

# Galaxy

Basic user manual of the Galaxy test installation of the Mr.Symbiomath project hosted by



Version 1 - 10/05/2016  
*Esteban Pérez-Wohlfeil*  
*Jose A. Arjona Medina*  
*Oscar Torreno*  
*Eugenia Ulzurrun*  
*and Oswaldo Trelles*

## Content of this document

[Getting used to Galaxy](#)

[Why Galaxy?](#)

[How do I start?](#)

[Uploading data and running modules and workflows](#)

[Retrieving produced data](#)

## Getting used to Galaxy

In this tutorial we will show, step by step, how to access our Galaxy instance and its basic usage. The aim is allowing users carrying out their own analyses using the mentioned instance. The paper published in the Future Generation Computer Systems journal [1] describes in detail the Galaxy deployment performed within the Mr.Symbiomath project. Please refer to the authors with any question.

### Why Galaxy?

[Galaxy](#) is an open, web-based platform for data intensive biomedical research. Whether on the [free public server](#) or [your own instance](#), you can perform, reproduce, and share complete analyses. After a careful analysis of the advantages and disadvantages, we chose using galaxy in the project, since, as a user, you will be able to both reproduce our galaxy instance and install it locally, or to use it via a Web browser on our server.

### How do I start?

Access <http://pistacho.ac.uma.es/galaxy/> and log in using the following guest user credentials:

User:            guest

Password:       guest

You may want to create a personal account for your own and private use.

Figure 1: How to log in our galaxy instance

After logging in, you will be redirected to the main site of the galaxy webpage.

Figure 2: Main page of the galaxy instance

There are three main sections that we should briefly address to improve the user's experience within this tutorial:

1. On the left side, section (1) lists all the available tools and workflows for data processing, which is divided into two subsections:
  - a. Section (1 a) shows all software modules organized by packages. For example, clicking on the package "METAGECKO" will display a drop down list containing all the individual modules needed for the METAGECKO workflow (part of the GECKO suite). Further clicking on any of the individual modules will allow for single data processing.
  - b. Clicking on section (1 b) will lead us to all available workflows. Each workflow is composed of several interconnected individual modules.
2. The middle view is reserved to display modules and workflows when requested, showing the required parameters for the executions.
3. Section (3) on the right side is the history manager. The history manager allows to see uploaded datasets, along with running executions of modules and workflows, and the datasets that these will generate. Galaxy keeps track of many histories, in order to switch and restore between them with no difficulty, helping the user to carry out different experiments at once.

## Uploading data and running modules and workflows

The first step to start processing is definitely gathering and uploading data. For such goal, we will click in the **Get Data** label (Section 1 a) and then on *Upload file from your computer*. This will allow us to upload files of size at most 1 GB. For files of larger size, please refer to the Galaxy Project tutorial itself [2].

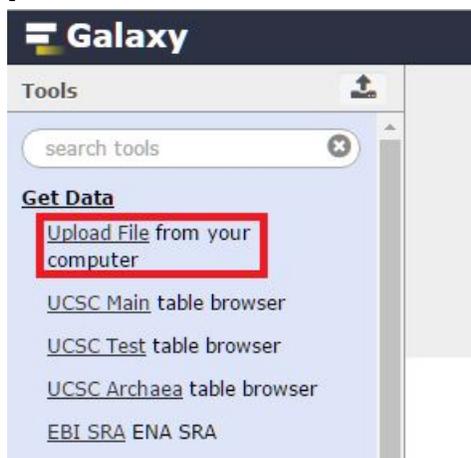


Figure 3: Opening the upload file dialog.

This will bring up the *Upload file* dialog. See the picture below:

#### Download from web or upload from disk

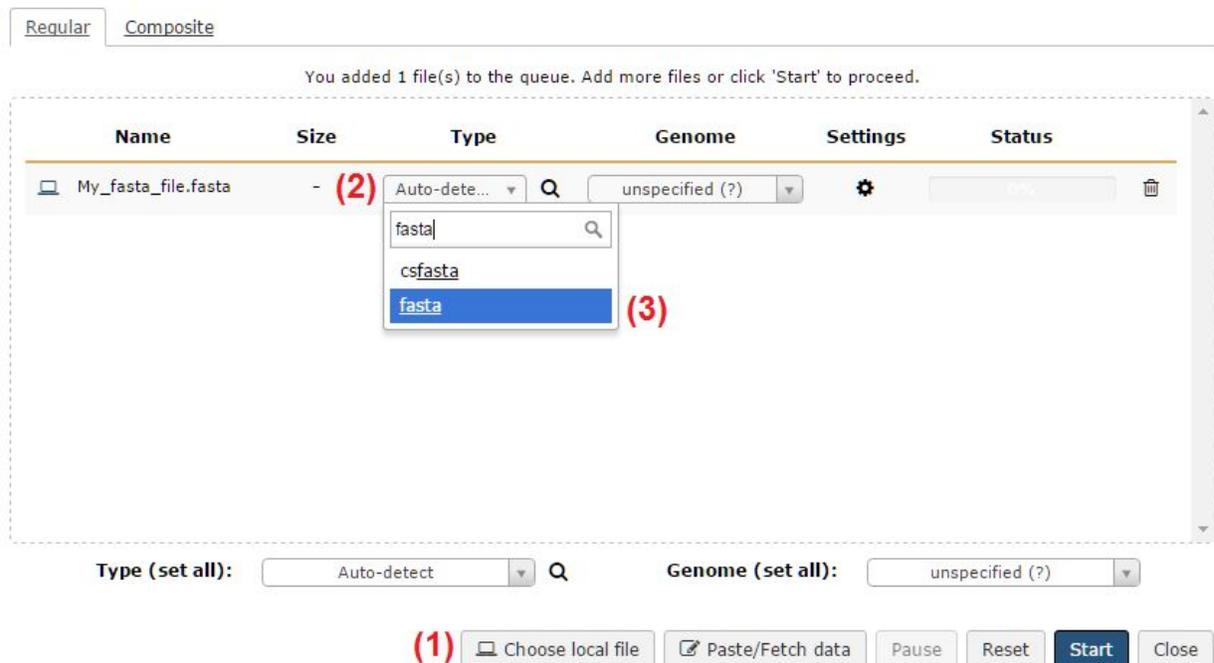


Figure 4: Uploading a file and choosing the format.

We will have to follow these instructions for each file we wish to add:

1. First click on the *Choose local file* button, marked in red with (1) in the picture above. This will let us select a particular file from our computer.
2. Secondly, click on (2), namely the *Auto-detect* form. This will open a small drop down menu. Manually enter the format of the file as shown in (3). For example, for a **.fasta** file type **fasta** in the search box. Galaxy is able to detect some file formats automatically, however, we recommend users to manually select them to avoid confusion.

Repeat the steps contained above for as many files as you wish to add. Then click on the blue *Start* button, on the bottom-right side of the panel. The upload will start and will take more or less time depending on the size of the files and the speed of your internet connection.

Once the files are completely uploaded, click on the *Close* button next to the *Start* button. All files will be immediately added to the current history and will be available for every workflow or software module, as shown in the picture above:

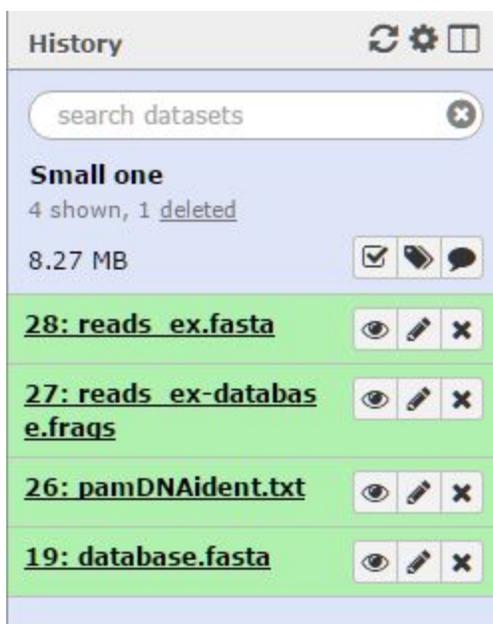


Figure 5: The history panel (Section 3) shows the current history and the just uploaded contents.

The history panel allows for multiple operations on datasets, some of them are (more details available in the Galaxy Project wiki):

- Download any file, as well as the whole history exported to a compressed format
- Visualise any file
- Edit file's attributes
- Delete datasets

Now that we have some datasets at our disposal, we can conveniently run individual modules and/or workflows. For example, choose METAGECKO on section (1 a). A drop down list of modules will be displayed (marked with (1) in the picture contained below):

The screenshot shows the Galaxy web interface. On the left, a 'Tools' sidebar lists several tools. 'METAGECKO' is highlighted with a red '(1)'. Below it, 'MGRindex' is listed with a red '(2)'. The main panel displays the configuration for the 'MGRindex' tool. It has a title 'MGRindex Creates indexes for the fast access of fasta files (Galaxy Version 1.0.0)'. Under the 'Fasta file' section, a dropdown menu is set to '28: reads\_ex.fasta' with a red '(3)' next to it. Below the dropdown is an 'Execute' button with a checkmark and a red '(4)' next to it.

Figure 6: Running a module using the datasets available in the current history.

Then click on (2) (for example), *MGRindex*, a tool that creates sorted and incremental indexes for metagenome processing. On section (2) (middle window) will appear the information related to the module/workflow. For this example, *MGRindex* only requires one parameter, a fasta file. Click on (3) to select one of the available datasets in the current history, and click on (4) to execute it.

The execution will now appear on the history panel:

The screenshot shows the Galaxy history panel. On the left, a green notification box with a checkmark icon contains the text: '1 job has been successfully added to the queue - resulting in the following datasets:'. Below this, two new datasets are listed: '29: Non sorted binary index of reads\_ex.fasta' and '30: Sorted binary index of reads\_ex.fasta' with a red '(1)' next to the second one. Below the list, there is a paragraph of text: 'You can check the status of queued jobs and view the resulting data by refreshing the History pane. When the job has been run the status will change from 'running' to 'finished' if completed successfully or 'error' if problems were encountered.' On the right, the 'History' panel shows a list of datasets. The top dataset is 'Small one' (8.27 MB). Below it are '30: Sorted binary index of reads\_ex.fasta' (2), '29: Non sorted binary index of reads\_ex.fasta', '28: reads\_ex.fasta', '27: reads\_ex-database.frag', '26: pamDNAident.txt', and '19: database.fasta'.

Figure 7: Adding jobs to the queue and generating new datasets.

Datasets that will be generated are displayed on (1). The job, will be displayed in the queue (See figure contained above). A yellow-beige background color (2) indicates that the job is currently being processed and has not finished yet. A green color indicates the job finished and generated the dataset successfully, while a red background indicates that something went wrong.

## Retrieving produced data

Generated datasets will be ready to download once the individual module or workflow has finished running. For instance, you can download one single file by clicking on its name (1) and then on the floppy disk icon (2). If there are more than a few datasets and you wish to download them all automatically, another possible procedure is to click on the top-right gear (3) and then on *Export history to file* (4).

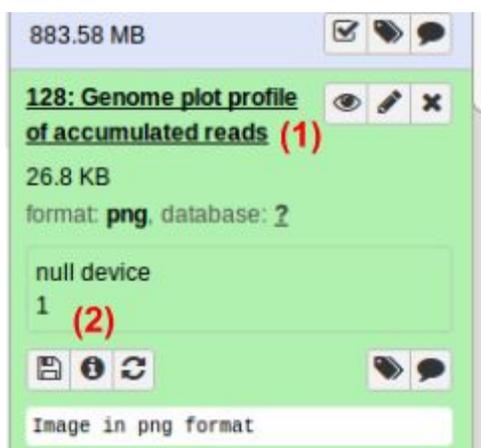


Figure 8: Downloading a single dataset.

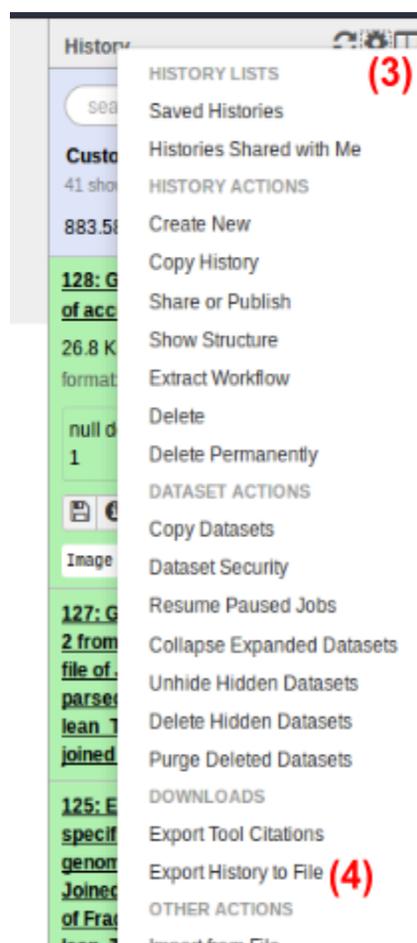


Figure 9: Download full history at once.



## References

1. Krieger, M. T., Torreno, O., Trelles, O., & Kranzlmüller, D. (2016). Building an open source cloud environment with auto-scaling resources for executing bioinformatics and biomedical workflows. *Future Generation Computer Systems*.
2. <https://wiki.galaxyproject.org/FTPUpload>