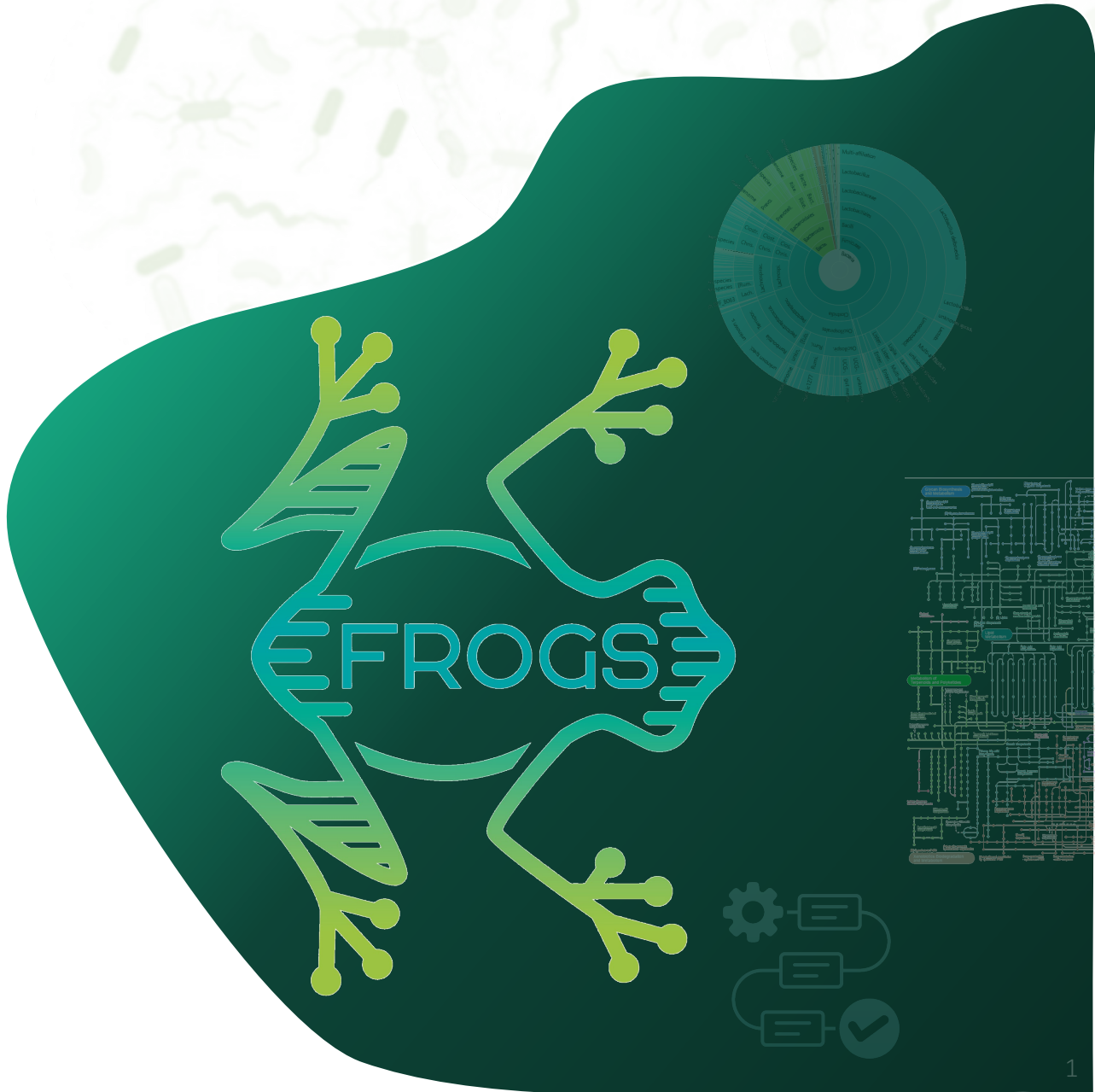
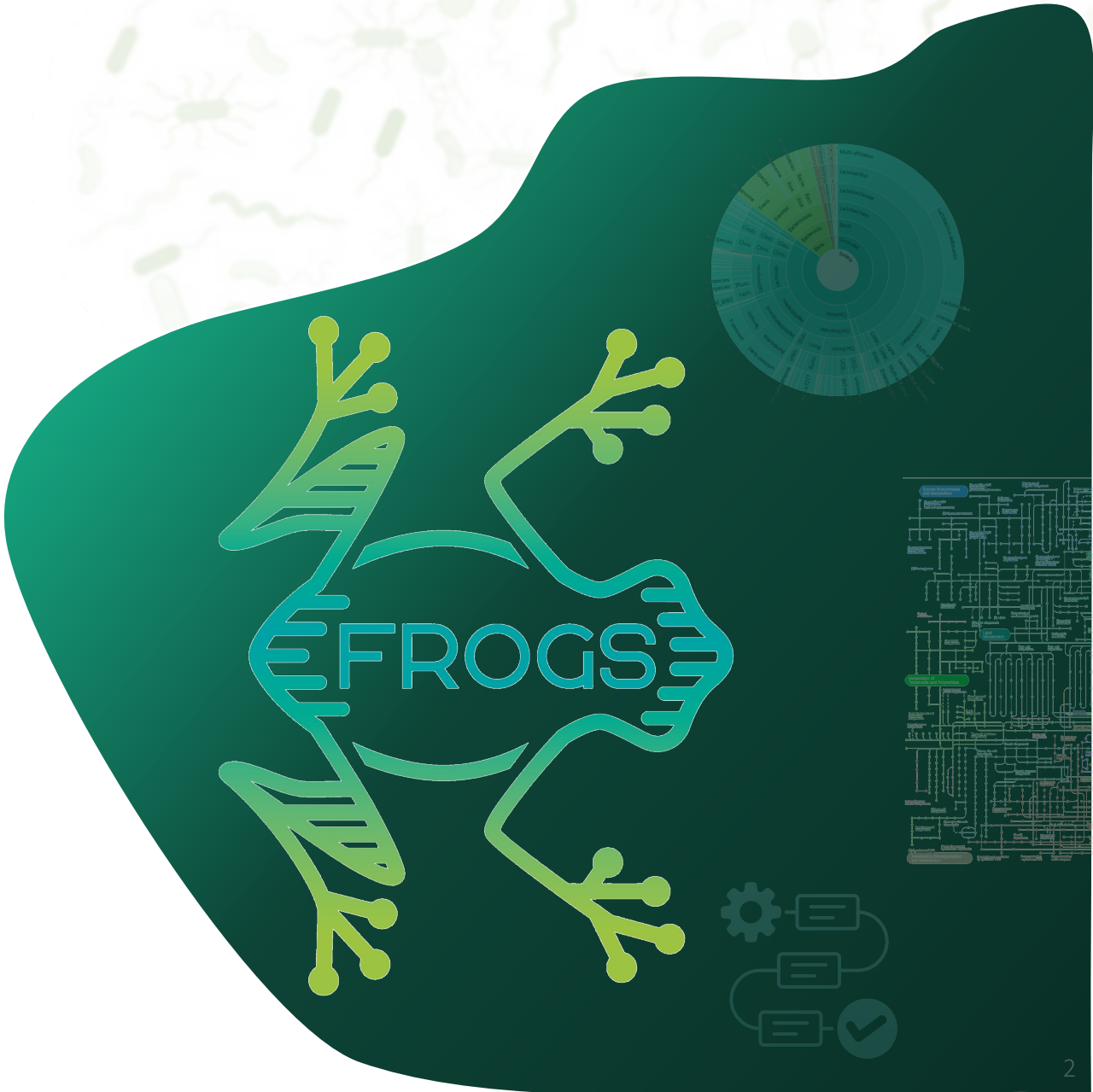
Get Started with Galaxy

Lucas AUER, Gabryelle Agoutin,
Maria BERNARD Géraldine PASCAL,
Maëlle POMIÈS & Olivier RUÉ



Get Started with Galaxy

Overview of the Galaxy platform



Overview of the Galaxy platform

FROGS is hosted on the **Galaxy platform**, an open-source web environment designed to make bioinformatics analysis **accessible, transparent, and reproducible**. Originally developed at Penn State University, Johns Hopkins University, Oregon Health & Science University, and the Cleveland Clinic, Galaxy has grown thanks to substantial contributions from an active international community.

As a web-based platform, Galaxy allows users to perform complex analyses without installing any software or writing code. Its versatile workflows make it easy to combine and reproduce tools for different sequencing projects.

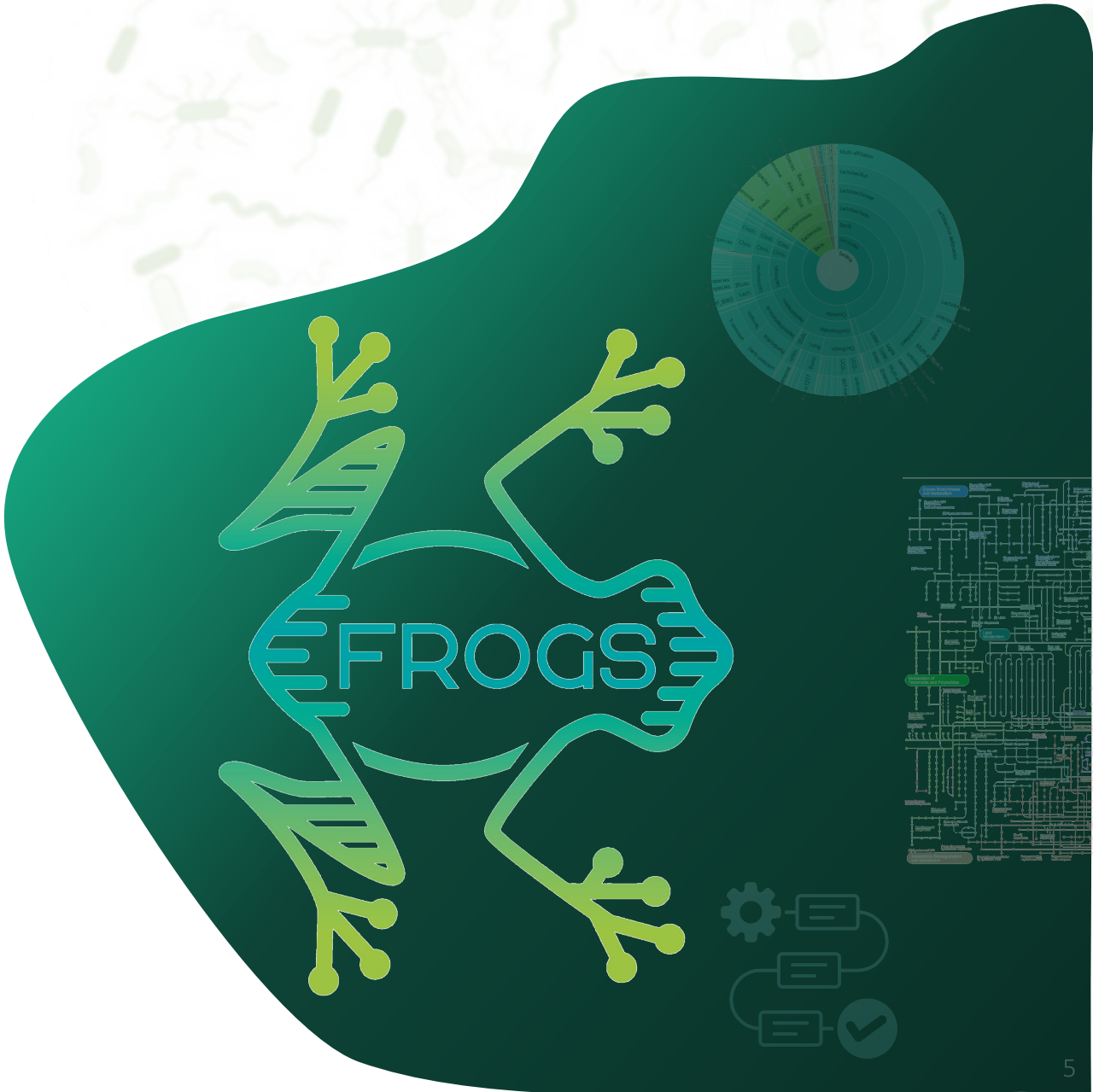
Overview of the Galaxy platform

- **Accessibility:** Users with no programming experience can easily upload or retrieve data, run workflows like FROGS, and visualize results directly in their browser.
- **Reproducibility:** Galaxy automatically records all parameters and inputs, so any analysis can be fully understood, reviewed, or repeated by others.
- **Transparency:** Users can share or publish their entire analyses — including histories, workflows, and visualizations — or even create Galaxy “Pages” to document methods for publications.

Together, these features make Galaxy an ideal platform for deploying FROGS and ensuring that every analysis is both reliable and shareable.

Get Started with Galaxy

Accessing the platform

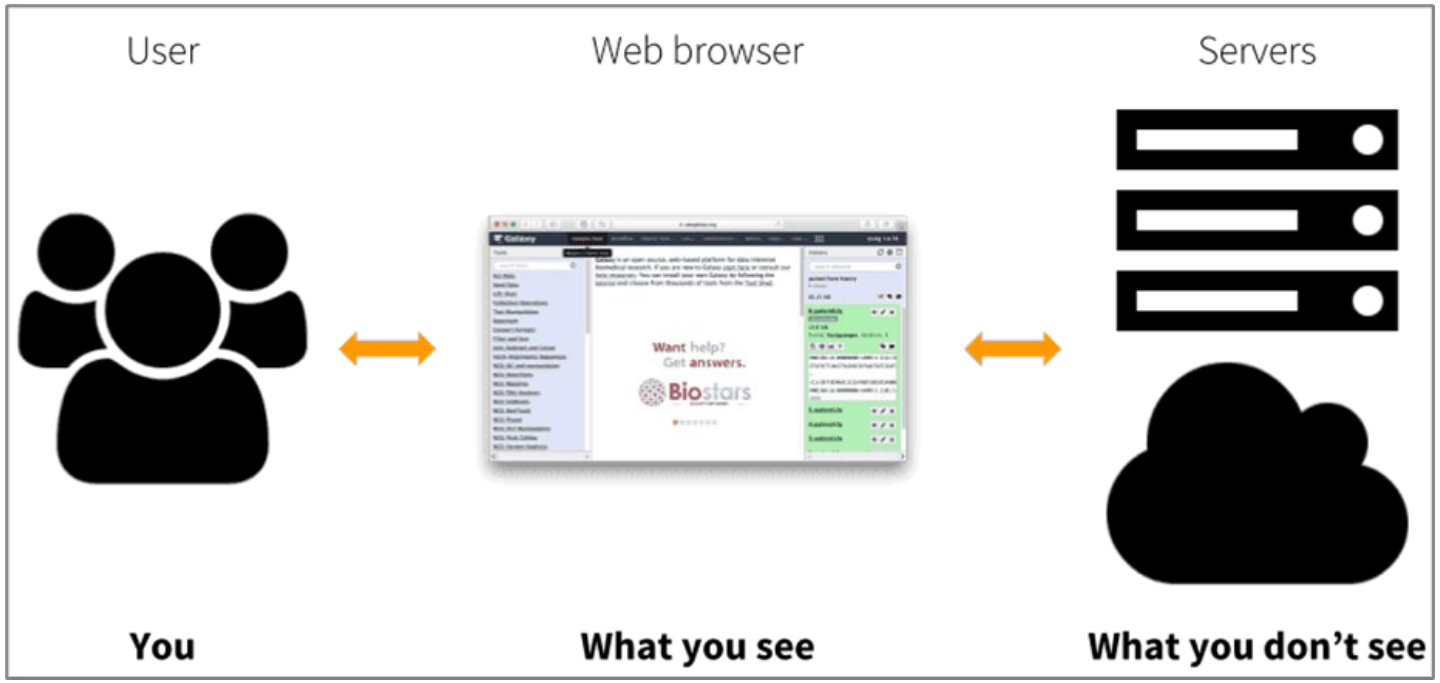


Accessing the platform

When using FROGS on the Galaxy platform, nothing runs on your own computer.

All computations are executed on remote servers, so your computer is only used to access the web interface.

This means you can run analyses without worrying about your computer's power or operating system — you only need a web browser and an internet connection.



FROGS is hosted on three main partner platforms

- Genotoul Bioinfo/Sigenae (INRAE Toulouse) https://vm-galaxy-prod.toulouse.inrae.fr
- Migale (INRAE Jouy-en-Josas) https://galaxy.migale.inra.fr
- IFB Core (French Bioinformatics Institute) https://metabarcoding.usegalaxy.fr

⚠ Important

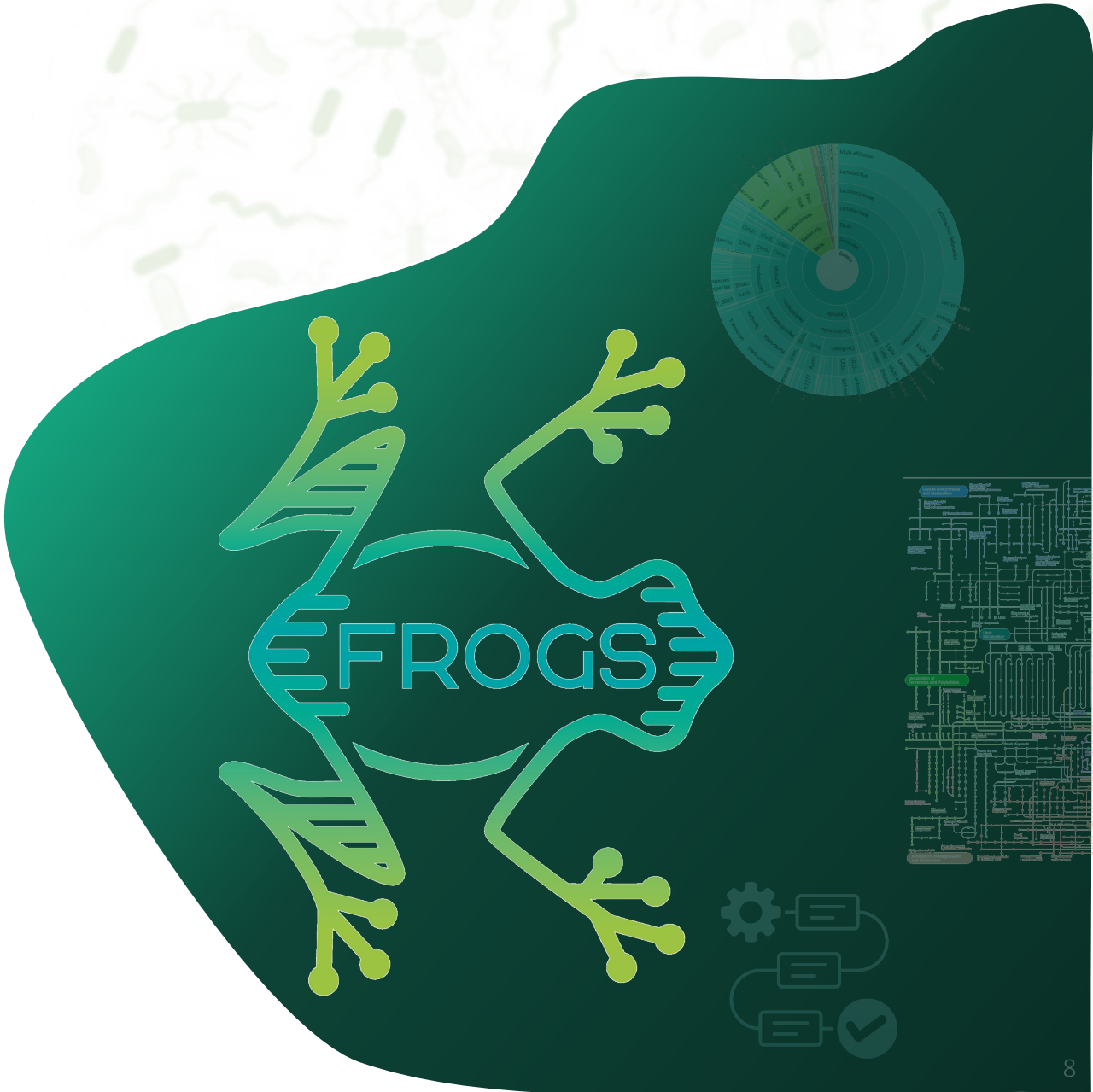
Each Galaxy platform operates independently. FROGS is available on several Galaxy instances but these platforms are not connected to each other.

This means that:

- Your data are stored only on the platform you are using and are not shared with other Galaxy servers.
- Tools and workflows may differ slightly between platforms; some tools available on one FROGS Galaxy instance may not exist elsewhere.
- Each instance relies on its own computing cluster, so job performance and storage capacity can vary from one platform to another.

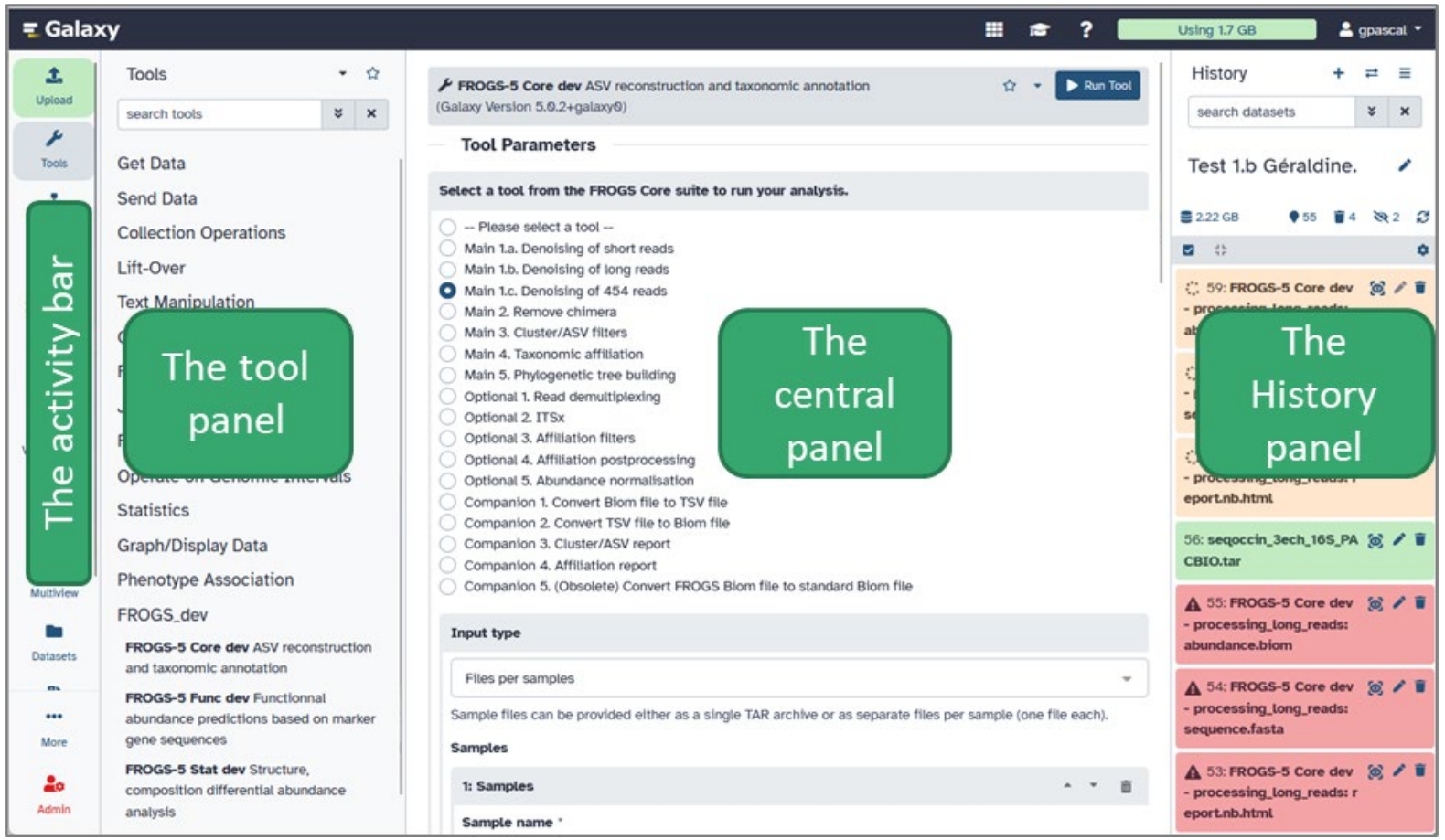
Get Started with Galaxy

The Galaxy Interface

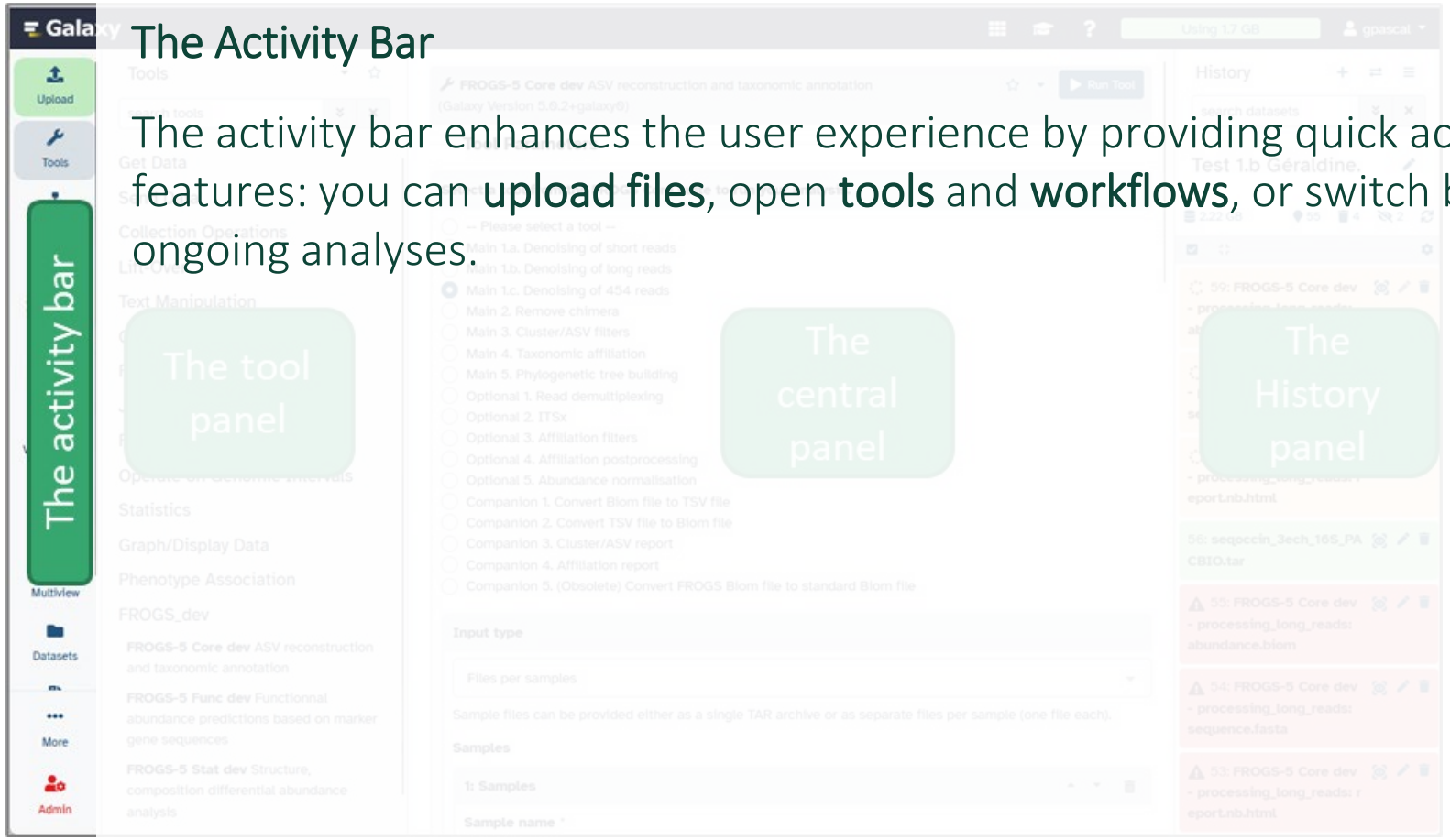


The Galaxy Interface Layout

The Galaxy web interface is divided into **four main panels**, each with a distinct role.



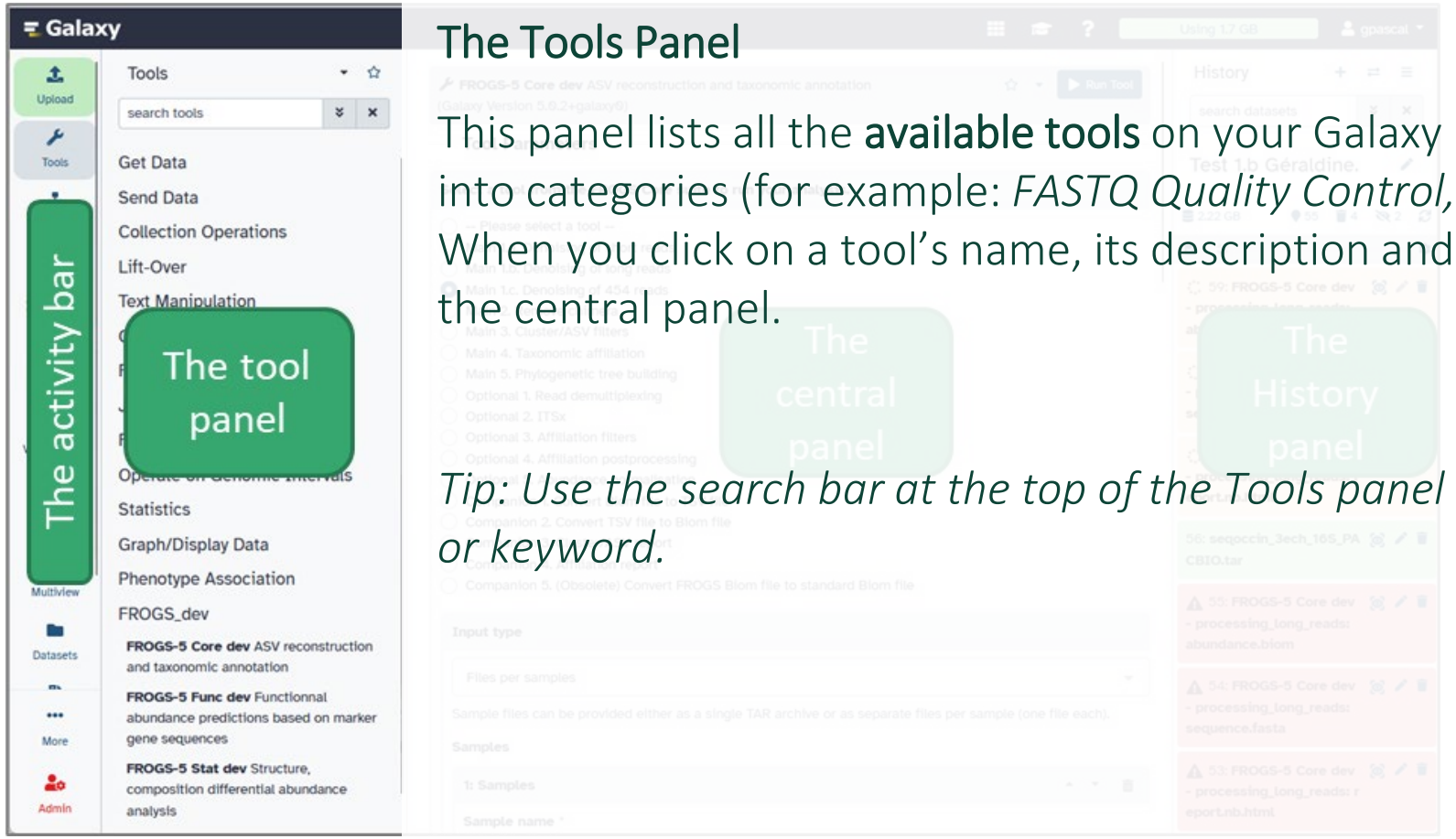
The Galaxy Interface Layout



The Activity Bar

The activity bar enhances the user experience by providing quick access to essential Galaxy features: you can **upload files**, open **tools** and **workflows**, or switch between your **histories** and ongoing analyses.

The Galaxy Interface Layout

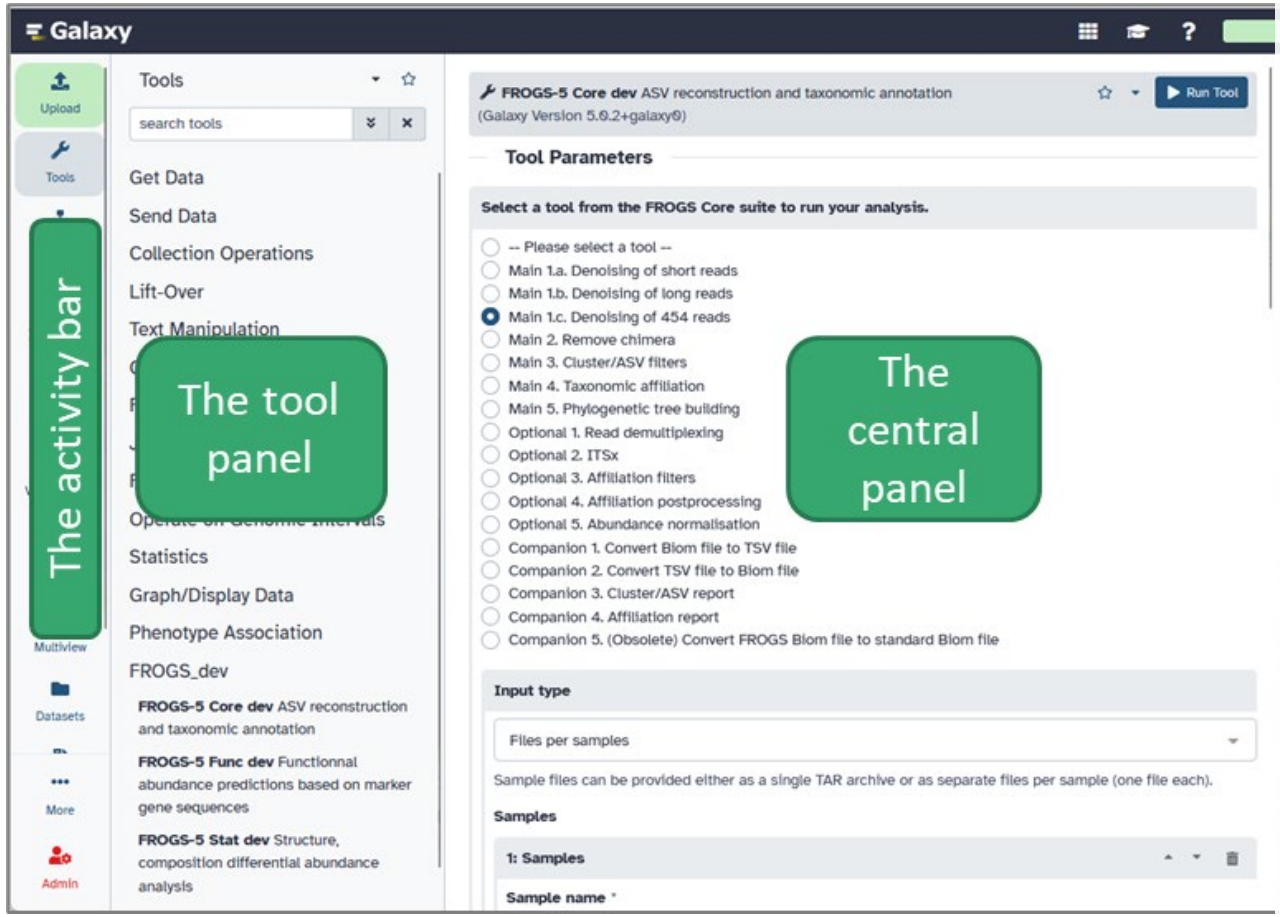


The Tools Panel

This panel lists all the **available tools** on your Galaxy platform. Tools are organized into categories (for example: *FASTQ Quality Control, FROGS-5 Core, Qiime,* etc.). When you click on a tool's name, its description and parameter options appear in the central panel.

Tip: Use the search bar at the top of the Tools panel to quickly find any tool by name or keyword.

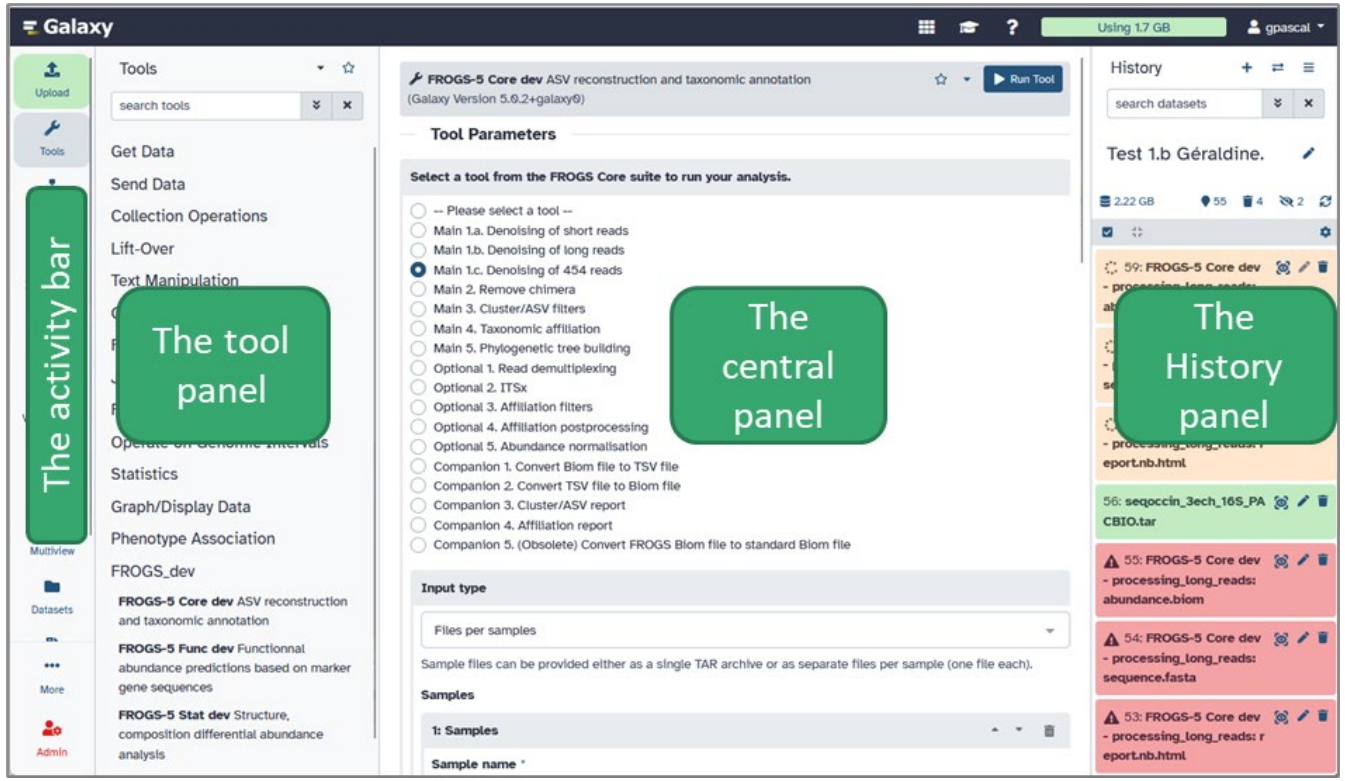
The Galaxy Interface Layout



The Central Panel (in the middle)

This is your **main working area**. It displays tool forms, results, visualizations, and documentation. Whenever you launch a tool or open a dataset, the details are shown here. You can also view the output of FROGS steps (for example, abundance tables, graphics, statistics report) directly in this area.

The Galaxy Interface Layout



The History Panel (on the right)

The History panel keeps track of all your **datasets, analyses, and results**. Each time you run a tool, its output appears here as a new dataset — color-coded to show its status:

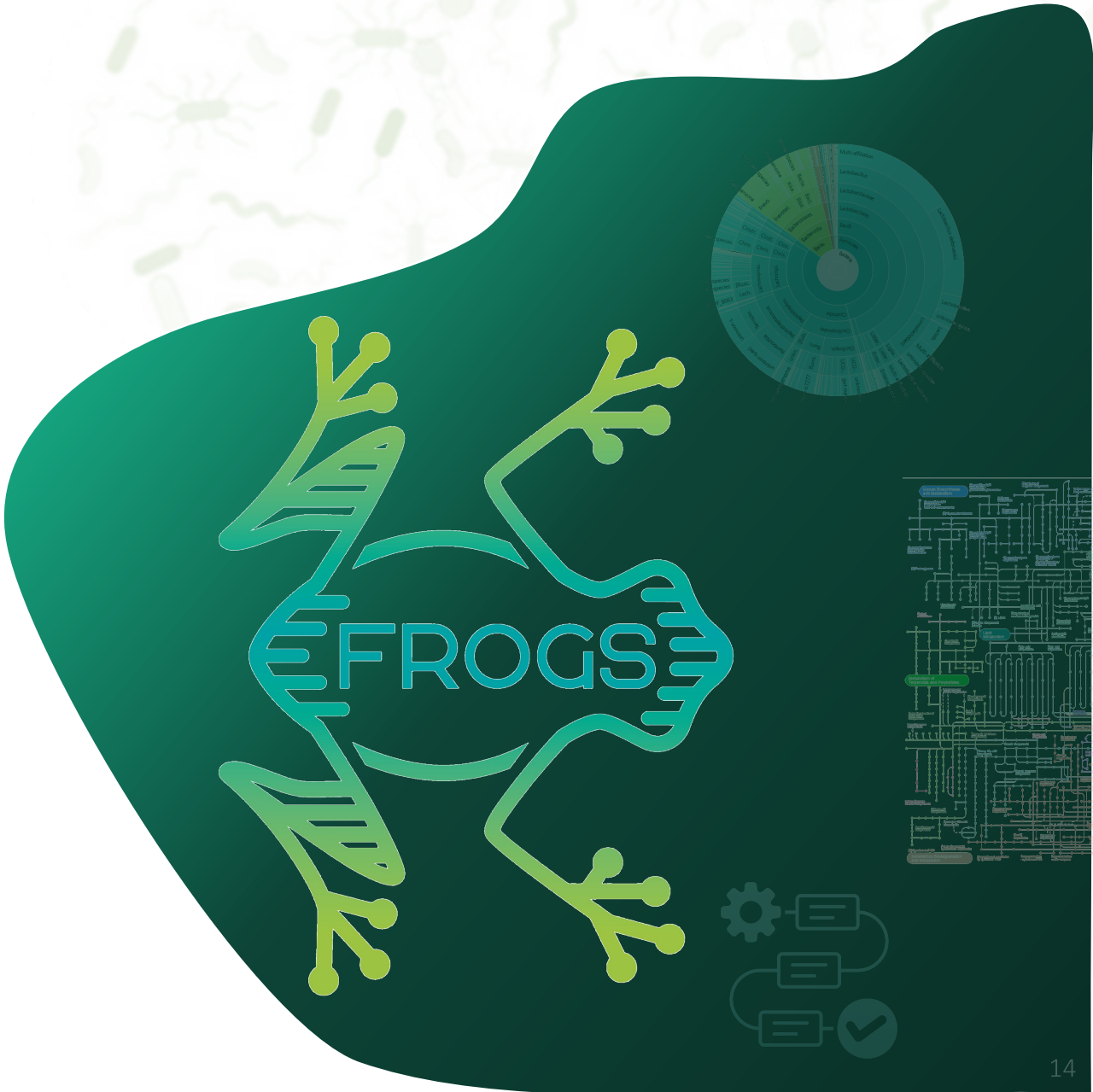
- Grey = waiting for server response
- Orange = running
- Green = finished successfully
- Red = failed

Histories are automatically saved and can be renamed, shared, duplicated, or downloaded.

Tip: You can click on a dataset name to view details or click the eye icon to preview its content.

Get Started with Galaxy

Create a *.tar* archive with your data



Why package your data into a .tar archive?

Uploading files one by one is slow, tedious, and error-prone.

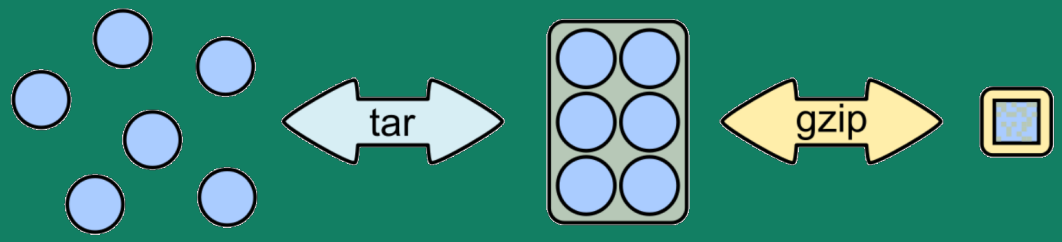
A better solution is to bundle your files into a single tar archive.

This makes the upload faster, cleaner, and much less likely to fail.

What is a .tar.gz file?

.tar bundles multiple files and folders in a single package

.gz compresses that package to reduce its size (effective if .fastq are not compressed)



How to package your data into a .tar achive?

Several solutions exists:

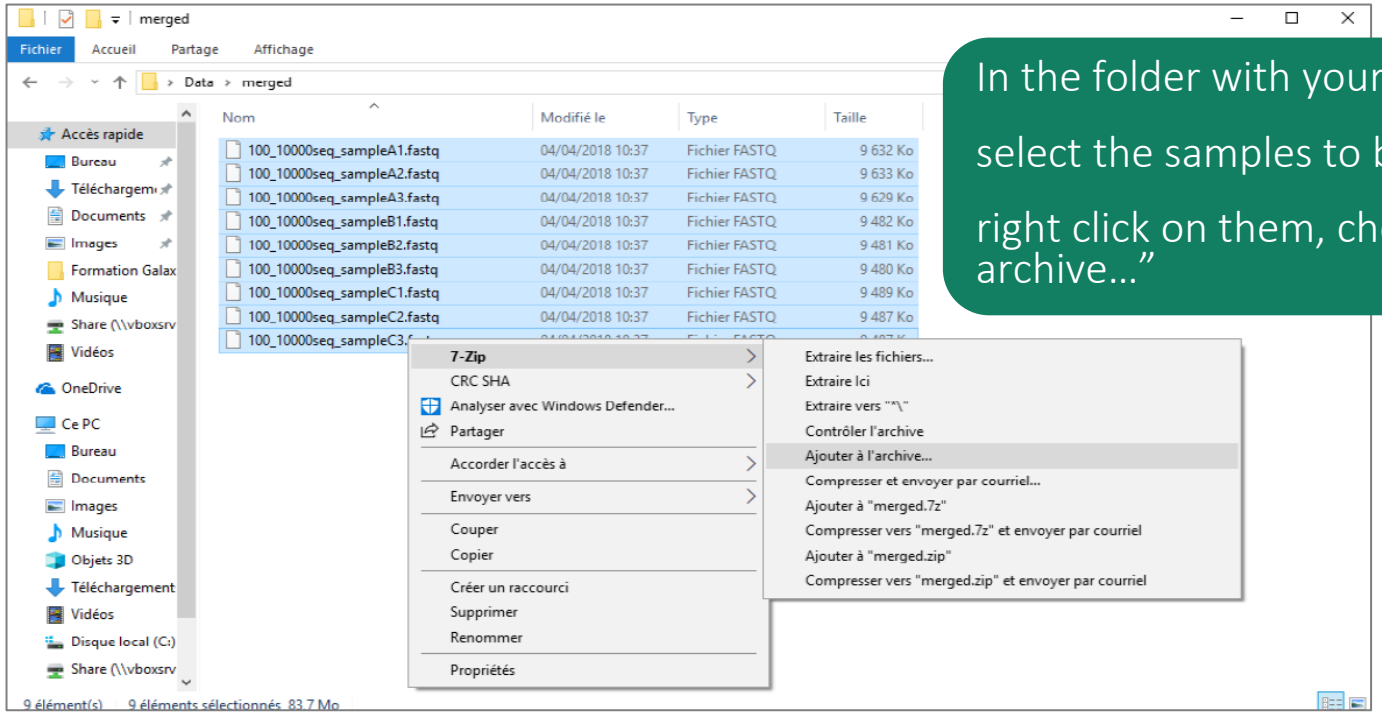
- without command lines: 7-zip (Windows only)
- with command lines: for macOS, Unix (or Windows with mobaXterm)

Create a tar file with 7-Zip (Windows)

7-ZIP is an open-source file archiver which can archive and compress files.

You can download it at: https://www.7-zip.org/

After installation, 7-Zip appears in the right-click context menu on files :

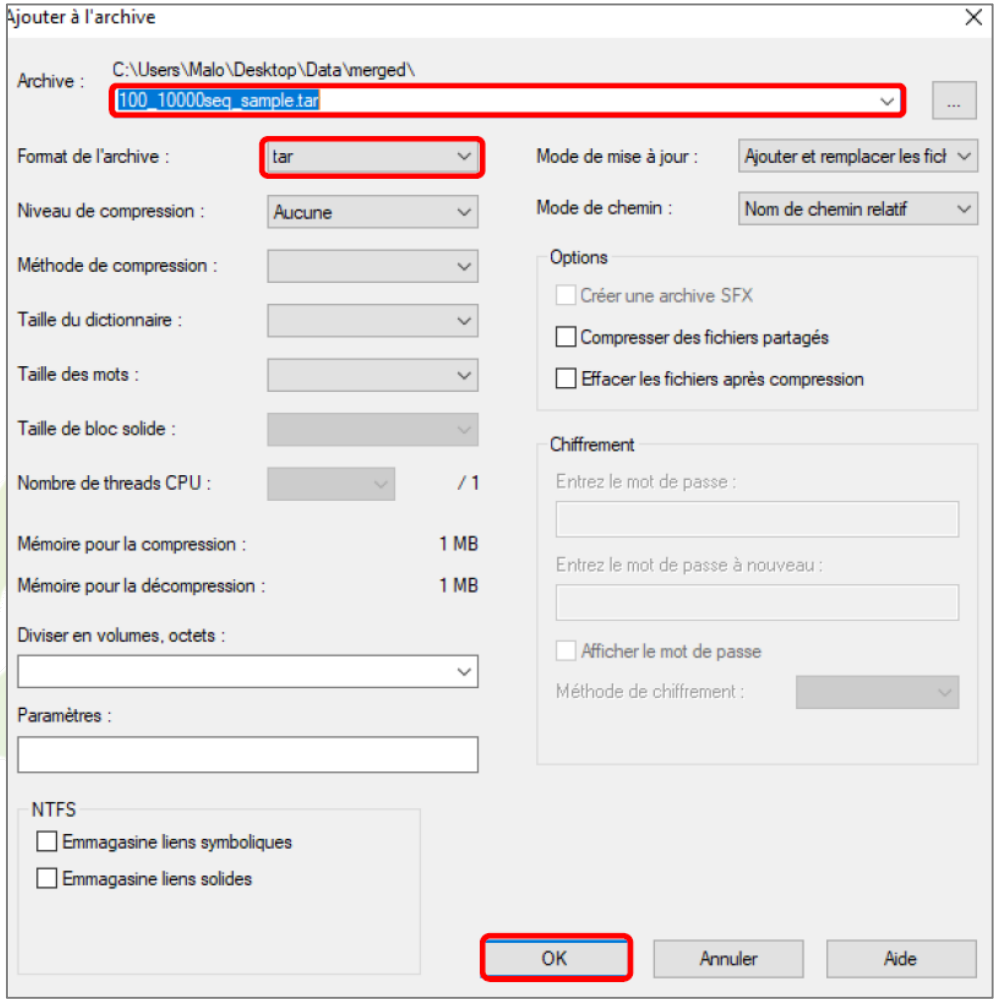


In the folder with your data,

select the samples to bundle

right click on them, choose 7-Zip, “add to archive…”

Create a tar file with 7-Zip (Windows)



Choose a name for your archive (ending with .tar)

Choose “tar” as archive format

Click on OK

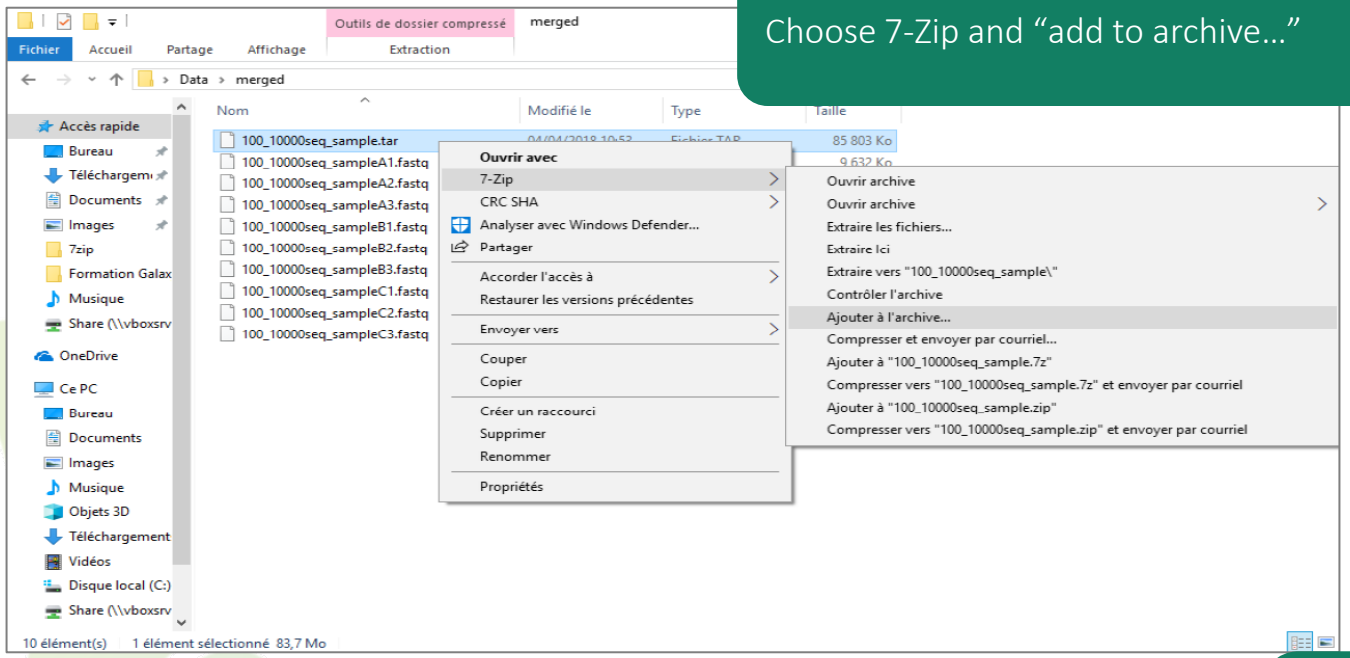
It creates a .tar archive which is not compressed.

If it needs to be compressed
(if the .fastq files were not already compressed in .fastq.gz format) it is required to compress it in a second step

Create a tar file with 7-Zip (Windows)

Right click your newly created .tar archive.

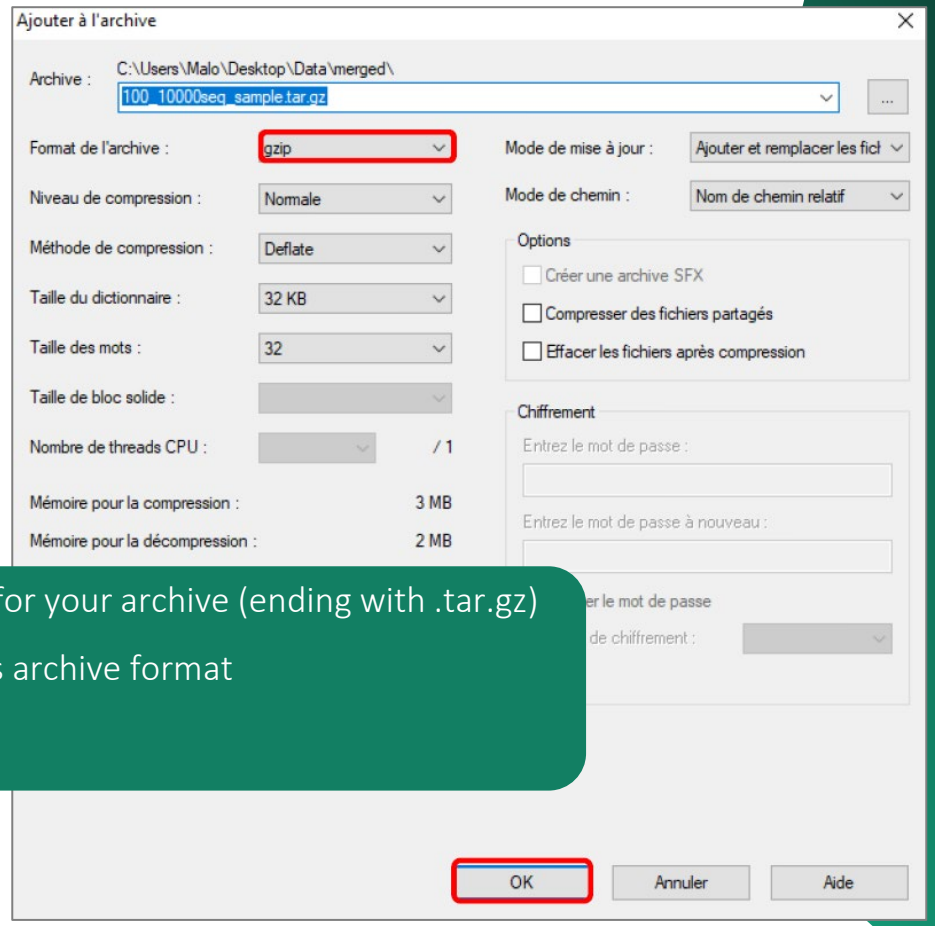
Choose 7-Zip and “add to archive…”



Choose a name for your archive (ending with .tar.gz)

Choose “gzip” as archive format

Click on OK



You now have a .tar.gz unique file containing all your data.

Create a tar file in command line

The tar software, which gave its name to the .tar format, is accessible through command-line.

Once located in the folder containing your data (using cd and ls to navigate):

```
tar –czv –f YourArchiveName.tar.gz *.fastq
tar –czv –f YourArchiveName.tar.gz
tar –czv –f YourArchiveName.tar.gz *.fq
tar –czv –f YourArchiveName.tar.gz *.fq.gz
```

run tar

use options:
-c: create archive
-z: use gzip compression
-v: list processed files

name of the output file

According to your file names, choose the files to archive:
*.fq.gz means “any chain of characters followed by .fq.gz”

macOS native tar installation can generated archives corrupted for other OS

Create a tar file with macOS

The tar software is natively available in the macOS Terminal, and thus there is no “point-and-click”, open-source and free solution to create an archive.

⚠ Warning: macOS native tar implementation can introduce UTF-8 encoding problems when archives created on macOS are read on other operating systems.

You need to use install and use GNU-tar:

```
> brew install gnu-tar
```

If brew (an installation manager) is not yet installed first install it using:

```
/bin/bash -c "$(curl -fsSL https://raw.githubusercontent.com/Homebrew/install/HEAD/install.sh)"
```

Once installed, use gtar the same way than tar on the previous slide:

```
> gtar –czv –f YourArchiveName.tar.gz *.fq.gz
```

run GNU-tar

(re)Naming your files

The filenames returned by sequencing platform usually contains a lot of machine-generated elements (sequencer, run or lane IDs), plus the _R1 and _R2 suffixes for the forward or reverse reads.

These names will be directly used by FROGS, thus a good practice is to clean and shorten them as much as possible before the archive creation.

Warning: keep the _R1 and _R2, which are required in downstream analyses.
You should not have any other _R1 or _R2 than machine-suffixes in your naming.

It can be done with command line or with renaming tools such as *ant renamer:*

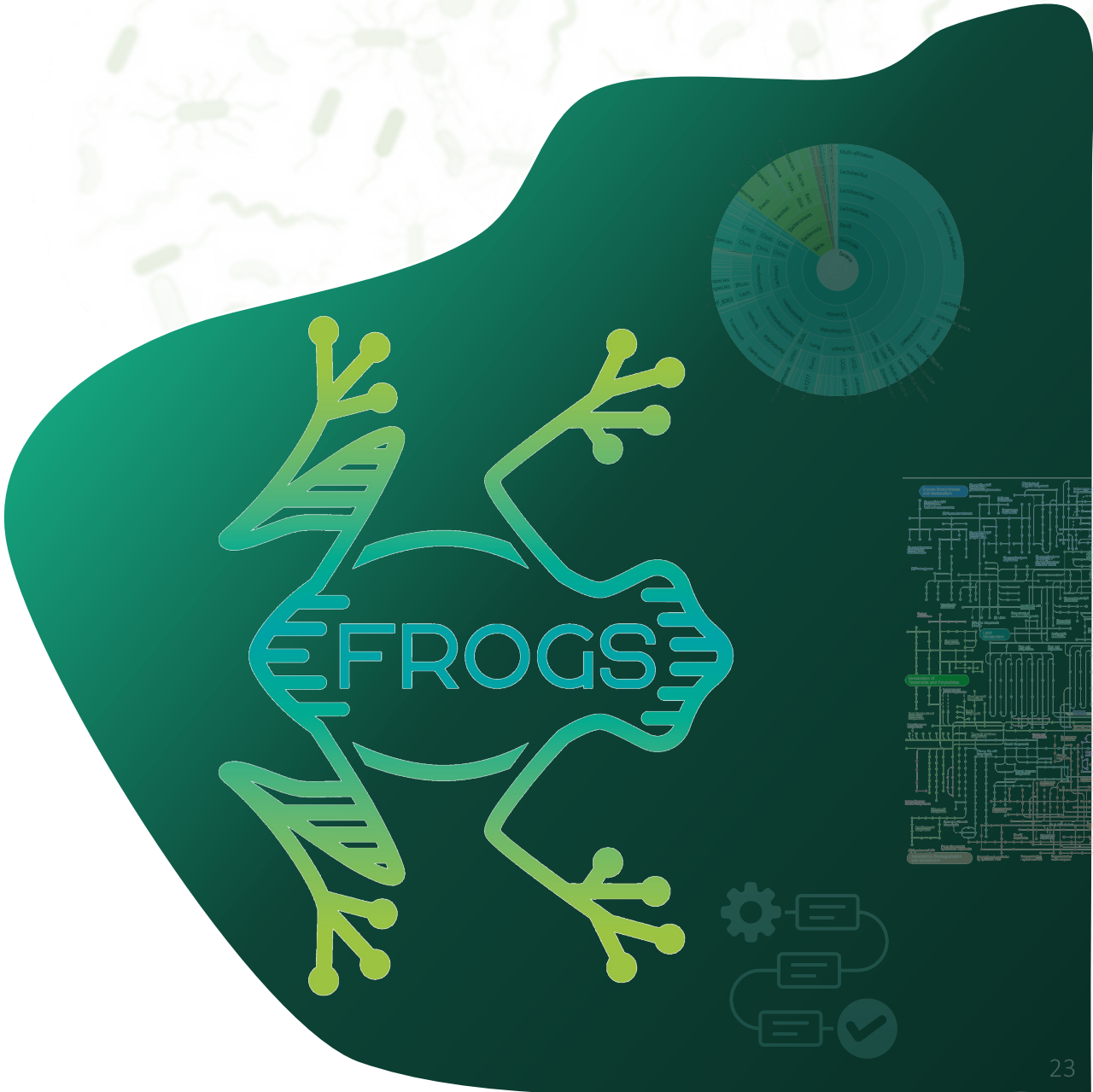
https://antp.be/software/renamer/fr

http://update.antp.be/renamer/antrenamer2_install.exe

Documentation: https://documentation.help/Ant-Renamer-2-fr/actions_fr.html

Get Started with Galaxy

Uploading your data



Uploading your data

Importing your data into Galaxy

Before running FROGS, you need to import your input files into Galaxy.

These are typically your **raw sequencing read**s in FASTQ format (single-end or paired-end).

How to upload your data

In the **Activity bar** (left side), click “Upload”.

Choose **one** of the following options:

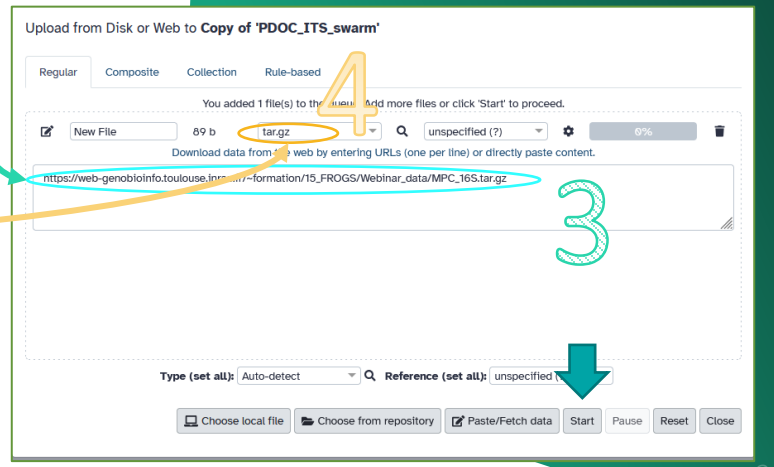
- Choose local file – drag and drop files or select them manually.

or

- Paste/Fetch data – paste a link/url to an online dataset.

Once uploaded, the files will appear in the **History panel** (on the right).

Their status will change from grey → orange → green when the upload is complete.

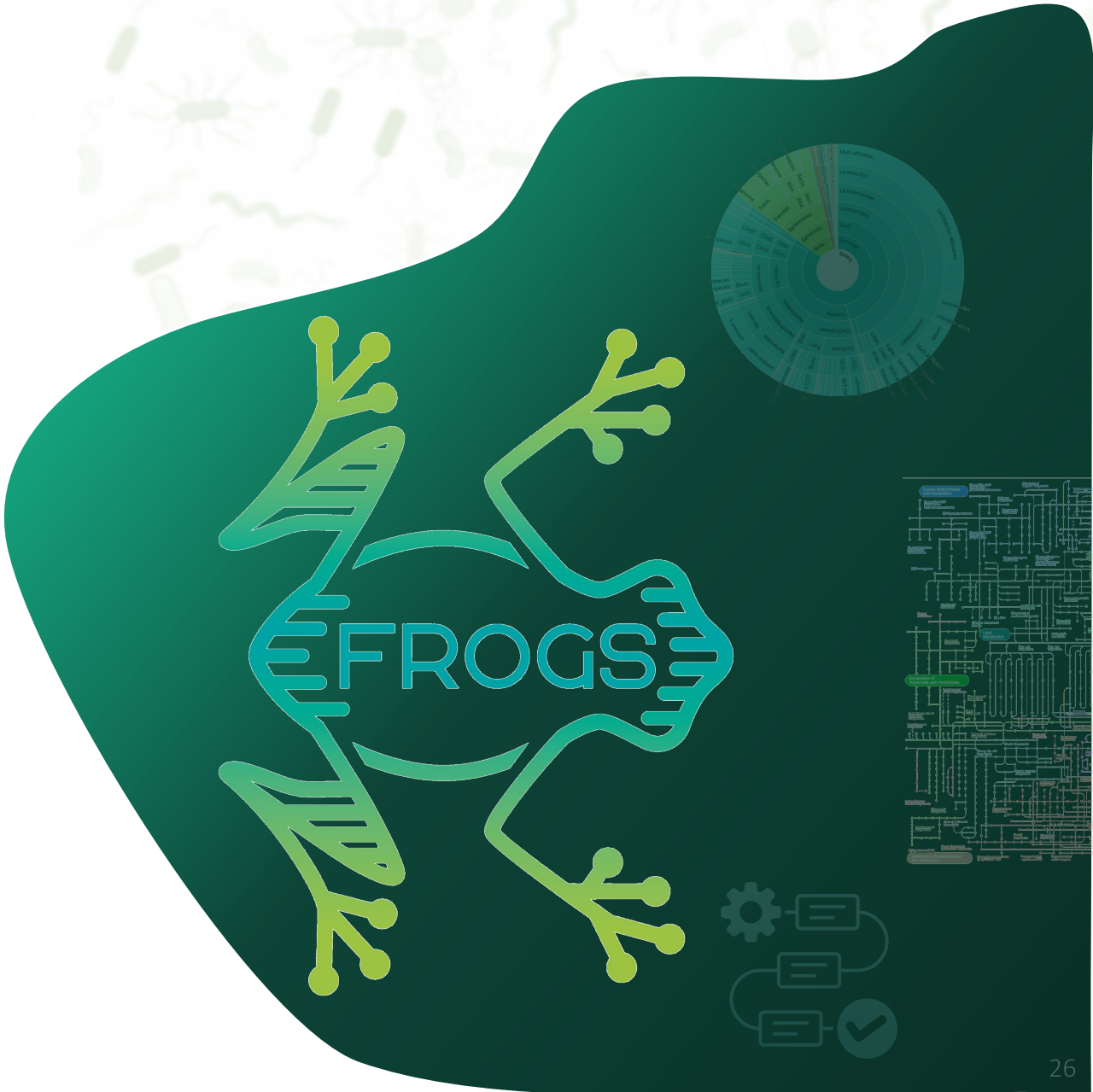


⚠ Important reminders

- Make sure your data correspond to the correct datatype (tar, fastq.gz, etc.).
- Tips: after uploading, if the datatype is incorrect,
  you can edit the metadata by clicking the ✎ (Edit attributes) icon.

Get Started with Galaxy

The tool interface



The tool interface

Every Galaxy tool — including each step of the FROGS workflow — is presented through a **tool form** in the central panel.

This interface allows you to select input data, adjust parameters, and start analyses.

A tool form is divided into several key sections:

Input datasets:
You can select one or more datasets from your history as inputs.
For example, in *FROGS Core Main 1.a*, you choose your raw FASTQ files.

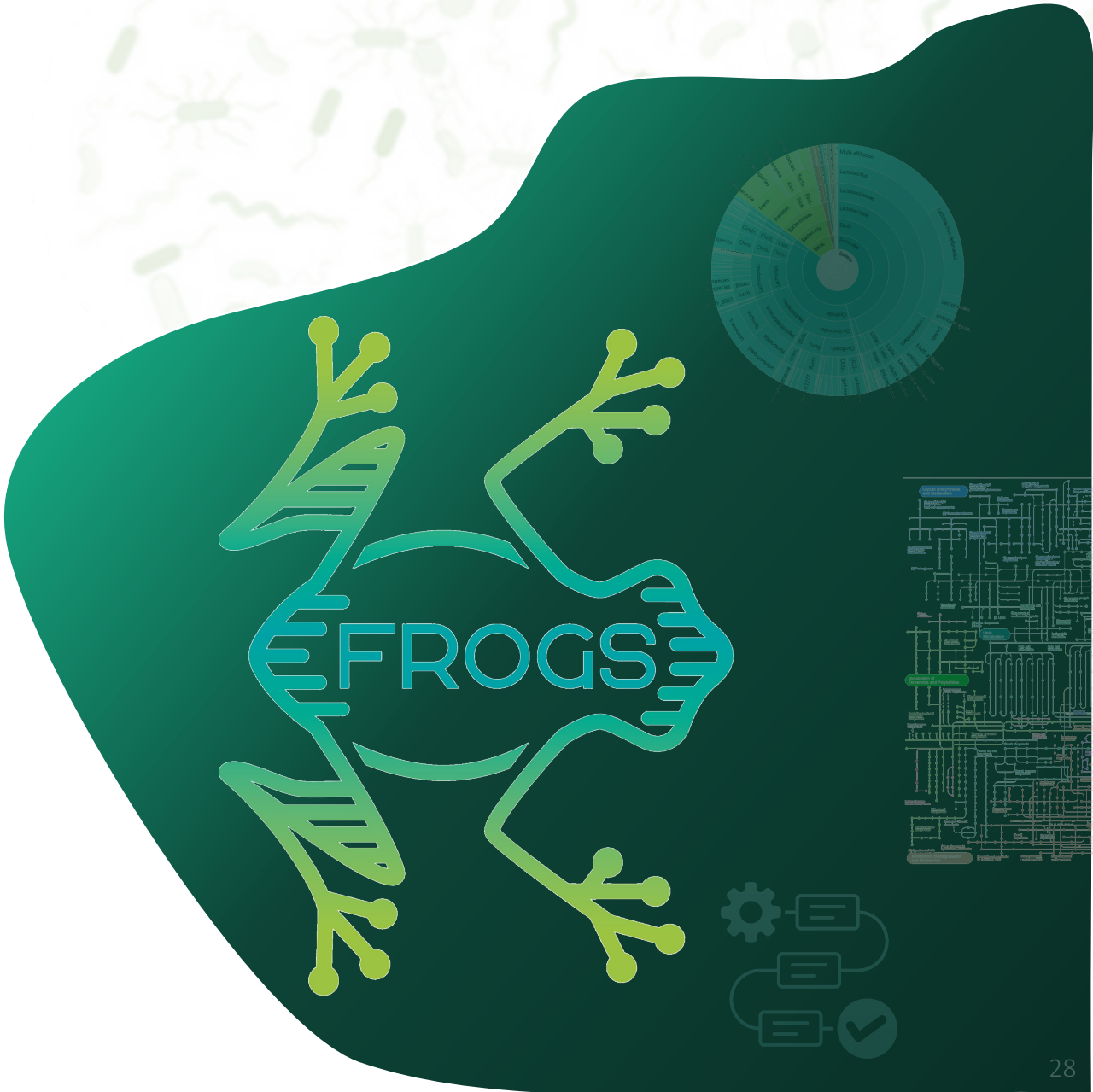
Parameters:
Below the inputs, you can configure tool-specific options such as quality thresholds, clustering methods, or reference databases.
Default parameters are already optimized for most use cases but can be adjusted if needed.

Run Tool button:
Once your inputs and parameters are set, click **“Run Tool”** to start the job.
The analysis will execute on the Galaxy server, and the resulting output datasets will appear in your **History panel**.

N.B.: New tool versions can be installed without removing old ones to ensure reproducibility.

Get Started with Galaxy

Running an analysis



Running an analysis

Once your data are uploaded, you can start a FROGS analysis using the available workflow or individual tools.

To launch FROGS tools:

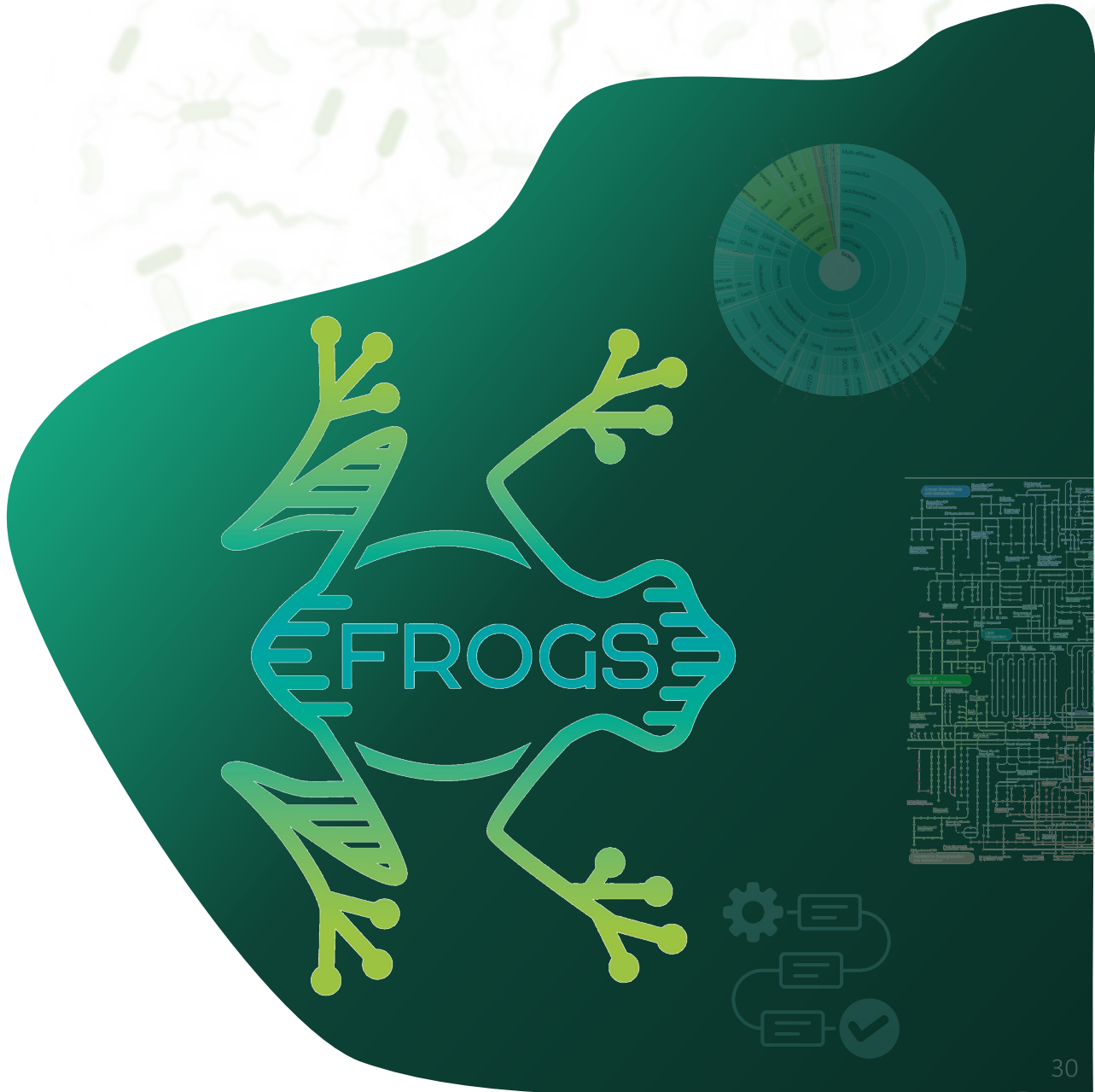
1. In the Tools panel, open the category FROGS (or similar, depending on your Galaxy instance).
2. Click on a tool of FROGS-5 Core, FROGS-5 Stat, FROGS-5 Func, or another tool adapted to your analysis.
3. In the central panel, select your input datasets (e.g. FASTQ files or abundance table).
4. Adjust parameters if needed — default values are usually well suited for standard datasets.
5. Click on the Run tool button.

Galaxy will automatically run each step in sequence. You can follow the progress in the History panel, where outputs will appear one by one. Each tool works independently and generates new datasets that can be used as inputs for the next step.

Tip: If a step fails (red color), click on the dataset to read the error message by clicking on the bug  and check input compatibility or parameter settings. If you do not understand where the error is, contact frogs-support@inrae.fr .

Get Started with Galaxy

Managing your histories



Creating and organizing histories

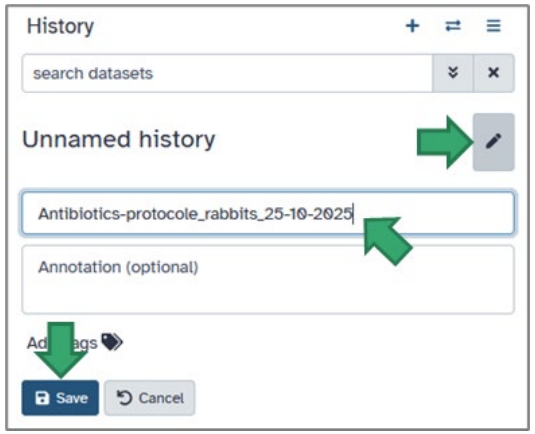
You can rename your current history by clicking on its title or on the pencil (top of the History panel).

To create a new one, click on the cross icon at the top of the right panel.

You can easily switch between histories from:

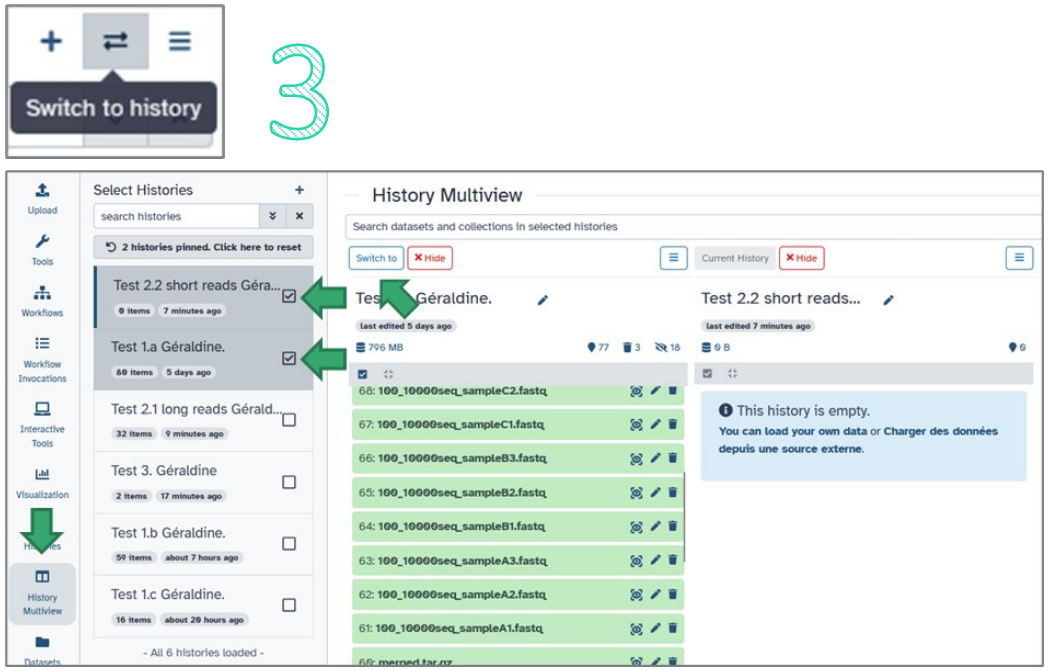
1. the **double arrow** on the top of the History panel
2. the **activity bar**. Simply select the **History Multiview** module.

1



2

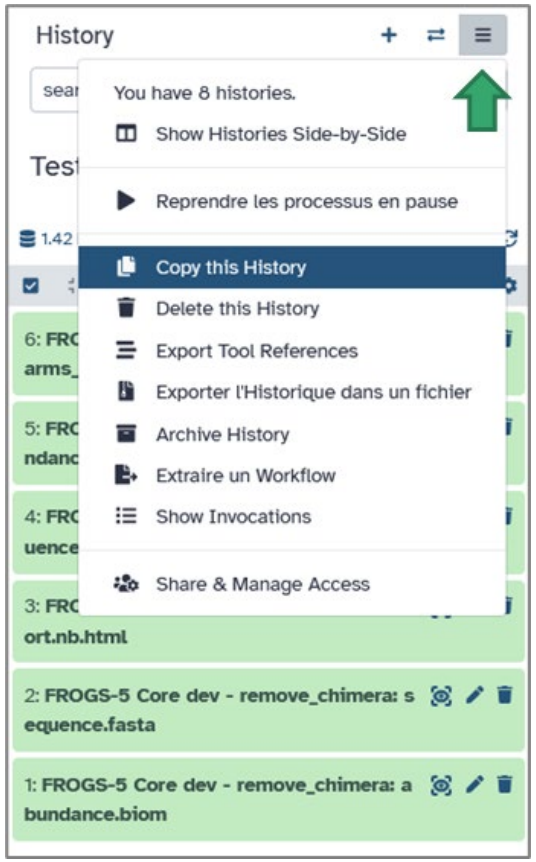
3



Duplicating and cleaning histories

1

Each history can be duplicated to reuse previous analyses.



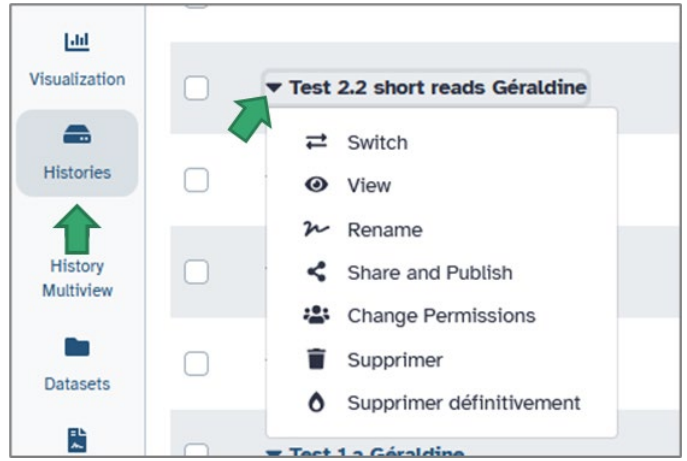
Click the history options icon → **Copy History**. You can then modify parameters or rerun analyses without altering the original version.

2

To save space, you may delete unnecessary datasets (click the trash icon), or obsolete history via history option icon.

3

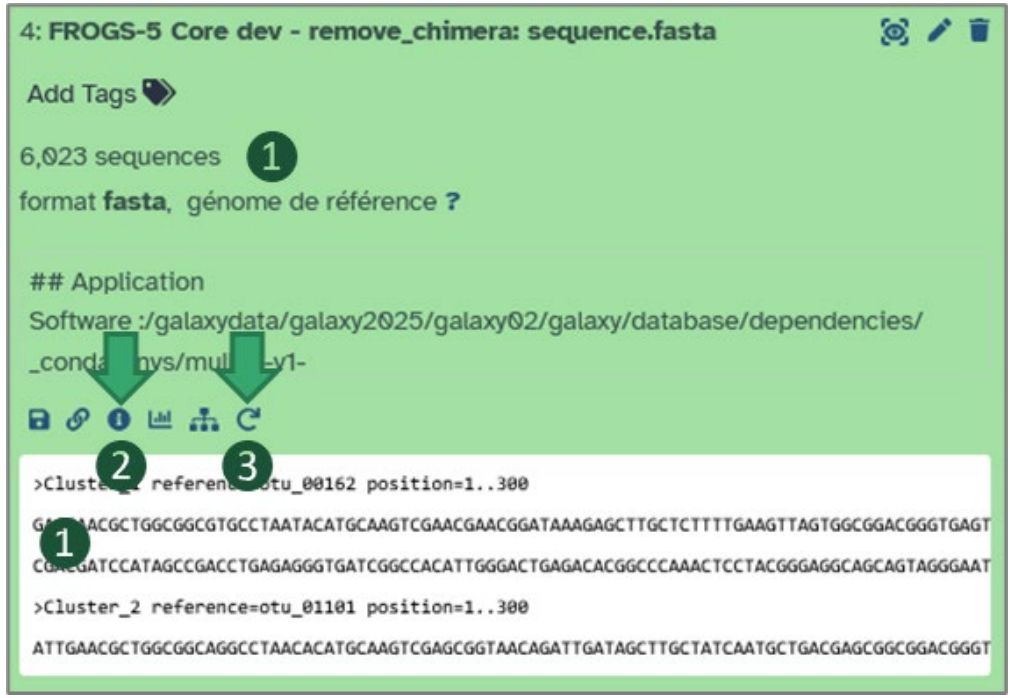
Via the **activity bar**, by selecting **Histories**, you will have access to many useful options.



Viewing and re-running tools

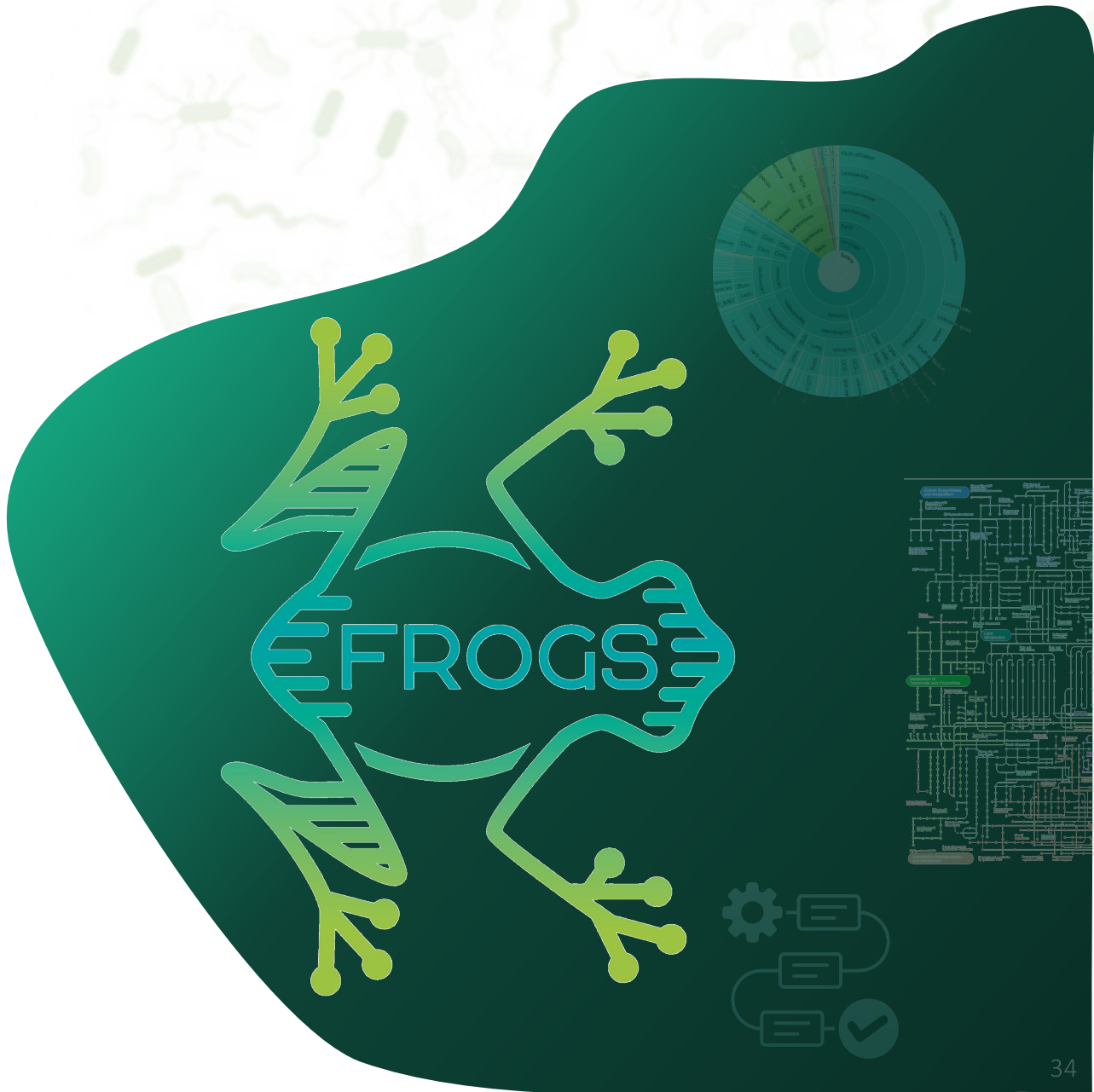
Clicking on a dataset allows you to:

1. Inspect metadata and outputs.
2. View the tool parameters used to generate it, click on “i”.
3. Re-run the same tool with identical or modified parameters using the Re-run button — ensuring complete reproducibility.



Get Started with Galaxy

Saving, sharing, and exporting your work



Saving and exporting

Your histories are automatically saved on the server, but you can also export them locally:

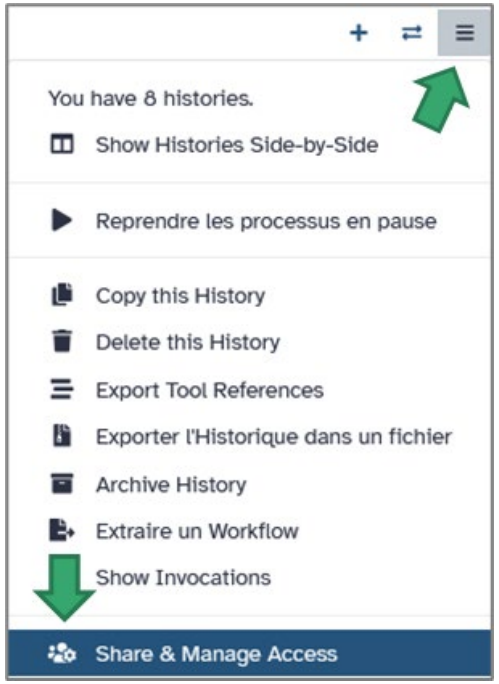
Open the **History options** menu → **Export to file**. This generates an archive (in .tar.gz) containing all datasets and workflow information, which you can download and re-import later or share with collaborators.

You can also export individual datasets via their specific menu → **Download**.

Sharing with collaborators

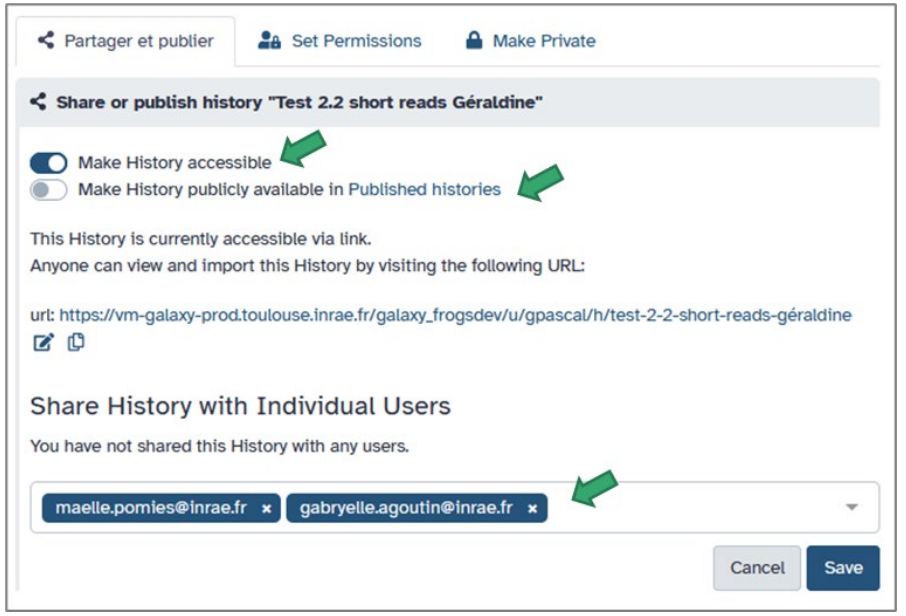
Galaxy makes it easy to share your work.

From the **History options**, select **Share & manage access**.



You can then:

- Generate a **private link** to share with specific collaborators.
- Make the history **public** so it's visible to anyone on the platform.
- Share it directly with another registered Galaxy user (by username or email).



When shared, others can view, copy, or re-run your analyses — ensuring transparency and reproducibility.

Working with shared histories

When another Galaxy user shares a history with you, it appears under your **Histories menu** but is not editable. You can view all datasets, inspect parameters, and even re-run tools, but you cannot directly modify the shared version.

To open a shared history: Go to the **Histories module** on the activity bar. Select **Histories shared with me**. Click on the history name to see it in your central panel.

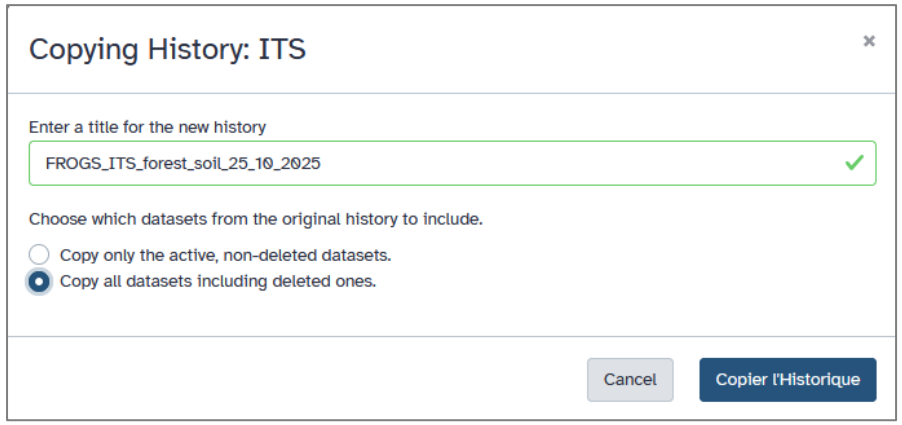
You will see all datasets and workflow steps exactly as they were produced by the owner, including tool versions, parameters, and metadata.

Working with shared histories

If you want to **reuse or modify** the shared analysis: Click on **View** the shared history and **Import this history**.

Galaxy will create a personal copy in your own account. You can then rename it, delete or add datasets, adjust parameters, or re-run the workflow using your own data.

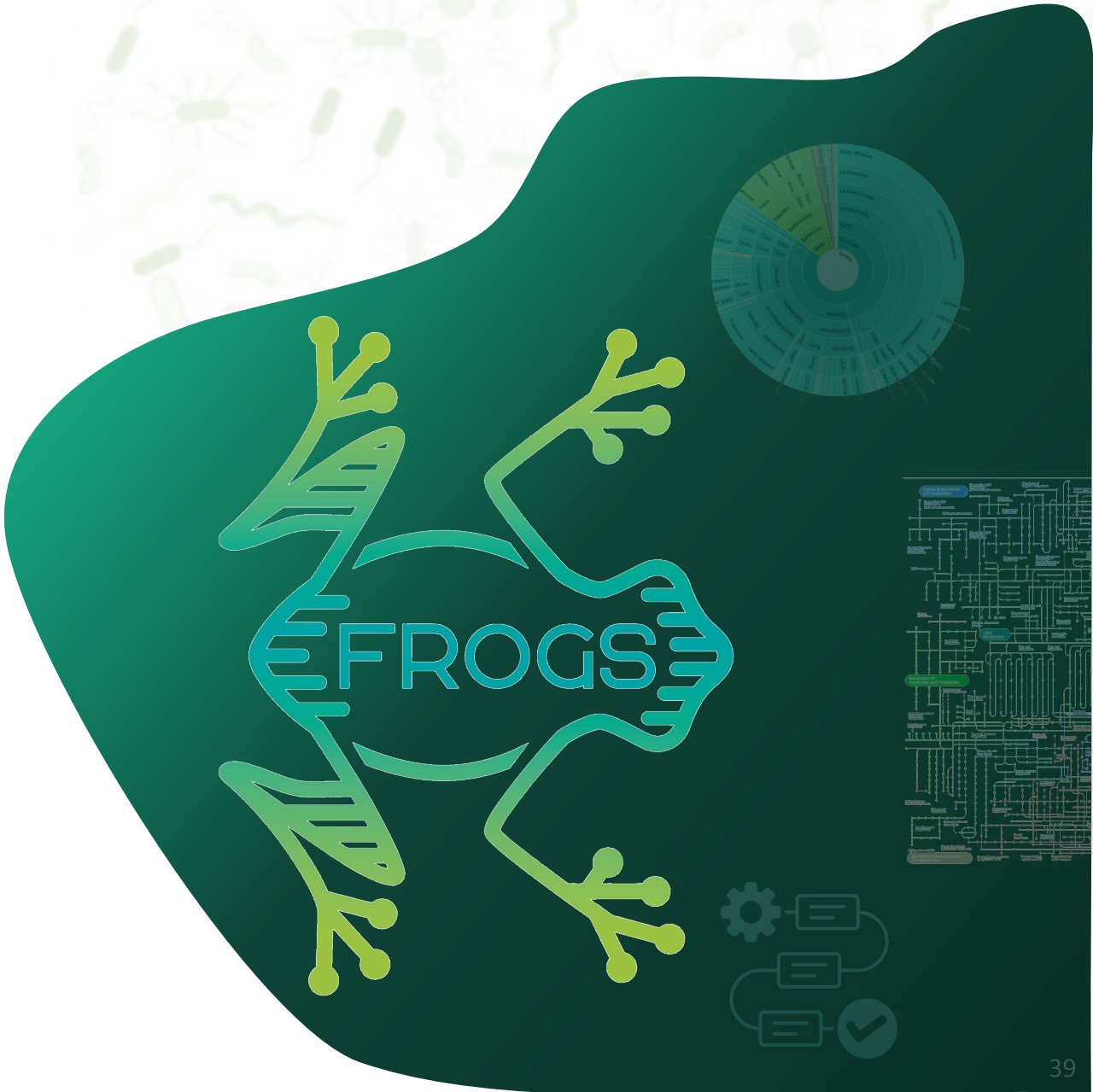
When the owner modifies their history, you will see the history update when it is shared.



Note: Any changes you make to your copy will not affect the owner's history. You can, in turn, share your modified copy with others using the same **Share or Publish** options.

Get Started with Galaxy

Building workflows from your histories



From history to workflow

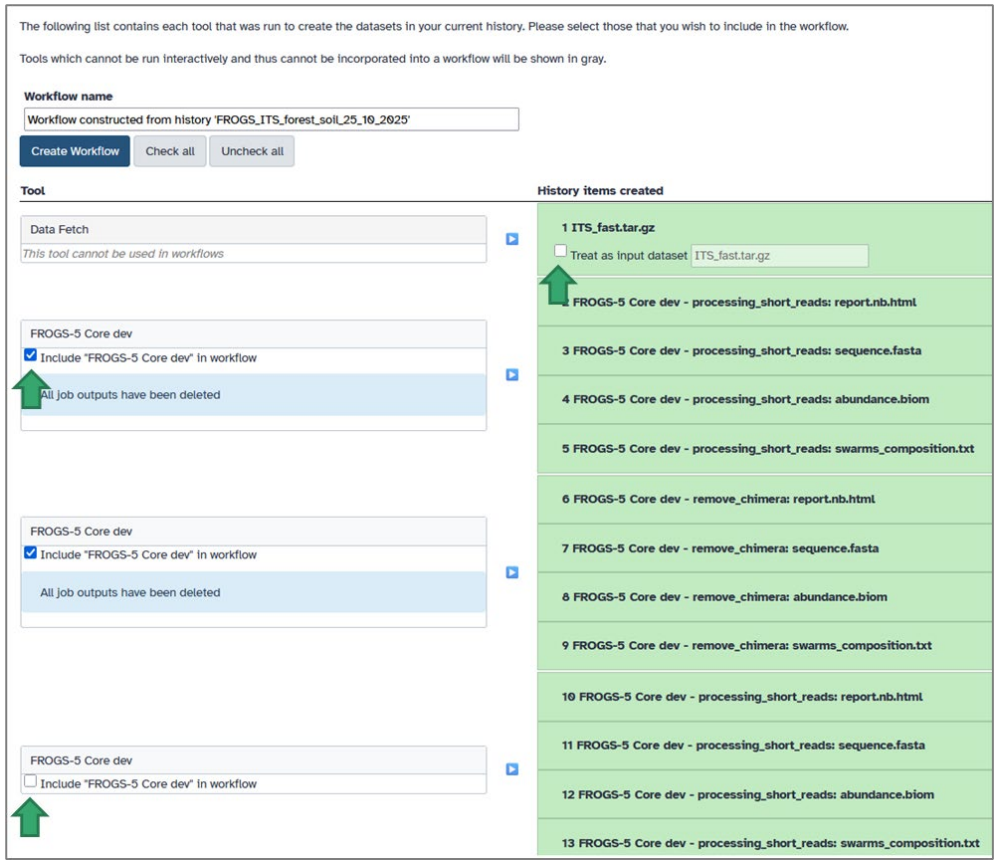
Once you've completed an analysis and are satisfied with the results: Open your **History option** (on the top of the right panel).

Choose **Extract Workflow**. Galaxy will display a new page showing all the tools that were run in this history, in their exact order.

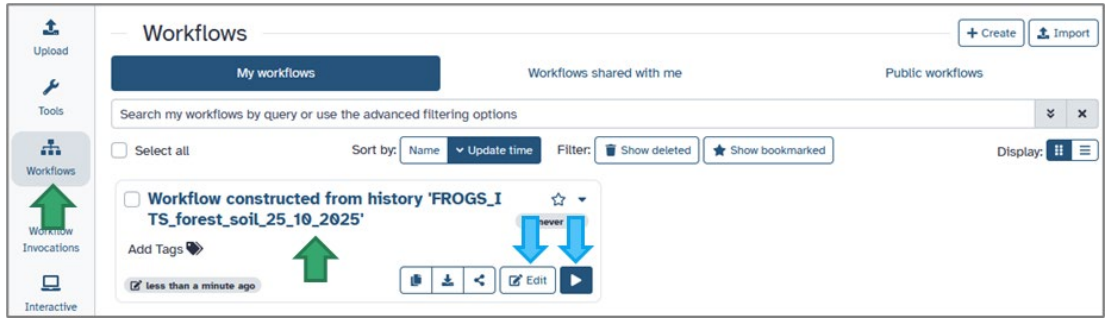
Selecting the relevant steps

In the workflow extraction view: Each line corresponds to a tool execution from your history. You can **select or deselect** individual steps to include only those you need.

For example, you may skip data import or statistical steps if they are not required for automation.



Give your workflow a **clear, descriptive name** (e.g. *Workflow constructed from history 'FROGS_ITS_forest_soil_25_10_2025'*), then click **Create Workflow**.
Your new workflow will appear in the **Workflow menu** (accessible from the activity bar).

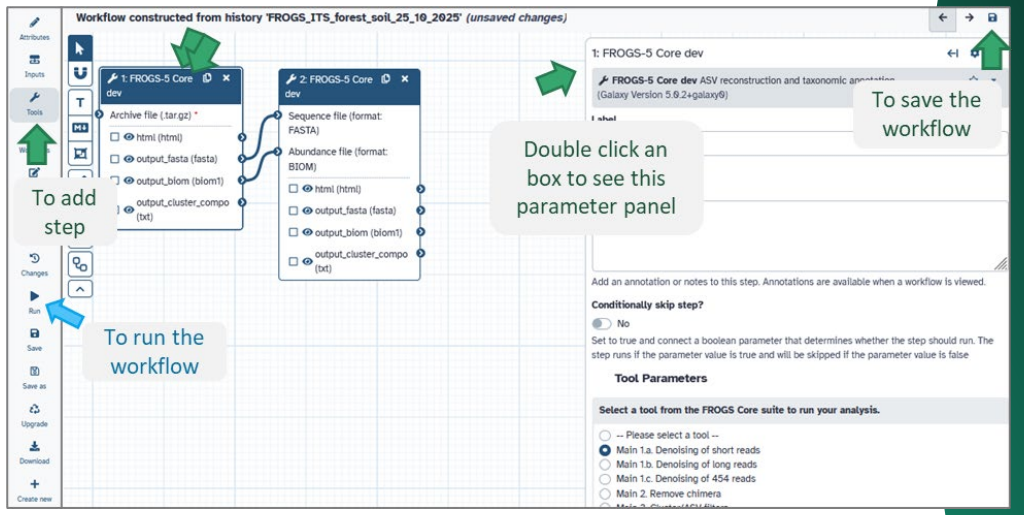


Editing and customizing your workflow

Once your workflow is created, you can modify it using Galaxy's **Workflow Editor**:

Open the **Workflow menu** (activity bar).

Click on your workflow name → **Edit**

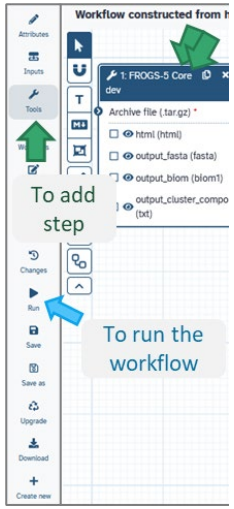


In the editor:

- Tools are shown as boxes connected by arrows (representing data flow).
- You can drag and drop tools, change parameter defaults, or add/remove steps.
- Each connection indicates how the output of one tool becomes the input of another.
- **Don't forget to save !**

Tip: Double-click a tool box to adjust its default parameters — this can save you time when re-running the workflow later.

Running your workflow

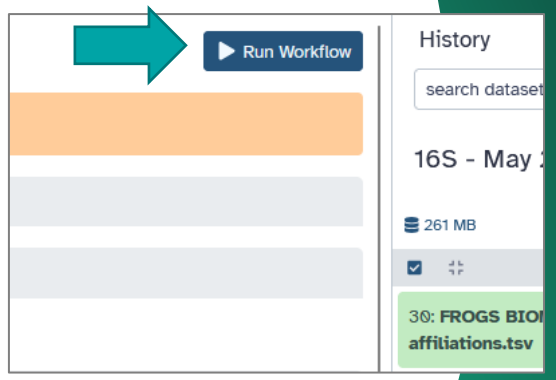


To use your newly created workflow:

Click **Run on the activity bar** next to your workflow panel.

In the History panel (on the right), **select your input datasets** from your current history (or another — click on switch to history).

Adjust parameters (on the central panel) if necessary, then **click Run Workflow** (on the top of the central panel).

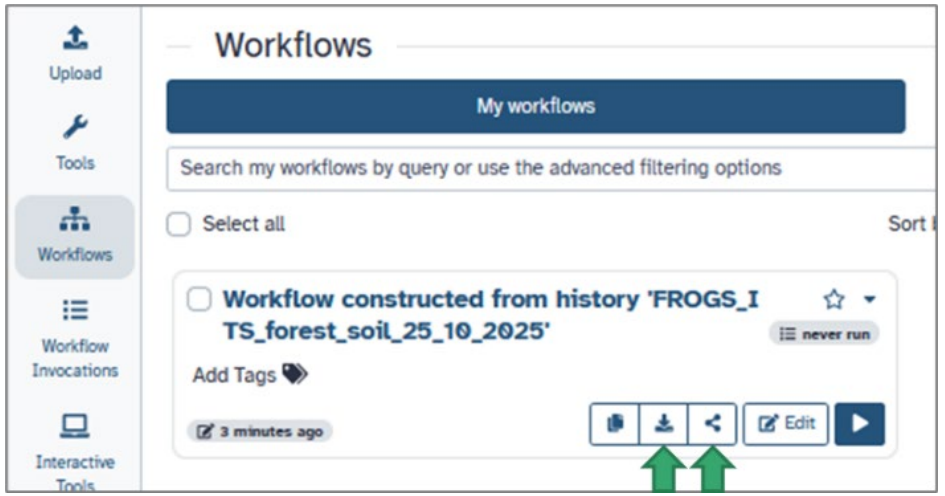


Galaxy will **execute all steps automatically**, following the same sequence as in your original analysis.

Outputs will appear in your current history.

Sharing and reusing workflows

Workflows can be **shared, exported, or published** just like histories: From the **Workflow menu**, click **Share** or **Download**.



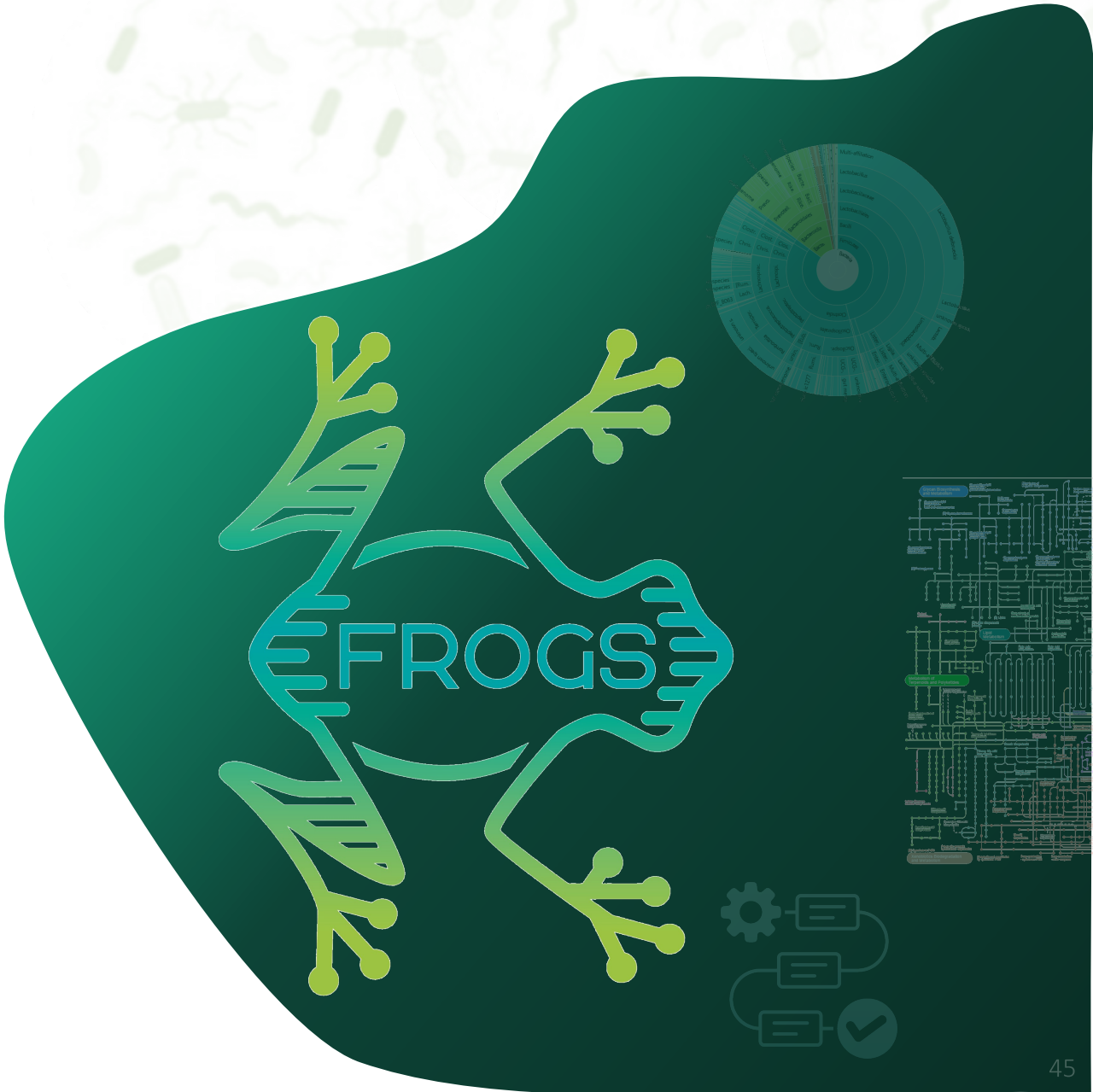
You can:

- Share a private link with collaborators.
- Publish the workflow on your Galaxy instance for the community.
- Download it as a .ga file and import it on another Galaxy server.

This is particularly useful for collaborative projects using FROGS, as all parameters, tool versions, and connections remain identical — ensuring perfect reproducibility.

Get Started with Galaxy

Best practices for stable analyses



Best practices for stable analyses

- **Organize your histories:** Keep one history per project or dataset to avoid confusion.
- **Use clear dataset names:** Short, informative names (e.g. “Soil_16S_PACBIO_summer2025”) help track inputs and results.
- **Avoid special characters or spaces** in dataset names — use underscores _ instead.
- **Regularly clean up old histories** or purge deleted datasets to free storage space.
- **Document your workflows:** Save and name your customized workflows clearly (e.g. “FROGS_pipeline_v5_customparams”).
- **Use the Re-run option** rather than reconfiguring tools from scratch — it guarantees parameter consistency.

Practice session

Now, we will prepare the histories using the data sets that are useful for training purposes.

We need to connect to the Galaxy platform.

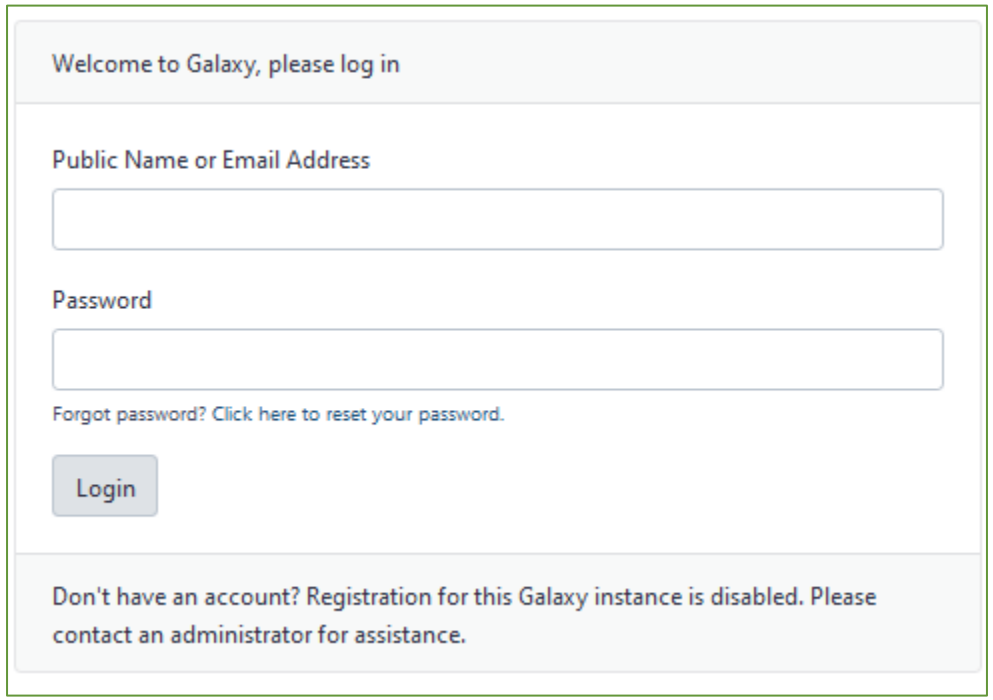
We need to create histories.

We need to download the datasets.

Practice session

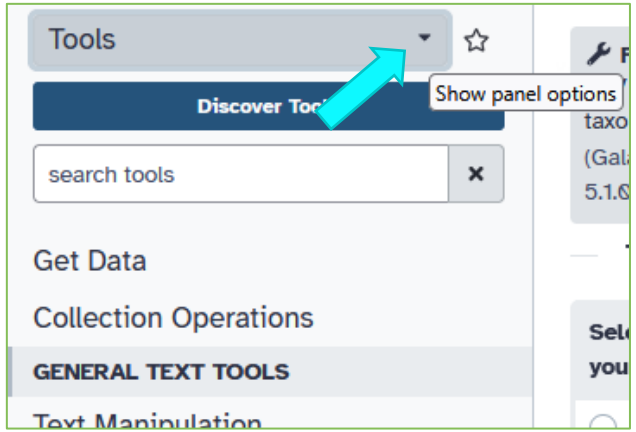
We need to connect to the Galaxy platform.

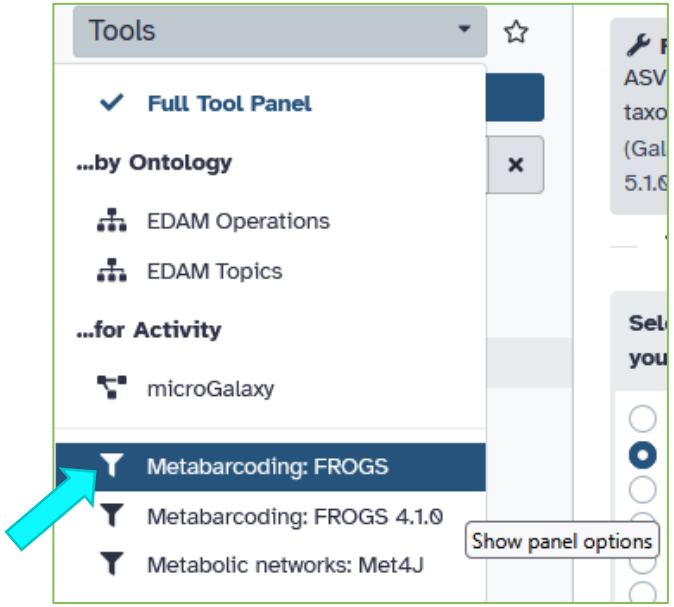
- Go to https://metabarcoding.usegalaxy.fr
- Enter your login and your password



Practice session

- Visit the Galaxy Platform.
- Look at the tool list.
- Display only FROGS tools.
- Display all tools concerning fastq files.
- Display the FROGS panel and choose FROGS



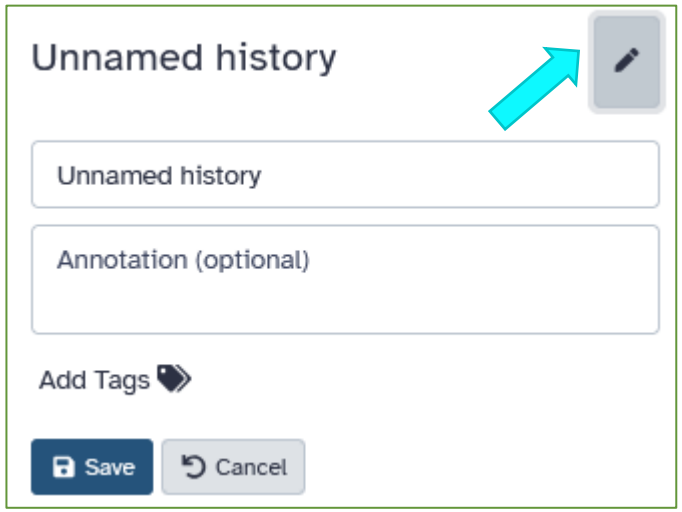


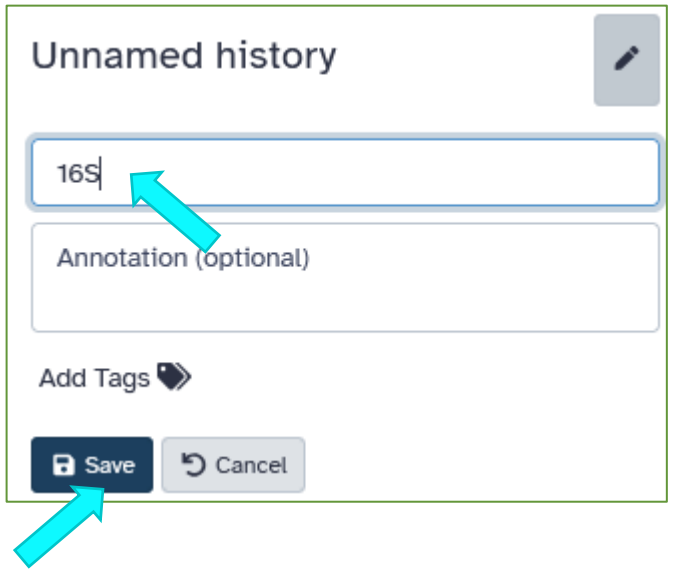
Practice session

We need to create histories.

Create a new history.

Rename this history.

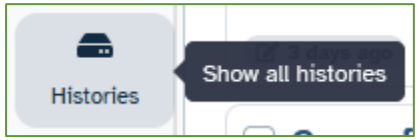




Practice session

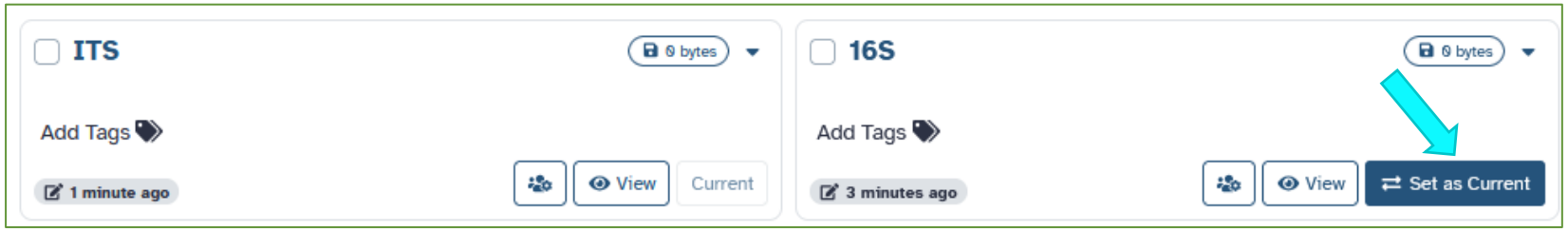
Create a new history named ITS

Look the 2 histories



Look the 2 histories

Return on the 16S history



Practice session

We need to download the datasets.

Together, we will examine what these files represent and the nature of this data.

However, please note that this is an analysis of cheese samples and includes data from both 16S and ITS metabarcoding sequencing of the same samples.

The metadata is therefore identical for both sets of sequence data.

Practice session

We need to download the datasets.

16S data are here :
https://web-genobioinfo.toulouse.inrae.fr/~formation/15_FROGS/30_March2026/

You have to upload
- MPC_16S.tar.gz: the data file
- MPC_complete_metadata.tsv: the metadata file

Do not hesitate to open the TAR archive, look at the sample names and see the sequences…

To upload files, you can follow protocol described in these slides.

Practice session

We need to download the datasets.

Return in the ITS history

ITS data are here :
https://web-genobioinfo.toulouse.inrae.fr/~formation/15_FROGS/30_March2026/

You have to upload

- MPC_ITS.tar.gz: the data file
- MPC_complete_metadata.tsv: the metadata file

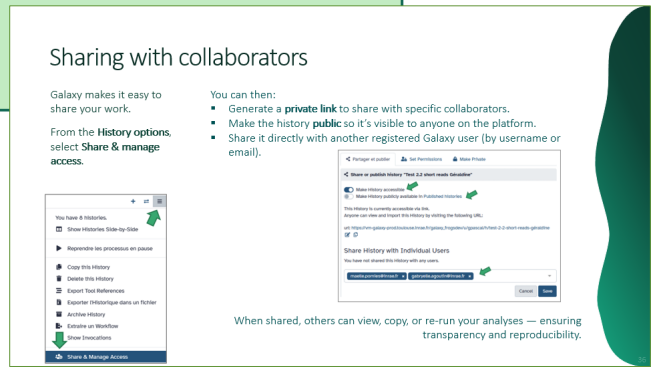
Practice session

Your histories should now be ready.

Please share them with us.

lucas.auer@inrae.fr ✖ geraldine.pascal@inrae.fr ✖

You can refer to this slide.



Please send us an email to let us know when everything is OK.

frogs-training@inrae.fr

Thank you for following this first step.
See you on monday 23th March!

Your trainers:

Géraldine & Lucas