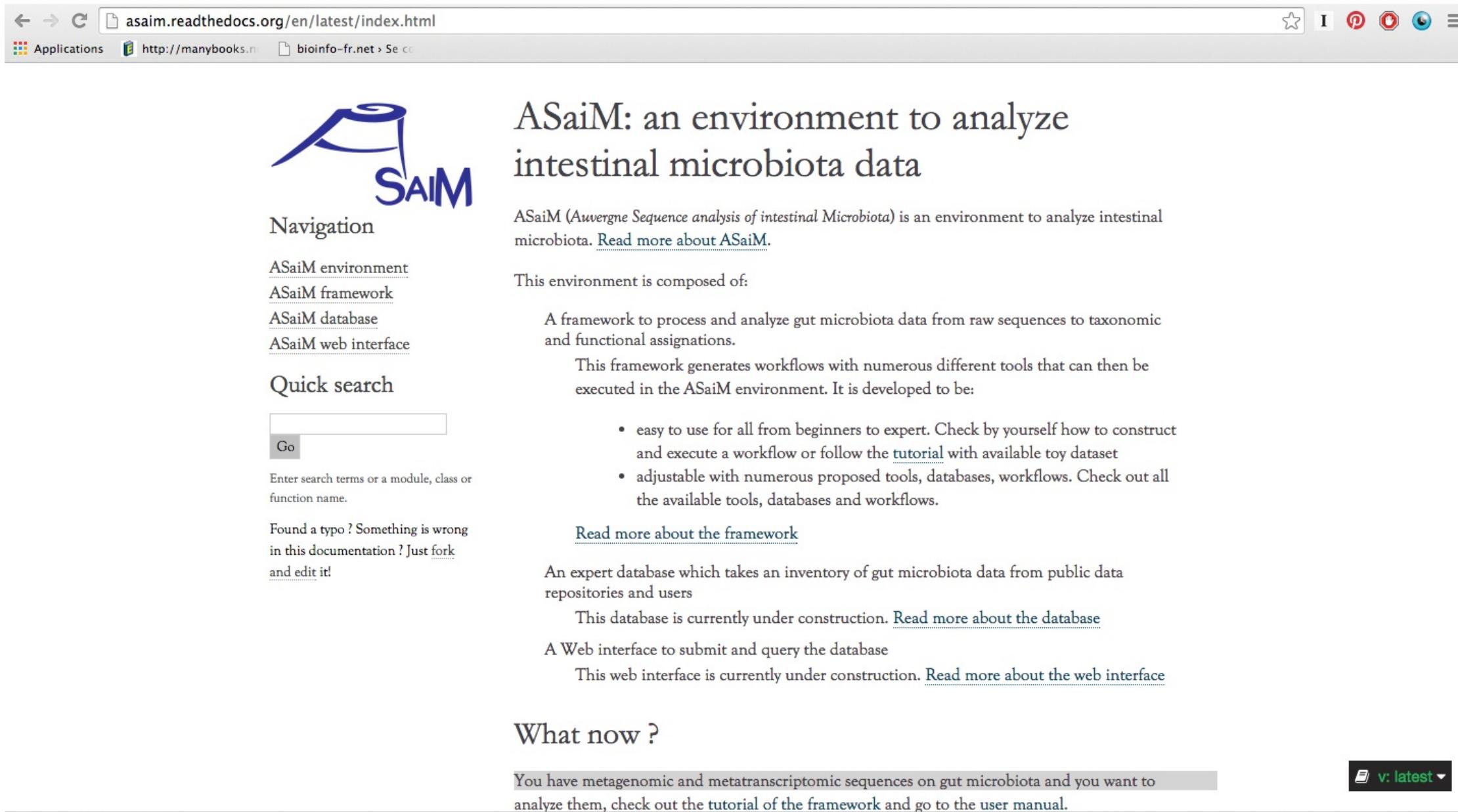
- From standard Galaxy instance
- From ToolShed
- Developed wrappers
 - Planemo
 - Integration in test Tool Shed
 - <https://github.com/ASaiM/galaxytools>

Workflows

Databases

- SortMeRNA ribosomal databases
- COG
- RefSeq
- *Catalog of reference genes in the human gut microbiome from Li et al. (2014)*
- Greengenes

Documentation



The screenshot shows a web browser window with the URL asaim.readthedocs.org/en/latest/index.html in the address bar. The page content is as follows:

ASaIM 

Navigation

- [ASaIM environment](#)
- [ASaIM framework](#)
- [ASaIM database](#)
- [ASaIM web interface](#)

Quick search

Enter search terms or a module, class or function name.

Go

Found a typo? Something is wrong in this documentation? Just fork and edit it!

ASaIM: an environment to analyze intestinal microbiota data

ASaIM (*Auvergne Sequence analysis of intestinal Microbiota*) is an environment to analyze intestinal microbiota. [Read more about ASaIM](#).

This environment is composed of:

- A framework to process and analyze gut microbiota data from raw sequences to taxonomic and functional assignations.
- This framework generates workflows with numerous different tools that can then be executed in the ASaIM environment. It is developed to be:
 - easy to use for all from beginners to expert. Check by yourself how to construct and execute a workflow or follow the [tutorial](#) with available toy dataset
 - adjustable with numerous proposed tools, databases, workflows. Check out all the available tools, databases and workflows.

[Read more about the framework](#)

An expert database which takes an inventory of gut microbiota data from public data repositories and users

This database is currently under construction. [Read more about the database](#)

A Web interface to submit and query the database

This web interface is currently under construction. [Read more about the web interface](#)

What now ?

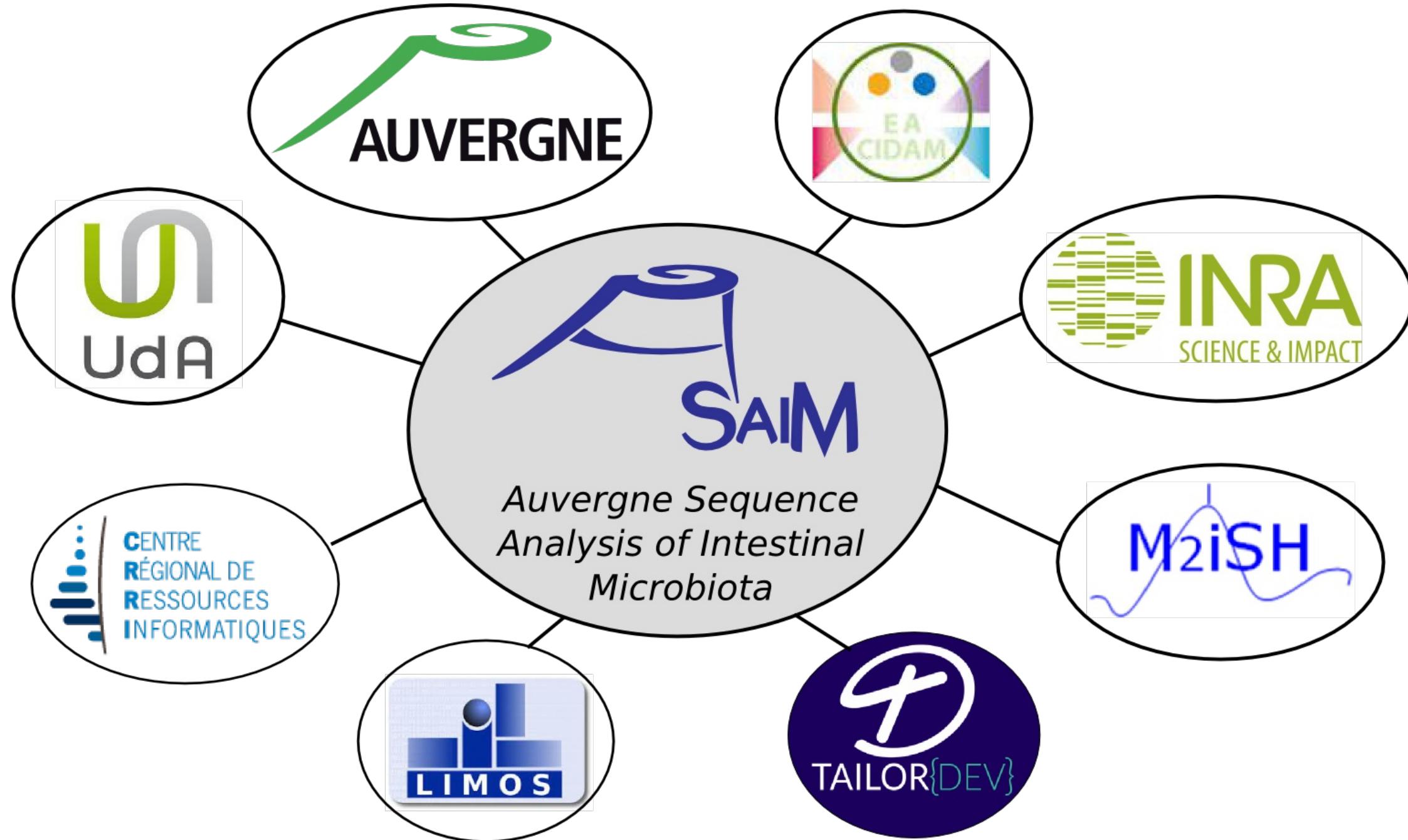
You have metagenomic and metatranscriptomic sequences on gut microbiota and you want to analyze them, check out the [tutorial of the framework](#) and go to the [user manual](#).

v: latest ▾

<http://asaim.readthedocs.org/>

To do

- ⊖ Automatize the configuration and deployment of the instance with Ansible
- ⊖ Add tools in development, databases, workflows
 - Validate workflows on datasets (local, mock, ...)
- ⊖ Integrate tools and workflows to the ToolShed
- ⊖ Automatize tool integration from ToolShed with Ansible
- ⊖ Complete the documentation



Thank You.

Questions?

<http://asaim.github.io>

-  bebatut.fr
-  github.com/bebatut
-  twitter.com/bebatut