

A - Training on Galaxy: Metabarcoding

March 2024 - webinar

GALAXY Practice

LUCAS AUER, MARIA BERNARD, LAURENT CAUQUIL, VINCENT DARBOT, MAHENDRA MARIADASSOU, GÉRALDINE PASCAL & OLIVIER RUÉ

Objectives of the 1st part

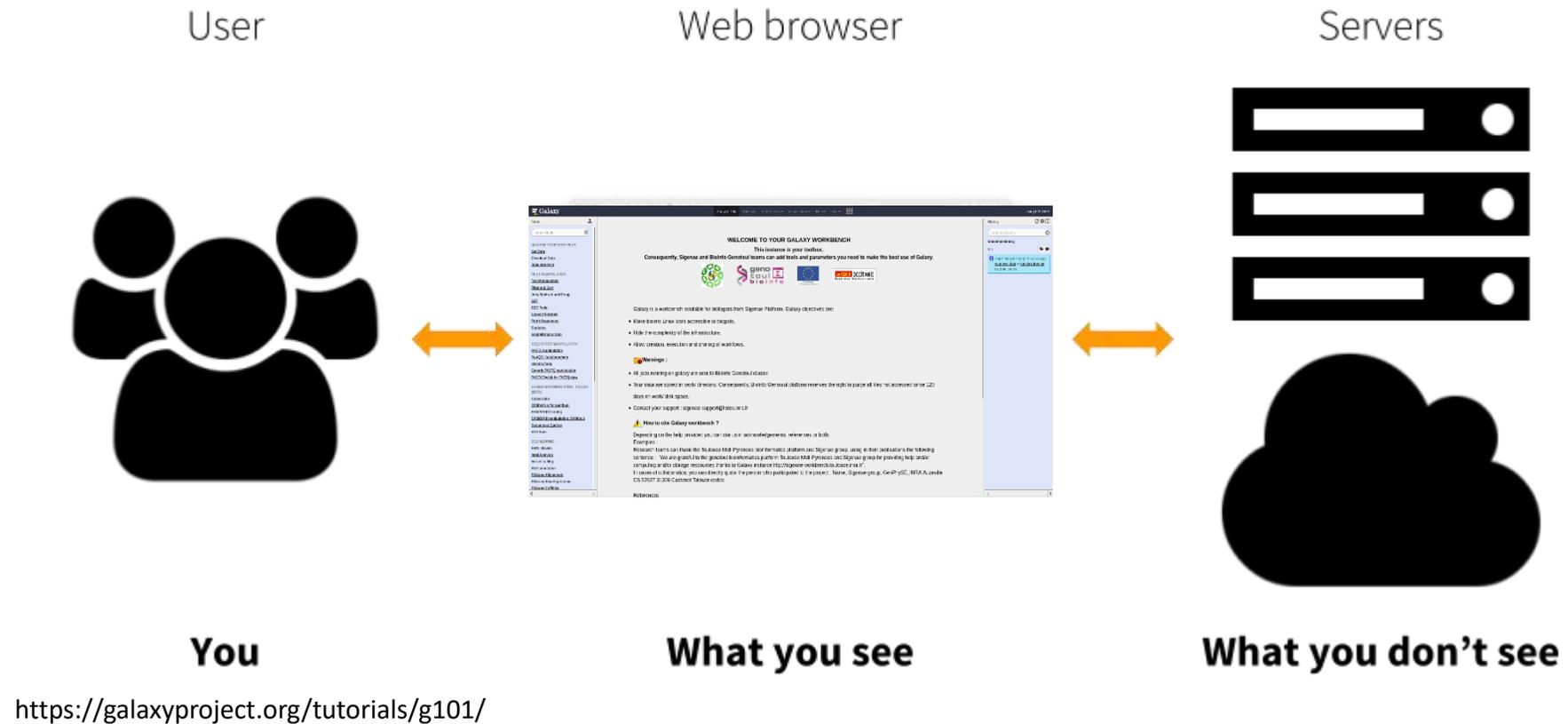
- Learn the basics of Galaxy
- Prepare the 1 dataset for FROGS formation

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



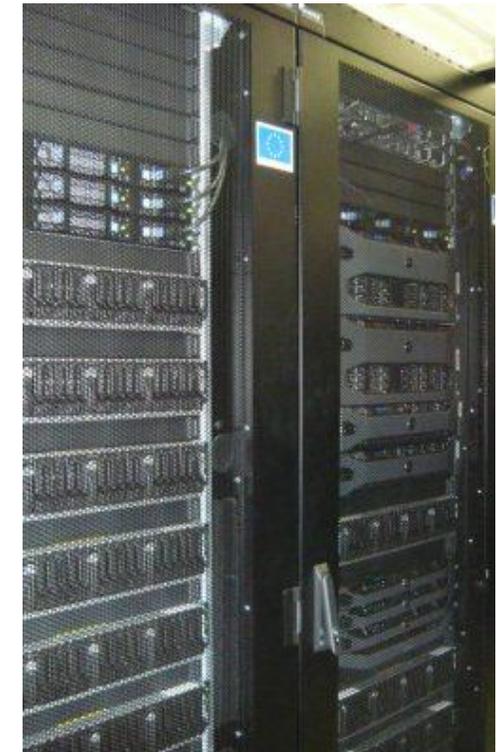
Where to use Galaxy?

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

The image displays three vertical panels, each representing a different Galaxy installation location. Each panel lists technical specifications and provides navigation buttons.

- Genotoul Bioinfo (Toulouse):**
 - Location: **Toulouse**
 - Compute (#CPU HT*): **6224**
 - Storage (#TB): **4400**
 - RAM (#GB): **36500**
 - RAM/core (#GB): **5.56**
 - GPU (#Card): **1**
- mig:ole (Jouy en Josas):**
 - Location: **Jouy en Josas**
 - Compute (#CPU HT*): **1016**
 - Storage (#TB): **350**
 - RAM (#GB): **7000**
 - RAM/core (#GB): **6.89**
 - GPU (#Card): **-**
- IFR CORE (Orsay (IDRIS)):**
 - Location: **Orsay (IDRIS)**
 - Compute (#CPU HT*): **4300**
 - Storage (#TB): **2400**
 - RAM (#GB): **20008**
 - RAM/core (#GB): **4.65**
 - GPU (#Card): **-**

Each panel includes a 'Demande de compte' button, a 'Support communautaire' button, and a 'Documentation' button. The bottom of each panel features the Galaxy logo and, in the case of IFR CORE, logos for R Studio and GitLab.



Where to use Galaxy?

- The Toulouse Galaxy platform: <https://vm-galaxy-prod.toulouse.inrae.fr>
- Institut Français Bioinformatique IFB (national server): <https://metabarcoding.usegalaxy.fr>
- Migale Platform: galaxy.migale.inra.fr
- Others...

 The different platforms of Galaxy are not connected together

- Galaxy is installed on many clusters across the world.
- Some tools are in our Galaxy platform but not in other platforms.
- Your data is not shared with other Galaxy platforms than ours.

Exemple of Galaxy platforms

Galaxy Migale | Analyse de données | Workflow | Visualize | Données partagées | Aide | Authentification

Tools | search tools

Get Data

Collection Operations

BASIC TOOLS

- Text Manipulation
- Convert Formats
- Filter and Sort
- Join, Subtract and Group
- Statistics
- Multiple Alignments
- ncbi_blast

NGS TOOLS

- Quality control
- FASTQ manipulation
- Mapping
- RNAseq
- Variant calling
- Variant analyses

Migale Tools

SEQUENCE ANALYSIS TOOLS

GENOME ANALYSIS TOOLS

- Genome annotation

METAGENOMICS TOOLS

- Metabarcoding

METAPROTEOMICS TOOLS

- Send Data
- Lift-Over

migale

Welcome to the Migale Galaxy instance!

Global load of the nodes dedicated to Galaxy jobs

5%

What's new

- **28th January 2021:** Added [Parnsnp](#) in Section *Phylogeny*.
- **28th January 2021:** Added [ECTyper](#) in Section *Genome Annotation*.
- **14th January 2021:** Added [Filter SPAdes Output](#) in Section *Assembly*.
- **27th November 2020:** Added [Seqtk](#) in Section *FAST/FASTQ Manipulation*.
- **27th November 2020:** Added [SeqSero](#) in Section *Genome Annotation*.
- **23d November 2020:** Added [staram](#) in Section *Genome Annotation*.
- **2nd October 2020:** Added [Shovill](#) in Section *Assembly*.
- **27th July 2020:** Added [Progressive Mauve](#) in Section *Multiple Alignements*.
- **27th July 2020:** Added [Mummer4](#) in Section *Multiple Alignements*.
- **24th July 2020:** Added [NCBI accession Download](#) in Section *Get Data*.

Exemple of Galaxy platforms

The screenshot displays the Galaxy Toulouse web interface. The top navigation bar includes 'Workflow', 'Visualize', 'Données partagées', 'Aide', 'Utilisateur', and a user profile icon. A status indicator shows 'Using 32%'. A notification banner at the top left states: 'New Galaxy server, needed tools/databanks are added on demand'.

The main content area features a large logo for 'Galaxy Toulouse Main' with a 'NEW' stamp. Below the logo, there is a 'go to Latest news' link and a paragraph of text: 'This new Galaxy server replaces the one set up in 2016 (version 16.01) and configured to submit jobs on a SGE HPC cluster using a global account.' Below this, there are sections for 'Support' (with email support.sigenae@inrae.fr) and 'How to cite use of our Galaxy instance?' (providing a citation example and references).

The left sidebar contains a 'Tools' section with a search bar and an 'Upload Data' button. Below this, there are two main categories: 'BASIC TOOLS' and 'ADVANCED TOOLS'. Under 'BASIC TOOLS', there is a list of tool categories: Monitoring, Get Data, Send Data, Collection Operations, Lift-Over, Text Manipulation, Convert Formats, Filter and Sort, Join, Subtract and Group, Fetch Alignments/Sequences, Operate on Genomic Intervals, Statistics, and Graph/Display Data. Under 'ADVANCED TOOLS', there is a 'FROGS' section with a sub-section 'OTUS RECONSTRUCTION'. Below this, there are two tool descriptions: 'FROGS Demultiplex reads' (Attribute reads to samples in function of inner barcode) and 'FROGS Pre-process merging, denoising and dereplication'.

The right sidebar shows a 'History' section with a search bar 'Rechercher des données' and a status '(empty)'. Below this, there is a blue information box: 'Cet historique est vide. You can Charger vos propres données or Charger des données depuis une source externe'.

Exemple of Galaxy platforms

The screenshot displays the Galaxy France web interface. At the top, a dark navigation bar contains the 'Galaxy France' logo, a home icon, and menu items for 'Workflow', 'Visualize', 'Données partagées', 'Aide', 'Utilisateur', and a grid icon. A 'Using 0%' indicator is on the right. The left sidebar, titled 'Tools', includes a search bar, an 'Upload Data' button, and a list of tool categories: 'Get Data', 'Collection Operations', 'GENERAL TEXT TOOLS', 'Text Manipulation', 'Join, Subtract and Group', 'FASTA/FASTQ', 'FASTQ Quality Control', 'METAGENOMICS', 'DADA2', 'FROGS', 'Qime', 'Obitools', 'PICRUST', 'Mothur', 'STATISTICS AND VISUALISATION', 'Graph/Display Data', 'Interactive Tools', 'WORKFLOWS', and 'All workflows'. The main content area features a 'Welcome to metabarcoding.usegalaxy.fr' message with a blue notification box stating: '13/01/2022: usegalaxy.fr is now running the release 21.09 of Galaxy. Please check the 21.09 user release notes.' Below this, it says 'Hosted by:' followed by logos for IFR, eLife, and a circular logo. A paragraph of text reads: 'The Galaxy Project is supported in part by NSF, NHGRI, The Huck Institutes of the Life Sciences, The Institute for CyberScience at Penn State, and Johns Hopkins University.' The right sidebar, titled 'History', shows a search bar 'Rechercher des données' and a list of recent jobs. The top job is 'test ITS' with 42 shown, 10 deleted, and 276.74 MB. Below it are several 'FROGSSTAT' jobs, including 'Deseq2 Visualisation: report.nb.html', 'DESeq2 Preprocess: dds.Rdata', 'Phylose q Multivariate Analysis Of Variance: manova.nb.html', 'Phylose q Structure Visualisation: structure.nb.html', 'Phylose q Sample Clustering: clustering.nb.html', 'Phylose q Beta Diversity: beta_diversity.nb.html (wunifrac.tsv)', 'Phylose q Beta Diversity: beta_diversity.nb.html (unifrac.tsv)', and 'Phylose'.

Practice:

CONNECT TO GALAXY WORKBENCH

Exercise

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

Welcome to Galaxy, please log in

Public Name or Email Address

Password

Forgot password? [Click here to reset your password.](#)

Login

Don't have an account? Registration for this Galaxy instance is disabled. Please contact an administrator for assistance.

Tools

search tools

Upload Data

FROGS

FROGS Demultiplex reads Attribute reads to samples in function of inner barcode.

FROGS Pre-process merging, denoising and dereplication.

FROGS Clustering swarm Single-linkage clustering of OTUs based on

AVAILABLE TOOLS

FROGS Remove chimera Remove PCR chimera from OTUs.

FROGS OTU Filters Filters OTUs on several criteria.

FROGS ITSx Extract the highly variable ITS1 and ITS2 subregions from ITS sequences.

FROGS Affiliation OTU Taxonomic affiliation of each OTU's seed by RDP tools and BLAST

Optional step to resolve inclusive amplicon ambiguities and to aggregate OTUs based on alignment metrics

FROGS Abundance normalisation

FROGS Pre-process merging, denoising and dereplication. (Galaxy Version 3.2.3+galaxy2)

Sequencer

Illumina

Select the sequencing technology used to produce the sequences.

Input type

Files by samples

Samples files can be provided in a single TAR archive or sample by sample (with one or two files each).

Are reads already merged ?

No

The inputs contain 1 file by sample : R1 and R2 pair are already merged in one sequence.

Samples

1: Samples

Name

The sample name.

Please provide a value for this option.

Reads 1

No fastq dataset available.

Reads 2

No fastq dataset available.

R2 FASTQ file of paired-end reads.

Insert Samples

History

Rechercher des données

16S

2 shown

34.86 MB

2: chaillou_withprimers_64renamedsamples_V1V3_10000seq_R1R2.tar.gz

1: metadata_chaillou.tsv

DATASETS HISTORY

RESULTS VISUALISATION AND TOOL PARAMETER WINDOW

This is your working environment for analysing your data

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.

Galaxy France

Workflow Visualize Données parta

Tools

search tools

Show panel options

Upload Data

FROGS_0 Demultiplex reads
Attribute reads to samples in function of inner barcode

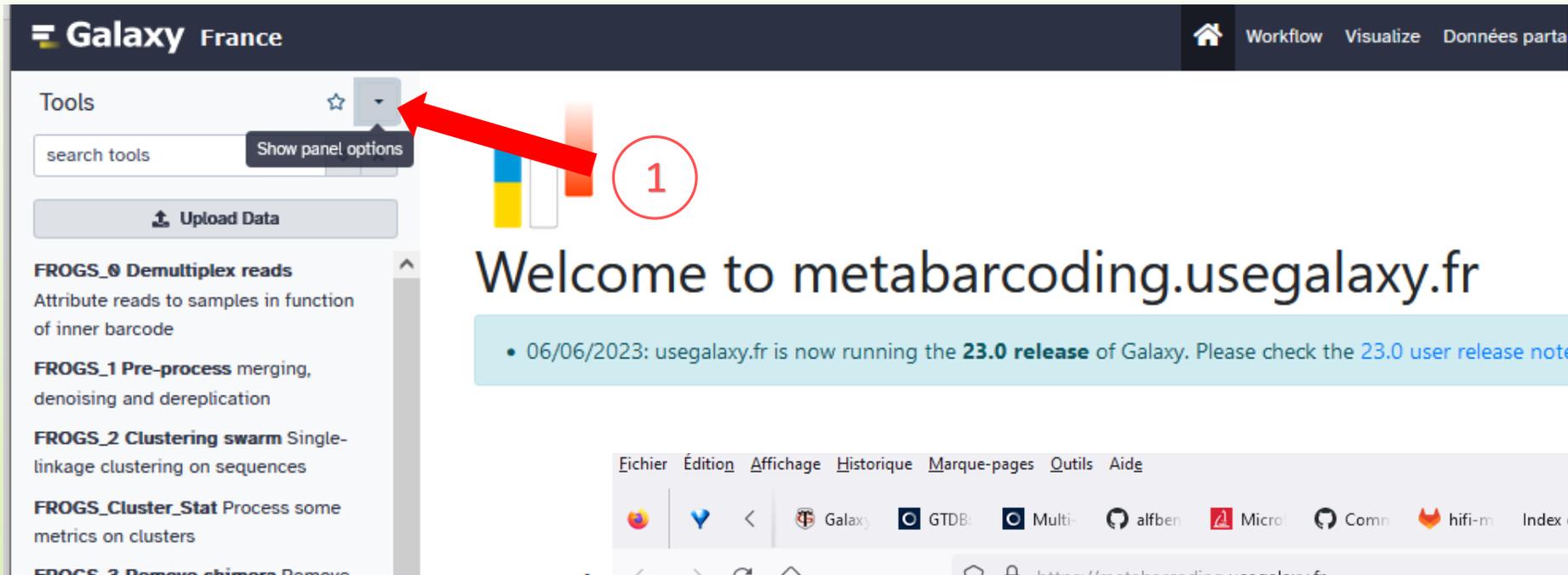
FROGS_1 Pre-process merging, denoising and dereplication

FROGS_2 Clustering swarm Single-linkage clustering on sequences

FROGS_Cluster_Stat Process some metrics on clusters

Welcome to metabarcoding.usegalaxy.fr

06/06/2023: usegalaxy.fr is now running the **23.0 release** of Galaxy. Please check the [23.0 user release note](#)



Fichier Édition Affichage Historique Marque-pages Outils Aide

Galaxy GTDB Multi- affben Micro Comn hifi-m Index of /f Jaweb auth0.ope EZ

https://metabarcoding.usegalaxy.fr

Les plus visités 22 02 Messagerie #Temps Galaxy-Prod Galaxy-dev Galaxy | France GitHub FROGS: home FROGS - Hac

Galaxy France

Workflow Visualize Donn

Tools

Full Tool Panel

...by Ontology

EDAM Operations

EDAM Topics

Metabarcoding: FROGS

Metabarcoding: FROGS 4.1.0

FROGS_2 Clustering swarm Single-linkage clustering on sequences

FROGS_Cluster_Stat Process some metrics on clusters

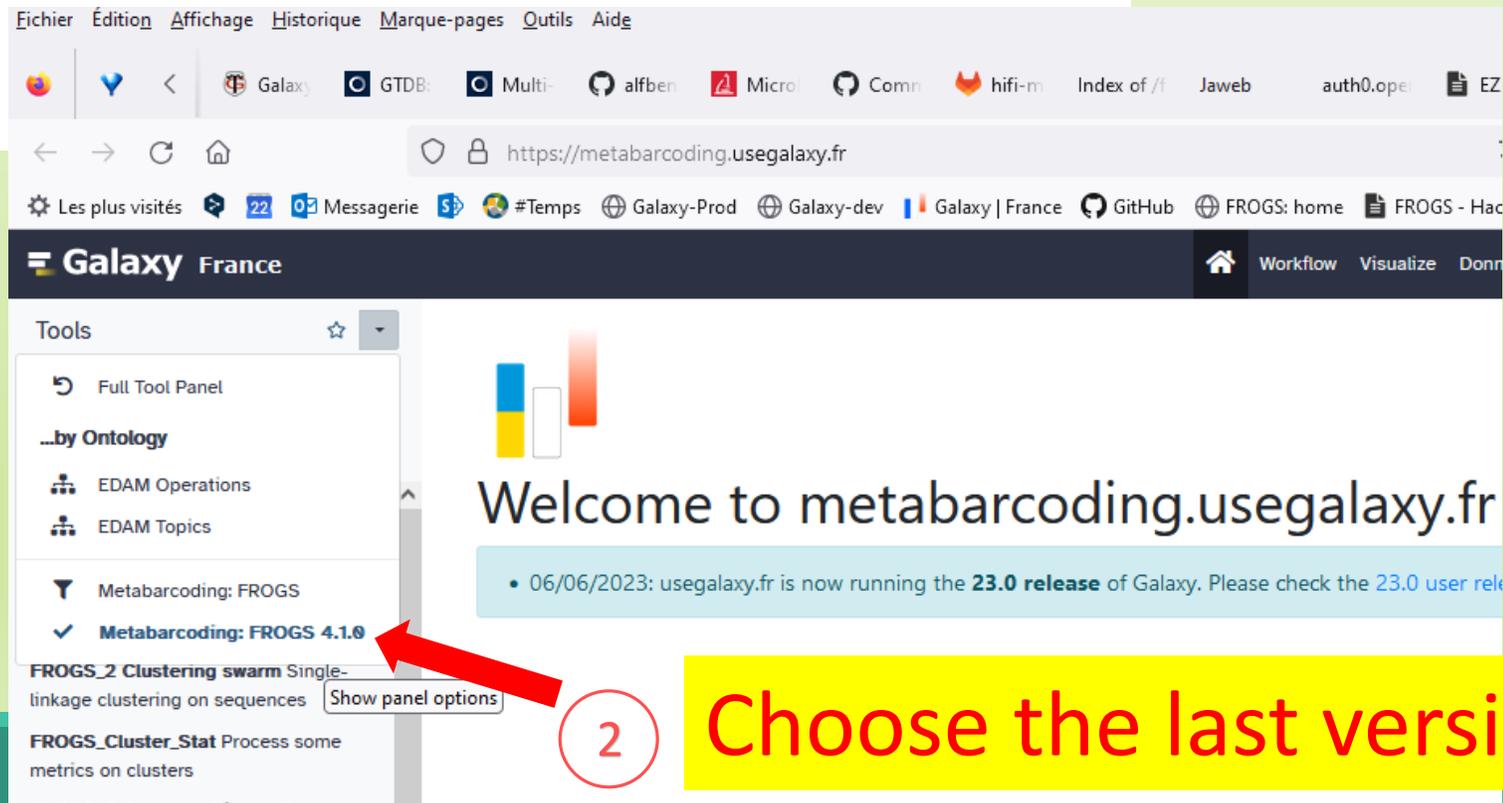
Show panel options

Welcome to metabarcoding.usegalaxy.fr

06/06/2023: usegalaxy.fr is now running the **23.0 release** of Galaxy. Please check the [23.0 user rel](#)

2

Choose the last version 4.1.0



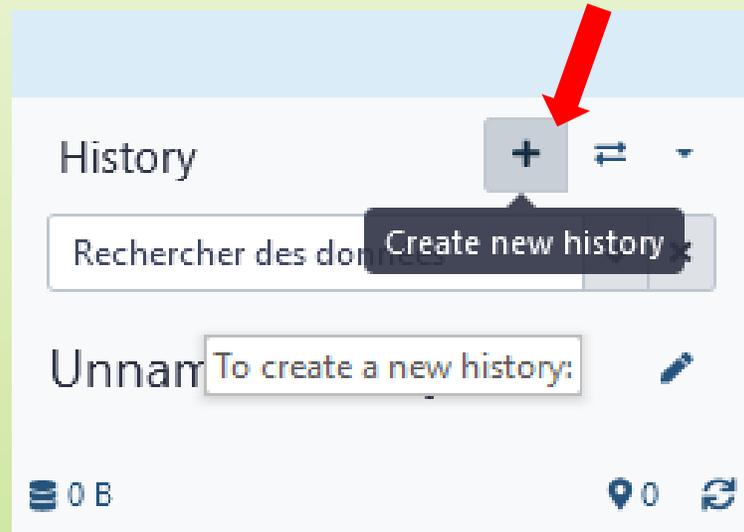
Manipulate Histories

Practice:

CREATE THE NECESSARY HISTORY FOR FROGS TRAINING

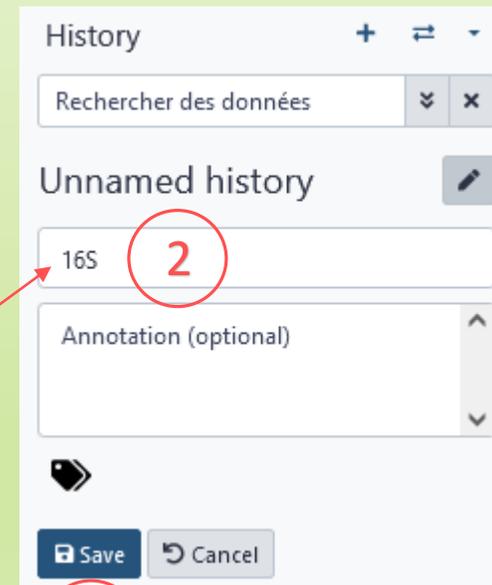
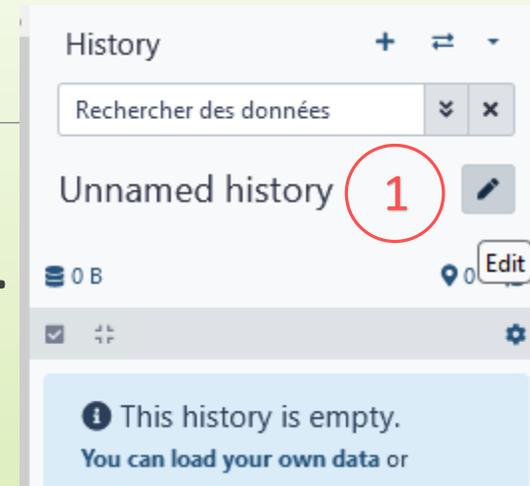
To create a new history:

- Click on the plus.



To rename a history:

- Click on the pen next to “Unnamed history” (top right).
- Enter « 16S ».
- Hit « **Enter** » to validate.

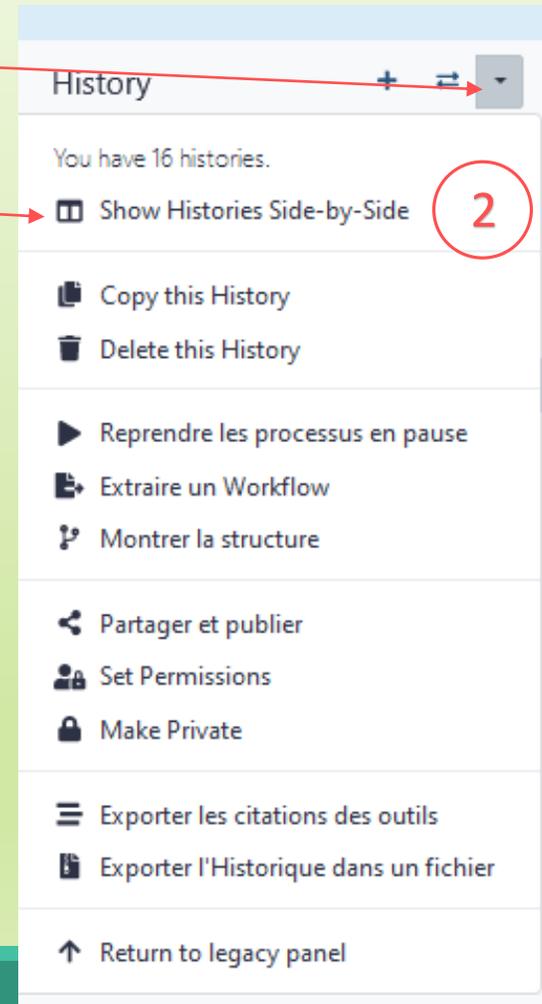


Exercise

- Create another history as before with 16S. This new history will be named **ITS**

How to list all histories?

1. To view all histories, click on this arrow
2. And click on



Explore the « View all histories » section

.fasta

16S

36.5 MB

No data found for selected filter.

- 37 : FROGS OTU Filters: otuFilter_sequences.fasta
- 27 : FROGS Affiliation Filters: affiFilter_sequences.fasta
- 15 : FROGS OTU Filters: otuFilter_sequences.fasta
- 10 : FROGS Remove chimera: non_chimera.fasta
- 6 : FROGS Clustering swarm: seed_sequences.fasta
- 3 : FROGS Pre-process: dereplicated.fasta
- 56 : FROGS ITSx: ITS_sequence.fasta
- 55 : FROGS ITSx: nonITS_sequence.fasta
- 52 : FROGS ITSx: ITS_sequence.fasta
- 51 : FROGS ITSx: nonITS_sequence.fasta
- 48 : FROGS ITSx: ITS_sequence.fasta
- 47 : FROGS ITSx: nonITS_sequence.fasta
- 44 : FROGS ITSx: ITS_sequence.fasta
- 43 : FROGS ITSx: nonITS_sequence.fasta
- 42 : Galaxy29-[FROGS_4_Cluster_filters_clusterFilters_sequences.fasta].fasta
- 35 : FROGSFUNC_step1_placeseqs: frogfunc_placeseqs.fasta
- 25 : FROGS ITSx: ITS_sequence.fasta

Select histories

Switch to Hide

Switch to Hide

Switch to Hide

Search dataset among displayed history by keywords

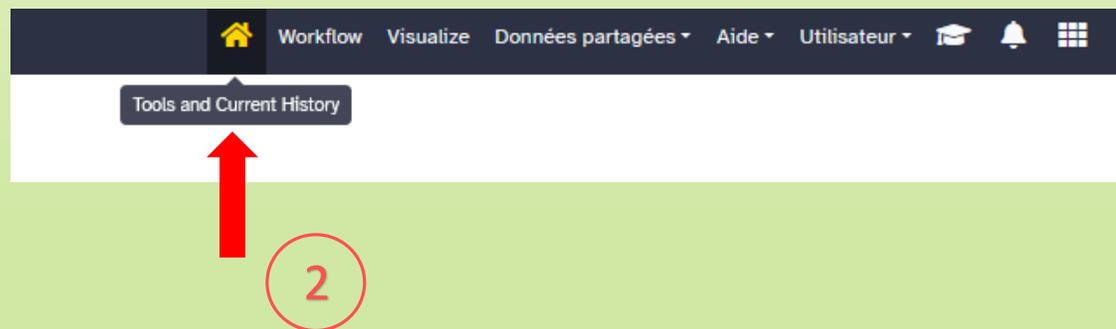
Switch to pass this history as your current history

Search among all your histories

Switch current history

1. When you are several histories, switch to the history named « 16S » as current history

2. You can click on “home” 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



drag-and-drop files

- By URL



Paste an url

- 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

<u>Name</u>	<u>Last modified</u>	<u>Size</u>	<u>Description</u>
 Parent Directory	-	-	-
 replicate information chaillou.tsv	2023-03-31 10:50	310	
 metadata chaillou.tsv	2023-03-31 10:50	1.4K	
 chaillou withprimers 64renamedsamples V1V3 10000seq R1R2.tar.gz	07-Mar-2021 15:34	35M	
 ITS fast replicates.tsv	31-Mar-2023 10:50	310	
 ITS fast metadata.tsv	31-Mar-2023 10:50	1.4K	
 ITS fast.tar.gz	31-Mar-2023 10:50	54M	
 How to put your data on genotoul server.pdf	11-Mar-2021 16:01	1.5M	
 How create an archive.pdf	05-Apr-2023 10:55	502K	

Datasets for 16S history

Datasets for ITS history

All samples are in archive .tar.gz

Remark: how create an archive ?

Not for this session, but for your own data, it will be important to know 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](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.

Practice: 16S history

UPLOAD FILE FROM AN URL

How to import your data to Galaxy ?

- Several ways to upload your data to Galaxy:

- From your computer

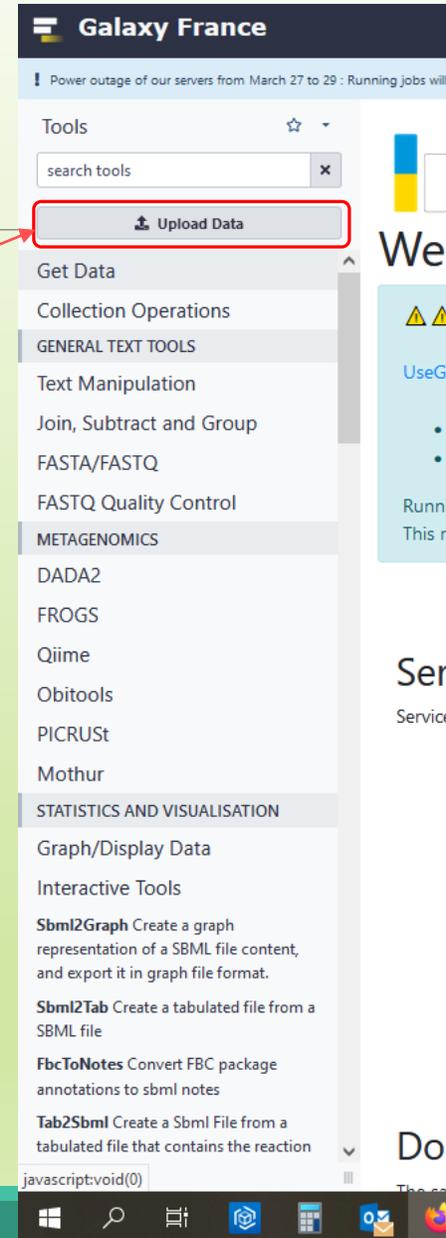


- By URL



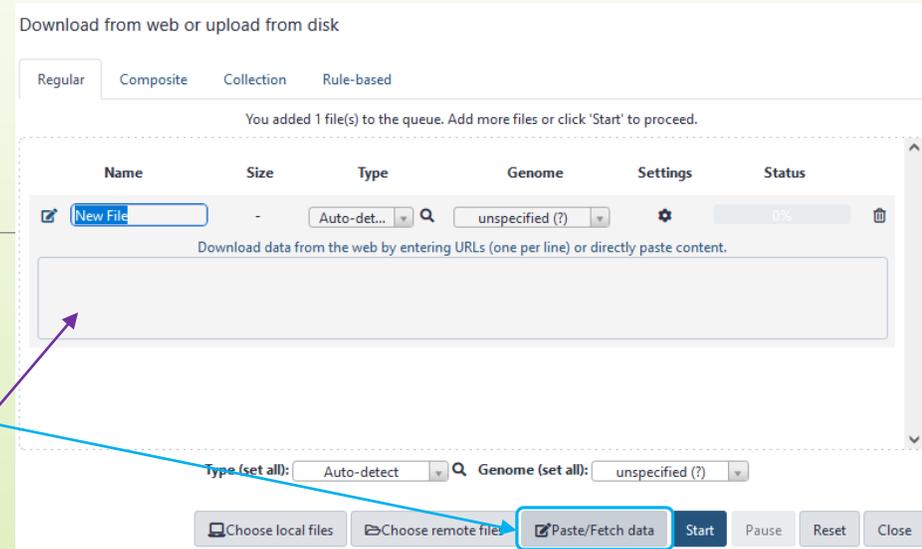
16S history creation

1. In your current history “16S” (If it not the case, switch to 16S history)
2. Go to **Upload Data** on the left panel



16S history creation

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

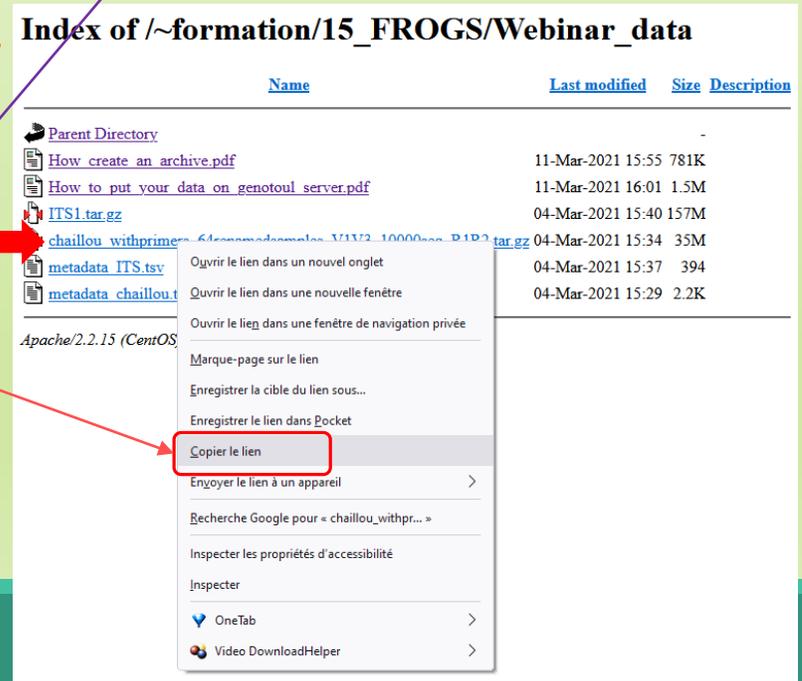


4. Collect your data address on [http://genoweb.toulouse.inra.fr/~formation/15 FROGS/Webinar data/](http://genoweb.toulouse.inra.fr/~formation/15_FROGS/Webinar_data/)

Click right on the name file

5. Copy the address of the link

6. Paste the link in the grey window

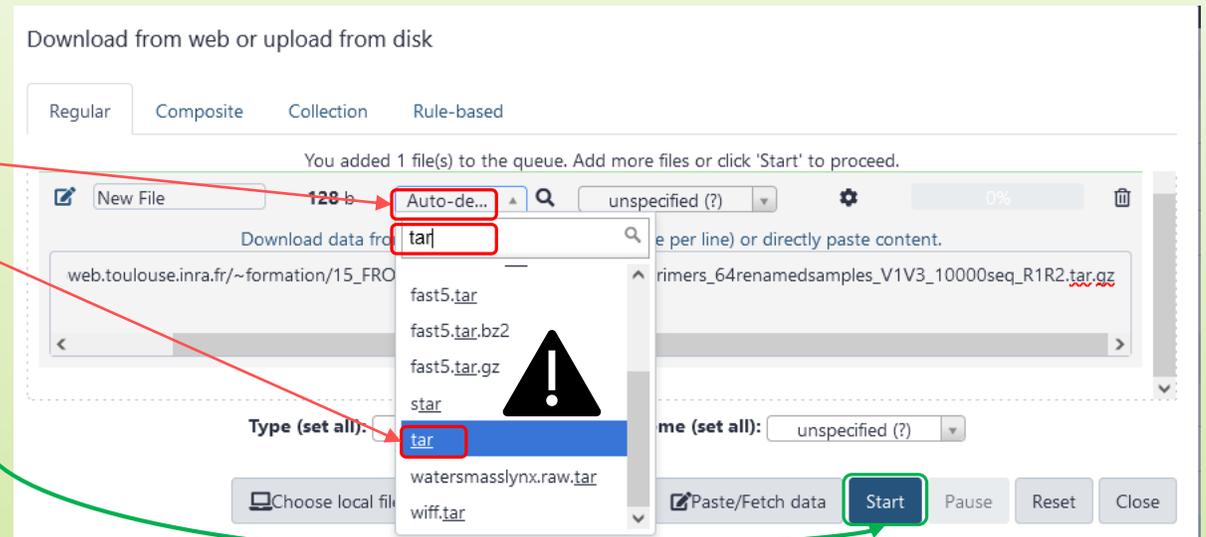


16S history creation

7.  Change the data type!

8. Click on Start

9. Start again with the file metadata_chaillou.tsv



Index of /~formation/15_FROGS/Webinar_data

	Name	Last modified	Size
	Parent Directory	-	-
	How create an archive.pdf	11-Mar-2021 15:55	781K
	How to put your data on genotoul server.pdf	11-Mar-2021 16:01	1.5M
	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	36K
	metadata_chaillou.tsv	04-Mar-2021 15:29	2.2K

**Remember to
change your
data type to TSV**

16S history creation

Share a History

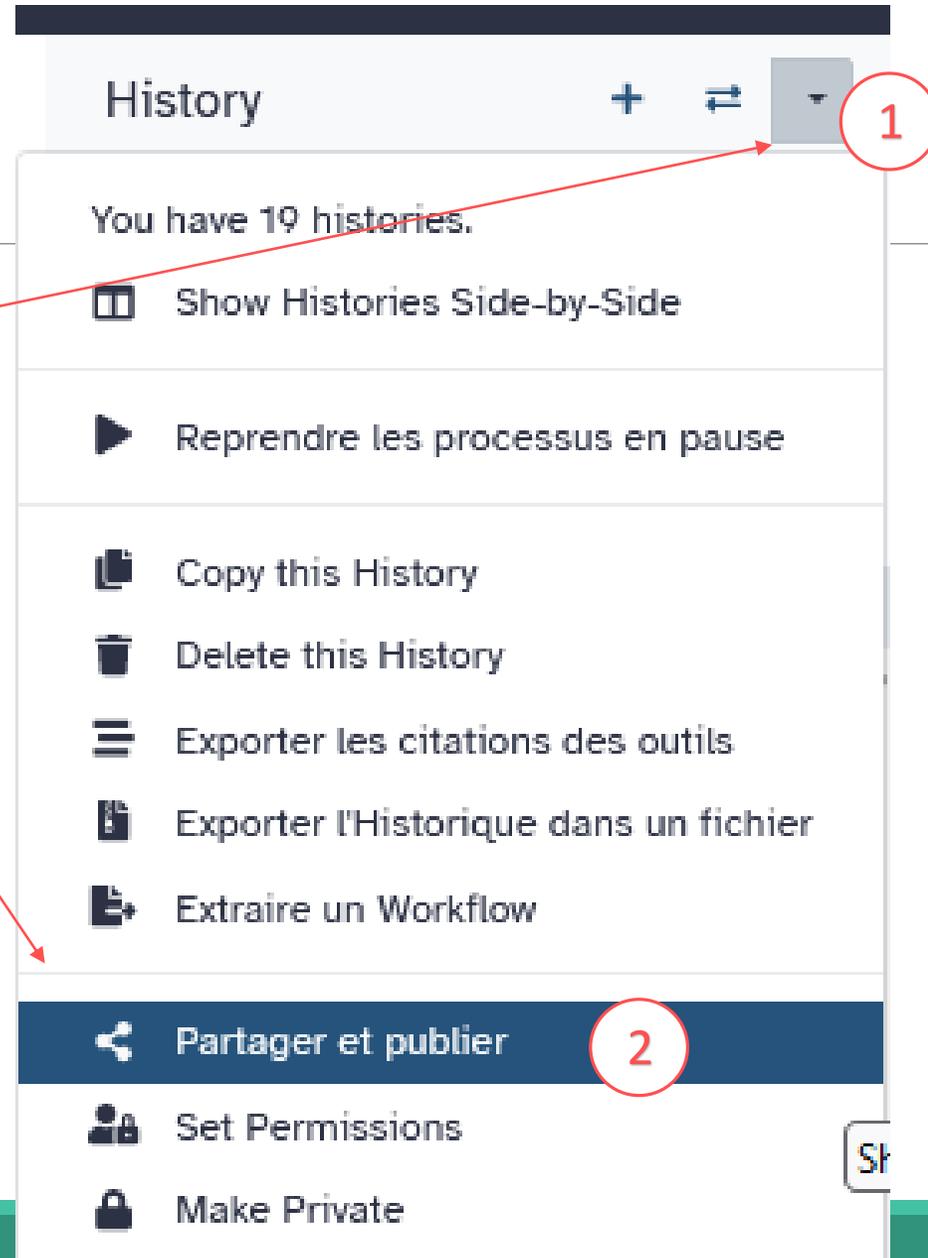
Share a history

- 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 a history

To share an history

1. Click on this arrow
2. Click on share and publish



Share a history

Share or Publish History `test frogs sans clustering`

Make History accessible

Access to this History is currently restricted so that only you and the users listed below can access it. Note that sharing a History will also allow access to all of its

To share click on and enter a user

Share History with Individual Users ▾

The following users will see this History in their History list and will be able to view, import and run it.

lucas.auer@inrae.fr ✕

Cancel Save

To unshare click on the cross

History + ⇄ ▾

Rechercher des données ▾ ✕

test FROGSFUNC ✎

8.64 MB 📍 11 ↻

☑ ⇄ ⚙

11 : FROGSFUNC_step2_cop 👁 ✎ 🗑
ynumbers: frogsfunc_copyn
umbers_marker.tsv

10 : FROGSFUNC_step2_co 👁 ✎ 🗑
pynumbers: frogsfunc_copy
numbers_predicted_functio
ns.tsv

Share a history

Share or Publish History `test frogs sans clustering`

- Make History accessible
- Make History publicly available in Published Histories

to make your history public and visible to the whole world, click here

This History is currently **accessible via link**.

Anyone can view and import this History by visiting the following URL:

  url: <https://metabarcoding.usegalaxy.fr/u/geraldine.pascal/h/test-frogs-sans-clustering>

Practice: share a history

SHARE 16S HISTORY WITH TRAINERS

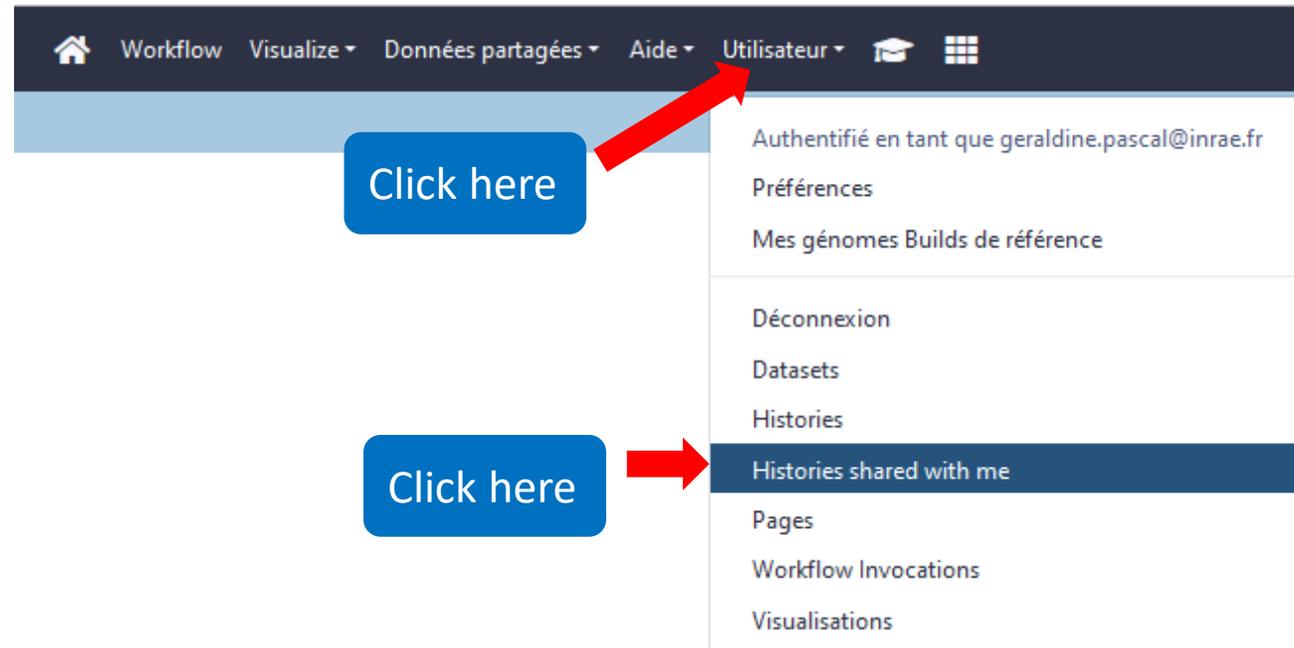
Share your data with us



- When your history “16S” is ready and contains the 2 files
 1. chaillou_withprimers_64renamedsamples_V1V3_10000seq_R1R2.tar.gz
 2. metadata_chaillou.tsv
- Then share with `lucas.auer@inrae.fr` × `geraldine.pascal@inrae.fr` ×
- Contact us by email frogs-training@inrae.fr when it is done.

Other useful things

Import a shared history



Histories shared with you by others

Name	Datasets	Created	Last Updated	Shared by
MG2 0.1	53 4	Feb 17, 2021	Feb 19, 2021	laurent.cauquil@inra.fr
16S_Test	54 3	Jul 09, 2020	Jul 10, 2020	auer@insa-toulouse.fr

View
Copy
Unshare

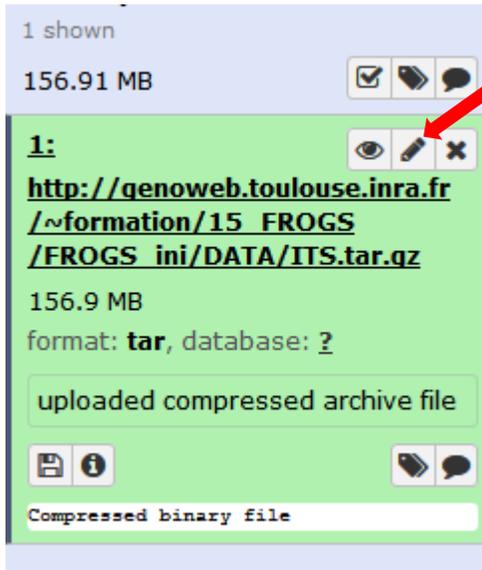
- Click on the name and click on « Copy » to import an history into your Galaxy account.
- To unshare click on unshare

The person who shared the history.

How manipulate datasets

To rename a dataset

Click here to display attributes and change the name.



1 shown
156.91 MB

1:

http://genoweb.toulouse.inra.fr/~formation/15_FROGS/FROGS_ini/DATA/ITS.tar.qz

156.9 MB
format: **tar**, database: ?

uploaded compressed archive file

Compressed binary file

The screenshot shows a dataset entry in a table. The entry is highlighted in green. It has a pencil icon next to the name '1:', indicating it can be edited. A red arrow points from a blue callout box to this pencil icon.

To rename a dataset

Change the name here

The screenshot shows a web interface for editing a dataset. At the top, there are three tabs: 'Attributes' (selected), 'Convert', and 'Permissions'. Below the tabs, the 'Name' field contains the text 'FROGS_2 Clustering swarm: seed_sequences.fasta'. A blue callout box on the left with the text 'Change the name here' has three red arrows pointing to the 'Attributes' tab, the 'Name' field, and the 'Save' button. Below the 'Name' field is the 'Info' section, which includes a '## Application' field with the value 'Software :/shared/ibfstor1/galaxy/mutable-data/dependencies/_cor'. Below that is an 'Annotation - optional' section with a text area. At the bottom, there is a 'Database/Build - optional' field with the value 'unspecified (?)' and two buttons: 'Save' and 'Auto-detect'.

Attributes Convert Permissions

Name

FROGS_2 Clustering swarm: seed_sequences.fasta

Info

Application
Software :/shared/ibfstor1/galaxy/mutable-data/dependencies/_cor

Annotation - optional

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

Database/Build - optional

unspecified (?)

Save Auto-detect

To change the datatype

Attributes Convert Permissions

Convert

Target datatype

tabular (using 'Convert FASTA to Tabular')

This will create a new dataset with the contents of this dataset

Create Dataset

Datatypes

New Type

fasta

This will change the datatype of the existing dataset but not modify its contents

Save Auto-detect

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

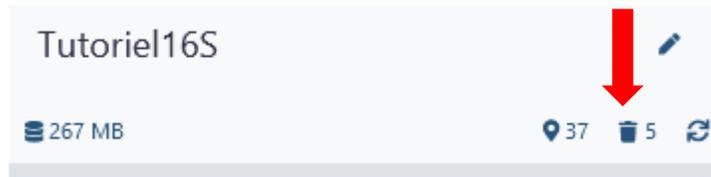
To delete a dataset

42 : FROGS BIOM to TSV: abundance.tsv	  
41 : FROGS Affiliations stat: report.html	  
40 : cleaned_biom-2023-02-13.biom	  



Click here to delete a dataset.

To see deleted dataset



Click here to see deleted dataset.



Here the complete list of deleted dataset.

To undelete dataset

History

deleted:true

Tutoriel16S

267 MB 37 5

42 : FROGS BIOM to TSV: abundance.tsv

35 : FROGS OTU Filters: report.html

34 : FROGS OTU Filters: excluded.tsv

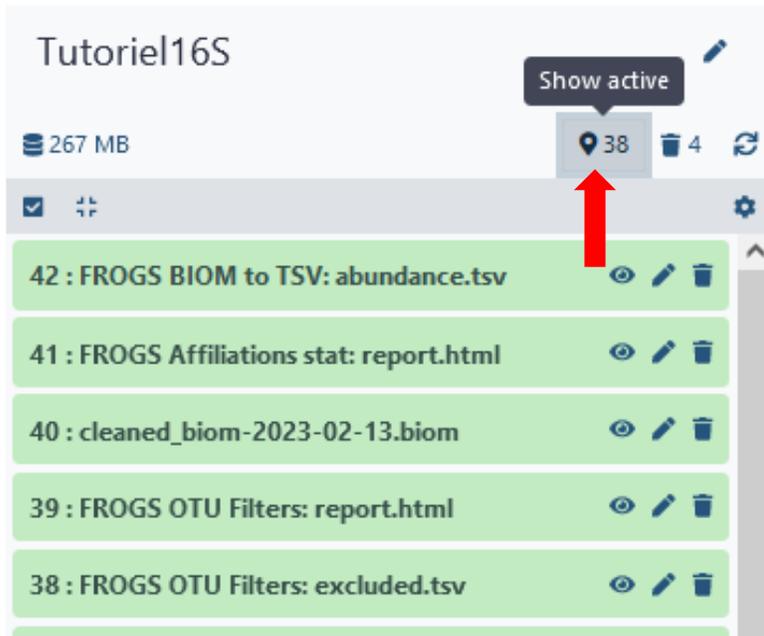
33 : FROGS OTU Filters: otuFilter_sequences.fasta

32 : FROGS OTU Filters: otuFilter_abundance.biom

Click here and this file return into the history

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

To undelete dataset



Click here to see again your active history

How to cite Galaxy workbench?

For Genotoul Bioinfo:

- 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.

For IFB:

- We are grateful to the Institut Français de Bioinformatique (ANR-11-INBS-0013) for providing help and/or computing and/or storage resources.

For Migale:

- Research teams that have used Migale services are expected to thank the Migale bioinformatics facility in the publications of their analyses with the following sentence: "We are grateful to the INRAE MIGALE bioinformatics facility (MIGALE, INRAE, 2020. Migale bioinformatics Facility, doi: 10.15454/1.5572390655343293E12) for providing help and/or computing and/or storage resources"
- In cases of collaboration, you can directly quote the person who participated to the project with the two following affiliations: Université Paris-Saclay, INRAE, MaIAGE, 78350, Jouy-en-Josas, France Université Paris-Saclay, INRAE, BioinfOmics, MIGALE bioinformatics facility, 78350, Jouy-en-Josas, France

Thank you for following this first step, see you on Monday 18 March!

YOUR TRAINERS:
GÉRALDINE & LUCAS

