A - Training on Galaxy: Metabarcoding

June 2021 - webinar

GALAXY Practice

LUCAS Auer, MARIA Bernard, LAURENT Cauquil, MAHENDRA Mariadassou, GÉRALDINE Pascal & OLIVIER Rué

.
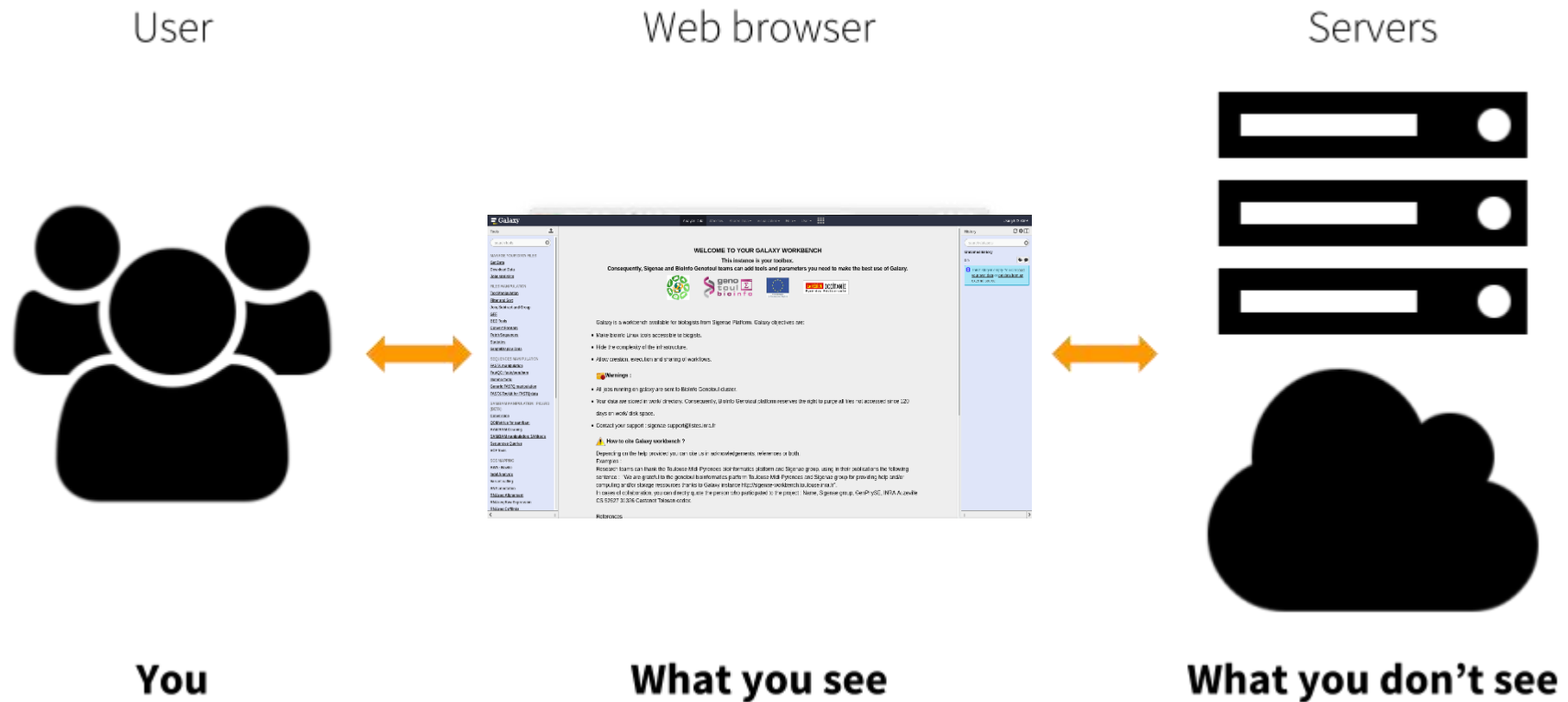
Objectives of the 1st part

- Learn the basics of Galaxy

- Prepare the 2 datasets for FROGS formation

What is Galaxy?

Galaxy software framework

- Galaxy is an overlay that allows researchers without computer expertise to perform computational analyses online, automate them and share the results easily.

- Developed by the Galaxy Team Project.

- Relies on a computational infrastructure (Server).

- Open source

Galaxy software framework

User

Web browser

Servers

You

What you see

What you don't see

https://galaxyproject.org/tutorials/g101/

Where to use Galaxy?

- Galaxy software can be installed and run on powerful server farms (Cluster) or on your individual PC.

- Genotoul Bioinfo

  - Cluster: Many computers tightly connected that work together

  - High performance computer:

    - More than 5000 cores

    - 34 TB of RAM

    - More than 1 Peta Byte (1024 TB) of hard drive

Where to use Galaxy?

- The Toulouse Galaxy platform is https://vm-galaxy-prod.toulouse.inra.fr/galaxy/

⚠ The different platforms of Galaxy are not connected together

- Galaxy is installed on many clusters across the world.

- Some tools are in <u>our</u> Galaxy platform but not in <u>other</u> platforms.

- Your data is not shared with other Galaxy platforms than ours.

Exemple of 2 INRAE Galaxy platforms

Exemple of 2 INRAE Galaxy platforms

Practice:

CONNECT TO OUR GALAXY WORKBENCH

Exercise

Ask trainers for the address and password.

All your data will be erased at the end of the week of the remote computer

Exercise

During this training, even you have a personal account, don't use it !
We use training accounts for more facilities

If you don't have a personal account, you can request a personal account later on :
http://bioinfo.genotoul.fr/index.php/ask-for/create-an-account/

All your data will be erased at the end of the week on the training account.

Exercise

Our Galaxy platform is: https://vm-galaxy-prod.toulouse.inra.fr/galaxy

Before starting, check if your browser is not already connected to a Galaxy account. If this is the case, log out or use a private browser window!

Risk of data duplication

Exercise

Our Galaxy platform is: https://vm-galaxy-prod.toulouse.inra.fr/galaxy/

Be careful, to fully login you must enter your credentials twice:

- The first time in this pop-up window:
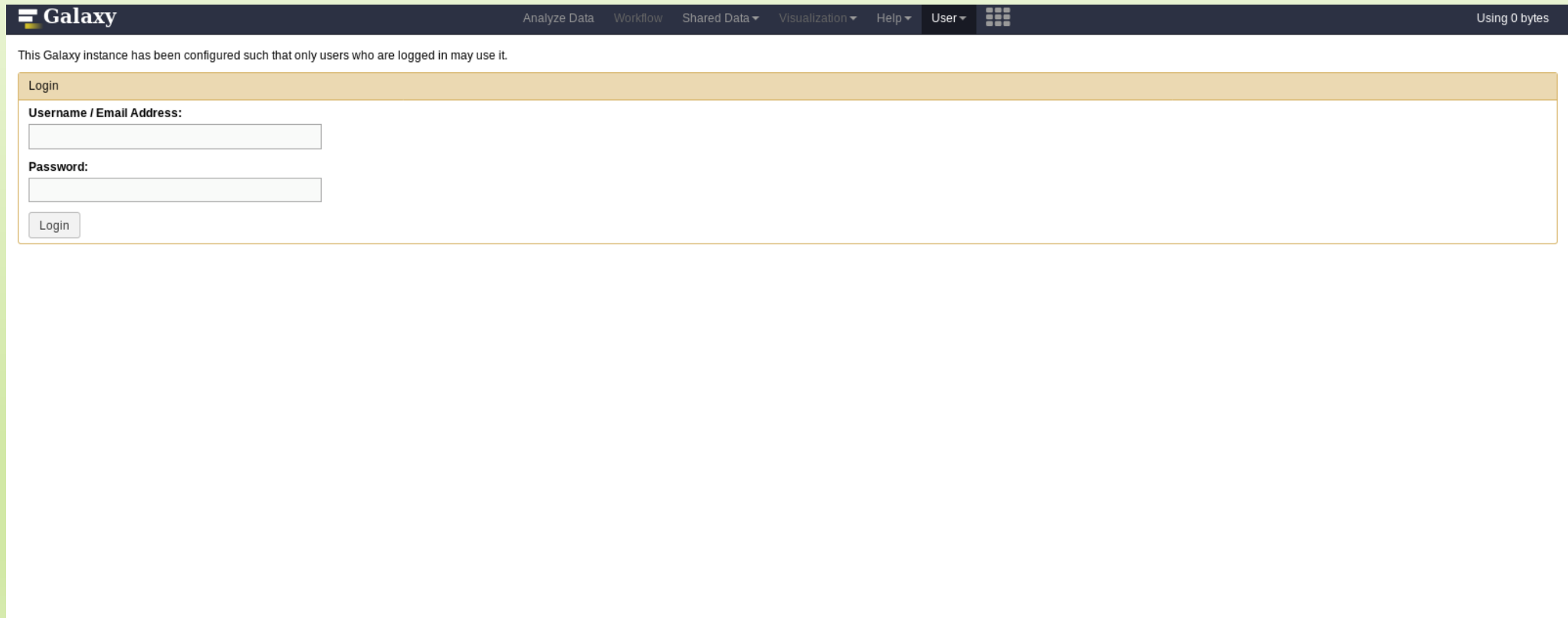
Exercise

- And a second time, in the browser:

Galaxy

Analyze Data | Workflow | Shared Data ▾ | Visualization ▾ | Help ▾ | User ▾

Using 0%

Tools

search tools

MANAGE YOUR DATA FILES
Get Data
Download Data
Jobs statistics

FILES MANIPULATION
Text Manipulation
Filter and Sort
Join, Subtract and Group
GFF
BED Tools
Convert Formats
Fetch Sequences
Statistics
Graph

AVAILABLE TOOLS

SEQUENCES MANIPULATION
FASTA manipulation
FastQC: fastq/sam/bam
Illumina fastq
Generic FASTQ manipulation
FASTX-Toolkit for FASTQ data

SAM/BAM MANIPULATION : PICARD (BETA)
Conversion
QC/Metrics for sam/bam
BAM/SAM Cleaning
SAM/BAM manipulation: SAMtools
Sequences Queries
VCF Tools

SGS MAPPING
BWA - Bowtie
Indel Analysis
Variant calling
SNP annotation
RNAseq Alignement
RNAseq Raw Expression
RNAseq Cufflinks

History

search datasets

Data Collection
3 shown

33 b

3: seq1.fasta

2: seq2.fasta

1: seq3.fasta

DATASETS HISTORY

How to cite Galaxy workbench ?

Depending on the help provided you can cite us in acknowledgements, references or both.

Examples :

Research teams can thank the Toulouse Midi-Pyrenees bioinformatics platform and Sigenae group, using in their publications the following sentence : "We are grateful to the genotoul bioinformatics platform Toulouse Midi-Pyrenees and Sigenae group for providing help and/or computing and/or storage ressources thanks to Galaxy instance http://sigenae-workbench.toulouse.inra.fr".

In cases of collaboration, you can directly quote the person who participated to the project : Name, Sigenae group, GenPhySE, INRA Auzeville CS 52627 31326 Castanet Tolosan cedex.

References
X. SIGENAE [http://www.sigenae.org/]

**RESULTS VISUALISATION
AND
TOOL PARAMETER WINDOW**

Ready to use Workflows

- 1/ FROGS - Find, Rapidly, Otus with Galaxy Solution
- 2/ Tax4Fun workflow
- 3/ Function Table for Tax4Fun matrix workflow

Galaxy News

- 1/ How to use datasets collections ?
- 2/ How to upload several files in Galaxy ?
- 3/ New tools : Tax4Fun (manual), VIP (manual), Salmon, multiQC.
- 4/ Tools availables for virologie analyses
- 5/ Virology tools
- 6/ How to use Function Table tool ?

E-learning

https://inra.classilio.com/Login
2 e-learning availables : Galaxy initiation and Quality sequences analysis

Vocabulary of Galaxy

- **Tools:**
  - A tool has a function which is explained when you click on it.
  - Each Galaxy platform has its own tools.

- **Datasets:**
  - A dataset is a file, uploaded to Galaxy by you or produced by a tool.
  - Be careful: a dataset has a datatype.

- **Histories:**
  - A tool generates datasets and these datasets are stored in the current history.
  - Everything is permanently saved.
  - If you log off your computer or browser, it's ok, everything will keep running and be saved!

Pratice:

DISCOVER GALAXY

Exercise

1. Visit the Galaxy Platform.

2. Look at the tool list.

3. Display only FROGS tools.

4. Display all tools concerning fastq files.

Exercise

Search a tool by name.
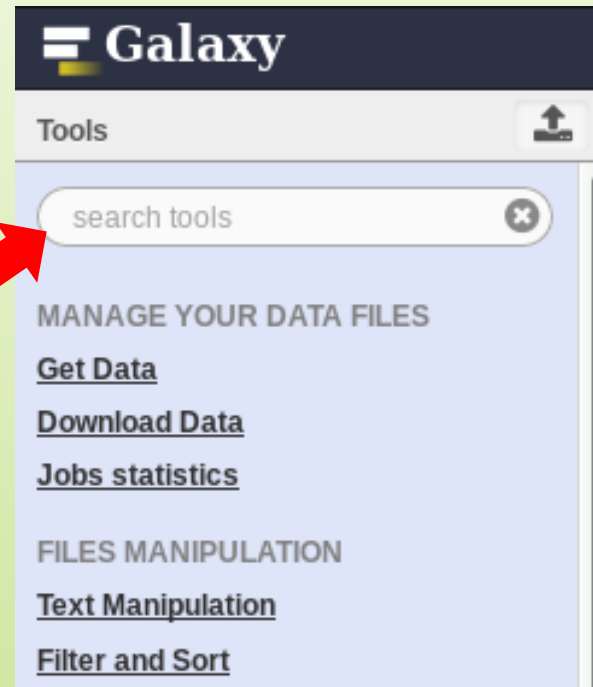
Manipulate Histories

Practice:

CREATE THE 2 HISTORIES NEEDED FOR THE FROGS FORMATION

To create a new history:

- Click on the wheel.

- Click on « create new ».
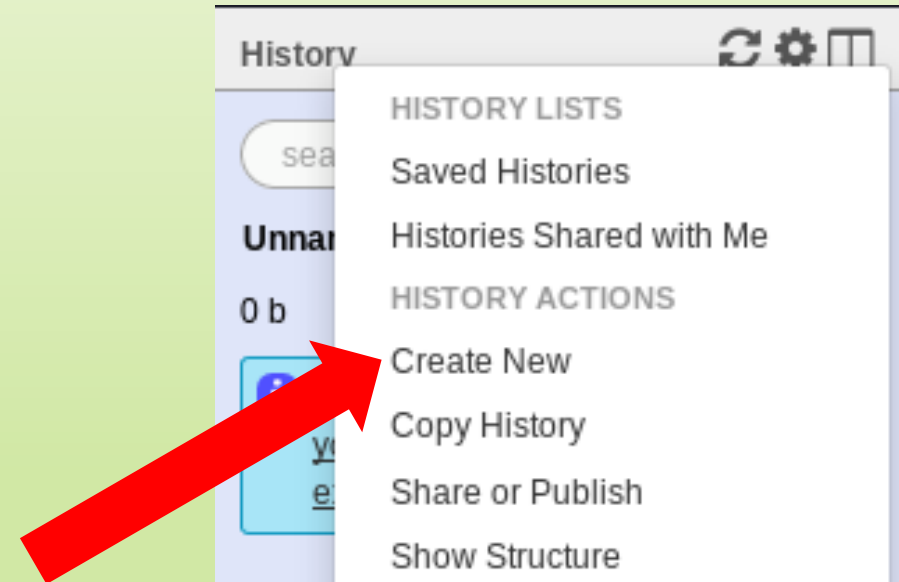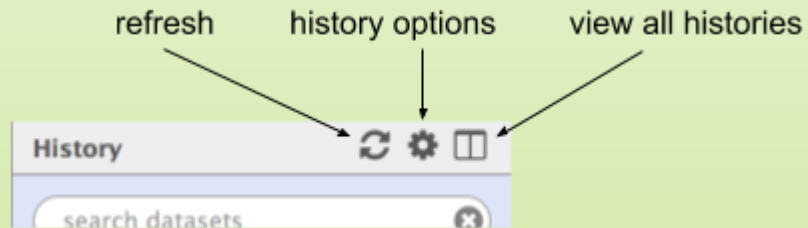
To rename a history:
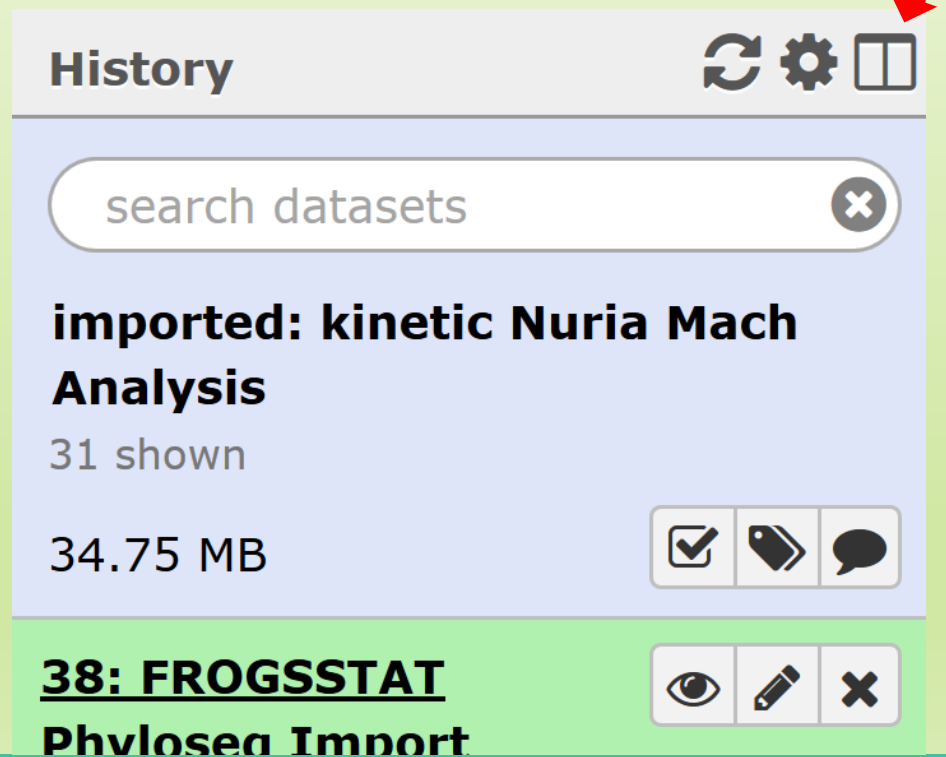
- Click on the history name (at the top).

- Enter « 16S ».

- Hit « Enter » to validate.
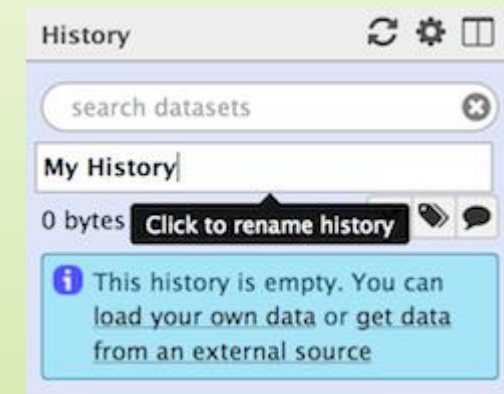
Don't use special characters or accents!

Exercise

- Create another history named ITS

- Switch to the history named « 16S » as current history.

- Go back to the main interface.

How to list all histories?

- To view all histories, click on this icon.

Explore the « View all histories » section

Switch current history

- Switch to the history named « ITS » as current history.

- Click on "Done" to go back to the main interface.

Data import

How to import your data to Galaxy ?

- Several ways to upload your data to Galaxy:

    - From your computer

    - By URL

    - From Genotoul Bioinfo clusters

    - Shared by other users of Galaxy

Formation data

- All the training data are online on:
  http://genoweb.toulouse.inra.fr/~formation/15_FROGS/Webinar_data/

Index of /~formation/15_FROGS/Webinar_data

Name	Last modified	Size	Description
Parent Directory		-	
ITS1.tar.gz	04-Mar-2021 15:40	157M	
chaillou_withprimers_64renamedsamples_V1V3_10000seq_R1R2.tar.gz	04-Mar-2021 15:34	35M	
metadata_ITS.tsv	04-Mar-2021 15:37	394	
metadata_chaillou.tsv	04-Mar-2021 15:29	2.2K	

All samples are in archive .tar.gz

Above all: how create an archive ?

To learn how create an archive
http://genoweb.toulouse.inra.fr/~formation/15_FROGS/Webinar_data/How_create_an_archive.pdf

Create an archive is
always possible !
It is our recommendation.

How to import your data to Galaxy ?

- Several ways to upload your data to Galaxy:

  - From your computer

  - By URL

  - From Genotoul Bioinfo clusters
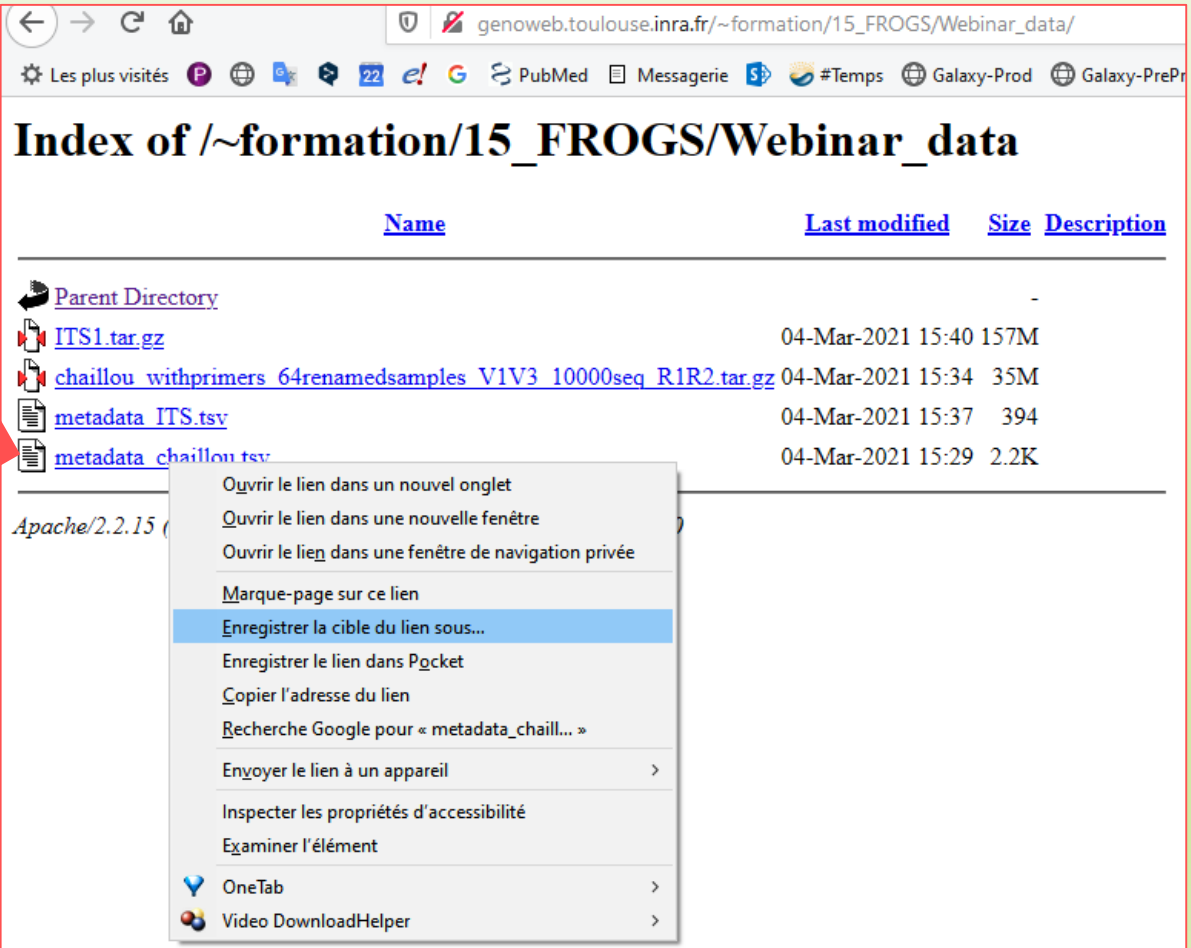
  - Shared by other users of Galaxy

We will learn how to import data with these 2 modalities.

Practice

UPLOAD FILE FROM AN URL

How to import your data to Galaxy ?

- Several ways to upload your data to Galaxy:

    - From your computer

    - By URL

    - From Genotoul Bioinfo clusters

    - Shared by other users of Galaxy

Upload file from URL

1. In your current history "16S"

2. Go to **Get Data** > **Upload File from your computer**

3. Click on **Paste/Fetch Data**

4. Copy the address of the file: http://genoweb.toulouse.inra.fr/~formation/15_FROGS/Webinar_data/chaillou_withprimers_64renamedsamples_V1V3_10000seq_R1R2.tar.gz

5. <u>Change the type</u>!

6. Click on Start

7. You can put one address per line for multiple uploads

Upload file from URL

Pratice:

UPLOAD FILE FROM YOUR COMPUTER

How to import your data to Galaxy ?

- Several ways to upload your data to Galaxy:

  - From your computer

  - By URL

  - From Genotoul Bioinfo clusters

  - Shared by other users of Galaxy

Upload a local file

- To simulate that you have already the file on your own computer, download the file "metadata_Chaillou.tsv" as:

Upload a local file

- In Galaxy, your current history must be 16S

- Click on the « Get Data » tool.

- Upload file from your computer.

⚠ For files smaller than 2 GB for old generations of galaxy platform
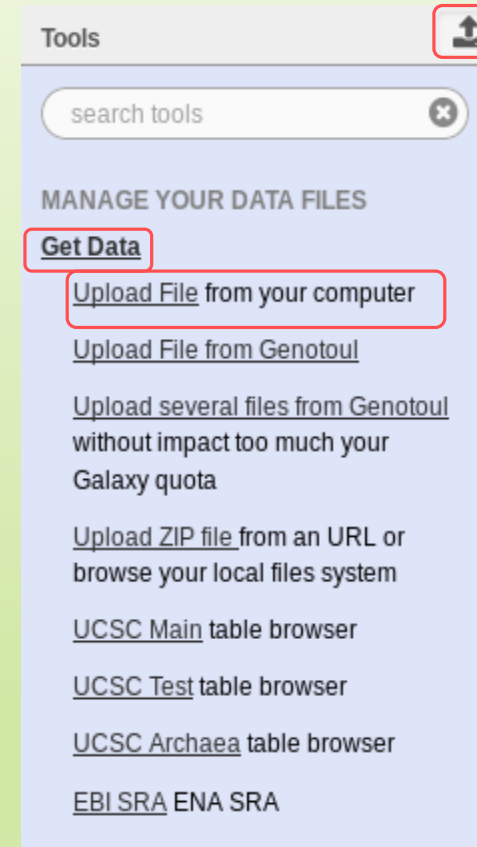
Upload a local file

Download from web or upload from disk

Regular   Composite

⊡ **Drop files here**

Click here to choose a file on your hard drive.

Type (set all):   Auto-detect 🔍   **Genome (set all):**   unspecified (?)

⬛ Choose local file   ✎ Paste/Fetch data   Pause   Reset   Start   Close

Upload a local file

The datatype of a file indicates the content of it.

.tsv = tabular

.fastq = fastq

.tar or .tar.gz = tar

Select the type corresponding to your file.

Do no not trust blindly the auto-detect!

Download from web or upload from disk

Regular   Composite

Name	Size	Type	Genome	Settings	Status
🖥 metadata_chaillou.tsv	2.2 KB	tabular ▾ 🔍	unspecified (?) ▾	⚙	100% ✔

1. Select the type of file (Do not leave on Auto-Detect!).

2. Select your other files the same way.

3. Begin upload.

Type (set all):   Auto-detect ▾ 🔍   Genome (set all):   unspecified (?) ▾

🖥 Choose local file   ✎ Paste/Fetch data   Pause   Reset   Start   Close

Upload a local file

Create and fill **yourself** the ITS history

Get data form here:

http://genoweb.toulouse.inra.fr/~formation/15_FROGS/Webinar_data/
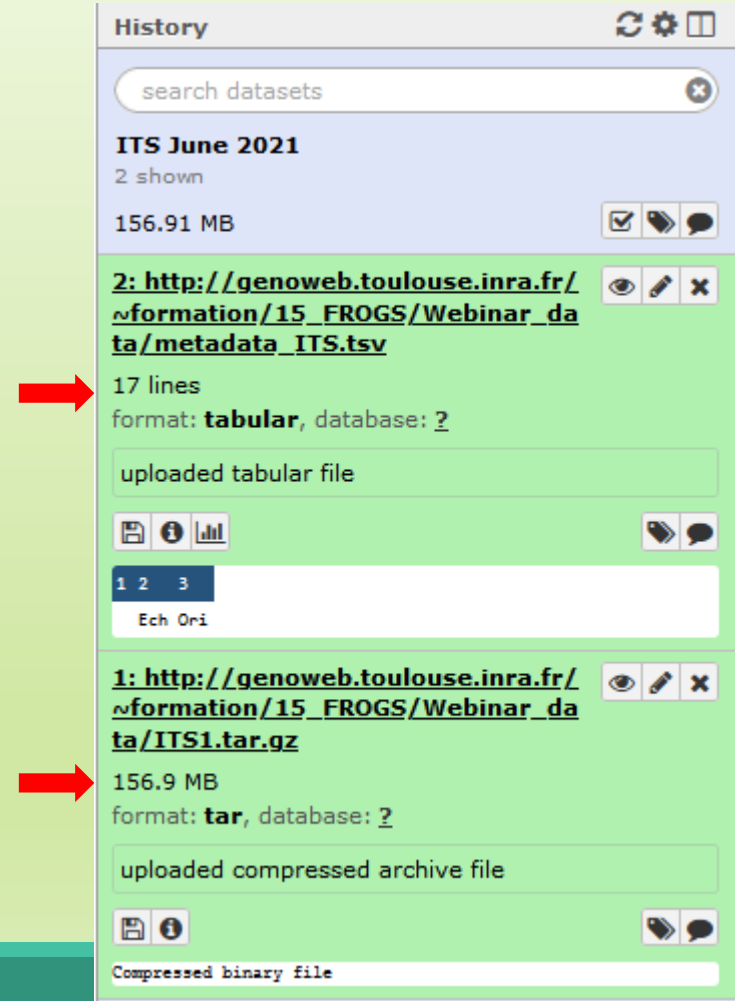
Create and fill **yourself** the ITS history

You have to obtain your second history:

How to import your data to Galaxy ?

- Several ways to upload your data to Galaxy:

    - From your computer

    - By URL

    - From Genotoul Bioinfo clusters

    - Shared by other users of Galaxy

Upload your data from Genotoul

To download data from the server « Genotoul of Bioinformatics of Toulouse », you must already have your data on this server.

To learn how put your data on genotoul
http://genoweb.toulouse.inra.fr/~formation/15_FROGS/Webinar_data/How_to_put_your_data_on_genotoul_server.pdf

Advantage to put your data on Genotoul:
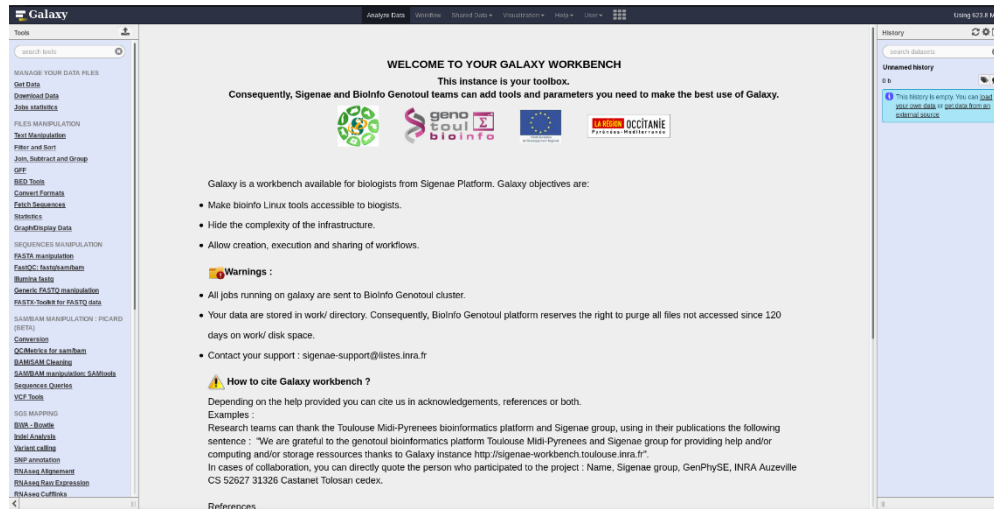
1TB of space

/work: 1TB (1024GB) of space
/save: 250GB of space

- This method allows you to have more space disk and to upload bigger files.

- Files not used in the last 180 days can be purged at any time.

Switch to a new history.

Next go to Get Data > Upload File from Genotoul.

After executing, the file is imported into Galaxy.

Share a History

Share a history

- You can share a history with another Galaxy user:

    - For working with your colleague.

    - For support, so we can help you better and faster.

- You can import shared history to your account too.

Share an history

Click on the wheel and click here.

History

HISTORY LISTS
Saved Histories
Histories Shared with Me
HISTORY ACTIONS
Create New
Copy History
Share or Publish
Show Structure
Extract Workflow
Delete
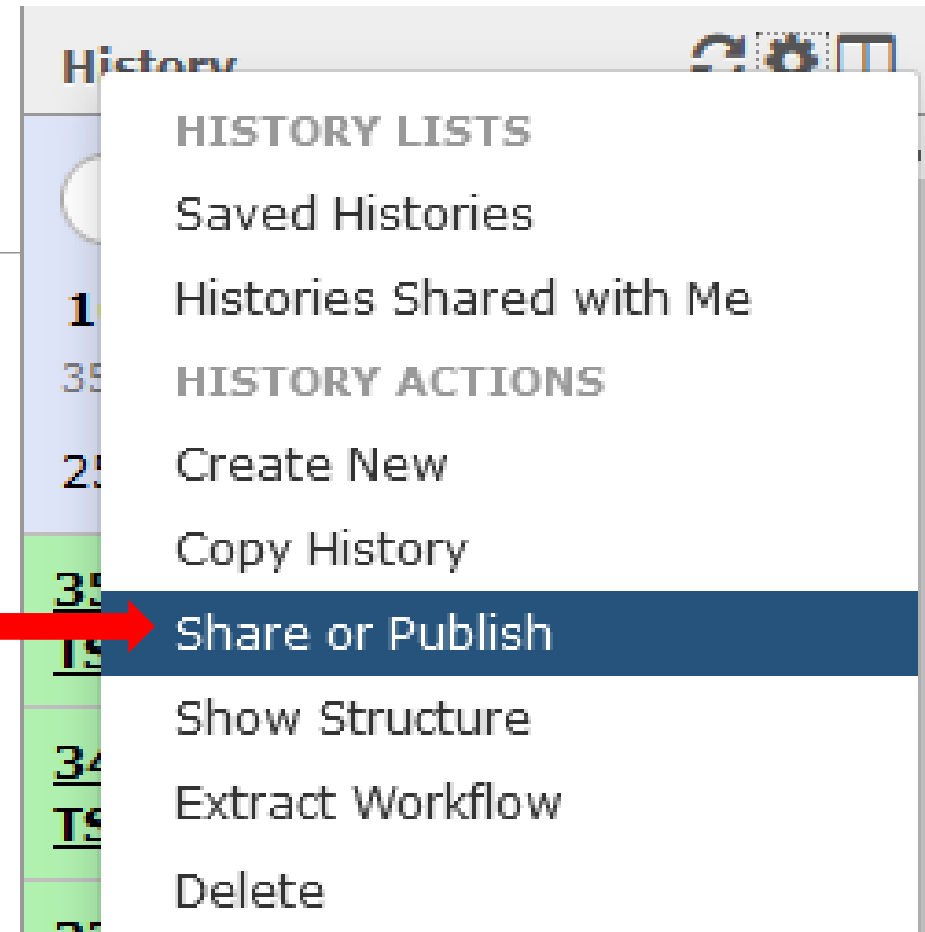
Share an history

Share or Publish History '16S'

Make History Accessible via Link and Publish It

This history is currently restricted so that only you and the users listed below can access it. You can:

| Make History Accessible via Link |

Generates a web link that you can share with other people so that they can view and import the history.

| Make History Accessible and Publish |

Makes the history accessible via link (see above) and publishes the history to Galaxy's Published Histories section, where it is publicly listed and searchable.

Share History with Individual Users

You have not shared this history with any users.

| Share with a user |

Back to Histories List

Share 1 histories

Histories to be shared:

History Name	Number of Datasets
ITS June 2021	26

Galaxy user emails with which to share histories

lucas

lucas.auer@inrae.fr

comma-separated list of addresses if sharing with multiple users

Submit

- You must click on the address in the dropdown menu.
- Click on submit.
- The history will appear in your co-worker's history.

Import a shared history

Click on the wheel and click here.

History

HISTORY LISTS

Saved Histories

Histories Shared with Me

HISTORY ACTIONS

Create New

Copy History

Share or Publish

Show Structure

How manipulate datasets

To rename a dataset

- Switch to ITS history

Click here to display attributes and change the name.

To rename a dataset

Attributes Convert Format Datatype Permissions

Edit Attributes

Name:

_FROGS/FROGS_ini/DATA/ITS.tar.gz

Info:

uploaded compressed archive file

Annotation / Notes:

Add an annotation or notes to a dataset; annotations are available whe

Database/Build:

unspecified (?)

Save

Auto-detect

This will inspect the dataset and attempt to correct the above column v

Change the name here and call it « ITS1.tar.gz »

To change the datatype

Attributes   Convert Format   Datatype   Permissions

Edit Attributes

Name:

http://genoweb.toulouse.inra.fr/~formatio

Info:

uploaded fastq file

If you put the wrong datatype in during the upload, you can change it here.

Annotation / Notes:

Add an annotation or notes to a dataset; annotations are available when a history is viewed.

Database/Build:

unspecified (?)

Save

Auto-detect

This will inspect the dataset and attempt to correct the above column values if they are not accurate.
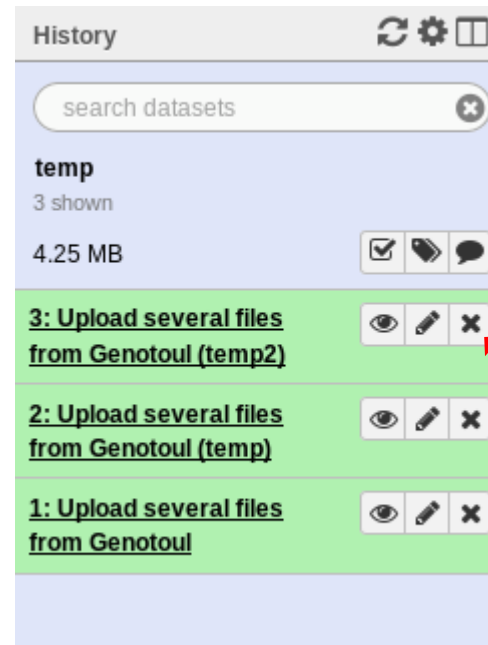
To delete a dataset

Click here to delete a dataset.

To delete a dataset

Click here.

Delete this dataset permanently.

Nothing is deleted in Galaxy until you delete it permanently (=purge).

How to delete a history?

- Stay in the temp history.

- Click on the wheel.

- Click on delete.

- A deleted item on Galaxy is recoverable.

- To definitively delete it:
  click on « Delete Permanently ».
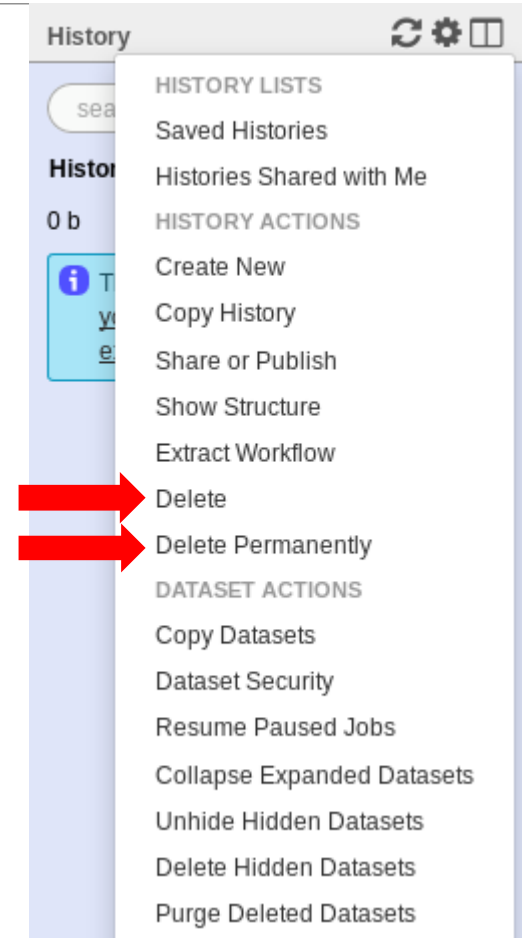
Other tools available on Galaxy

- Text Manipulation

- Filter and Sort data

- FASTA manipulation

Galaxy support

- Mail: support.sigenae@inrae.fr

- If you need more training about bioinformatics and Galaxy, please connect to Sigenae e-learning platform:
http://sig-learning.toulouse.inra.fr/

How to cite Galaxy?

- Goecks, J, Nekrutenko, A, Taylor, J and The Galaxy Team. *"Galaxy: a comprehensive approach for supporting accessible, reproducible, and transparent computational research in the life sciences."* Genome Biol. 2010 Aug 25;11(8):R86.

- Blankenberg D, Von Kuster G, Coraor N, Ananda G, Lazarus R, Mangan M, Nekrutenko A, Taylor J. *"Galaxy: a web-based genome analysis tool for experimentalists"*. Current Protocols in Molecular Biology. 2010 Jan; Chapter 19:Unit 19.10.1-21.

- Giardine B, Riemer C, Hardison RC, Burhans R, Elnitski L, Shah P, Zhang Y, Blankenberg D, Albert I, Taylor J, Miller W, Kent WJ, Nekrutenko A. *"Galaxy: a platform for interactive large-scale genome analysis."* Genome Research. 2005 Oct; 15(10):1451-5.

How to cite Genotoul Galaxy workbench?

▪ Research teams can thank the Toulouse Occitanie bioinformatics platform and Sigenae group, using in their publications the following sentence : " We are grateful to the genotoul bioinformatics platform Toulouse Occitanie (Bioinfo Genotoul, doi: 10.15454/1.5572369328961167E12) and Sigenae group for providing help and/or computing and/or storage ressources thanks to Galaxy instance http://sigenae-workbench.toulouse.inra.fr".

▪ In cases of collaboration, you can directly quote the person who participated to the project : Name, Sigenae group, GenPhySE, INRAE Auzeville CS 52627 31326 Castanet Tolosan cedex.