

A - Training on Galaxy: Metabarcoding

April 2023 - webinar

GALAXY Practice

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

Objectives of the 1st part

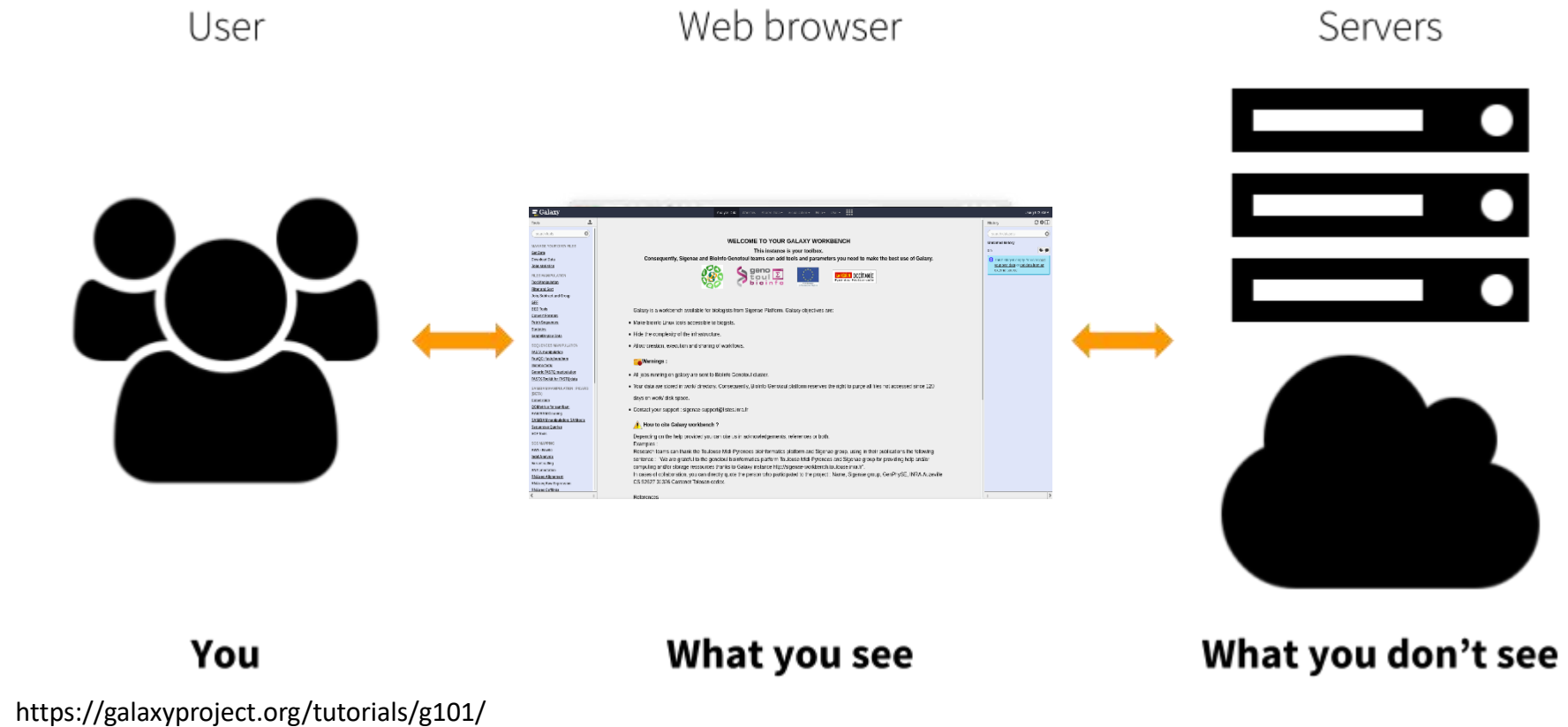
- Learn the basics of Galaxy
- Prepare the 1 dataset for FROGS formation

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!

5%

Global load of the nodes dedicated to Galaxy jobs

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 Alignments*.
- **27th July 2020:** Added [Mummer4](#) in Section *Multiple Alignments*.
- **24th July 2020:** Added [NCBI accession Download](#) in Section *Get Data*.

Exemple of Galaxy platforms


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

Tools Panel (Left):

- Search tools:
- Upload Data:
- BASIC TOOLS**
 - 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
 - Graph/Display Data
- ADVANCED TOOLS**
- FROGS**
 - OTUS RECONSTRUCTION**
 - FROGS Demultiplex reads** Attribute reads to samples in function of inner barcode.
 - FROGS Pre-process** merging, denoising and dereplication.

Main Content Area:

Welcome to new main Galaxy-Toulouse server



[go to Latest news](#)

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.

Support
Support is configured in Galaxy to send help requests to support.sigena@inrae.fr
Use this address to contact us in case of problems

How to cite use of our Galaxy instance?
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 Sigena group, using in their publications the following sentence:
"We are grateful to the genotoul bioinformatics platform Toulouse Midi-Pyrenees and Sigena group for providing help and/or computing and/or storage resources thanks to Galaxy instance <https://vm-galaxy-prod.toulouse.inra.fr>".
In cases of collaboration, you can directly quote the person who participated to the project, contact him to have his accurate contact details for quote.
References: X. SIGENAE [<http://www.sigena.org/>]

The main changes are:

- new version of Galaxy server: 21.05
- reorganization and cleaning of tool list with recent versions
- submit jobs on the new Slurm HPC cluster, using connected Galaxy user's account and CPU quota.
- new storage volume for Galaxy datasets, with a quota of 250GB per user

History Panel (Right):

History:

Unnamed history
(empty)

Message: 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, the navigation bar includes 'Workflow', 'Visualize', 'Données partagées', 'Aide', 'Utilisateur', and a 'Using 0%' indicator. The left sidebar contains a 'Tools' section with a search bar and an 'Upload Data' button, followed by categories like 'Get Data', 'Collection Operations', 'GENERAL TEXT TOOLS', 'METAGENOMICS', 'STATISTICS AND VISUALISATION', and 'WORKFLOWS'. The main content area features a 'Welcome to metabarcoding.usegalaxy.fr' message and a notification about the release 21.09. Below this, it lists the hosting institutions: IFB, eLife, and the University of Pennsylvania. The right sidebar shows a 'History' panel with a search bar and a list of recent jobs, including '52: FROGSSTAT Deseq2' and '49: FROGSSTAT DESeq2'.

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:

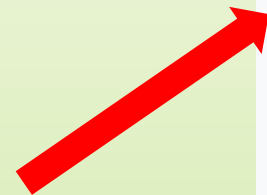
DISCOVER GALAXY

Exercise

1. Visit the Galaxy Platform.
2. Look at the tool list.
3. Display only FROGS tools.
4. Display all tools concerning fastq files.

Exercise

Search a tool by name.



Tools ☆

 ✕

📁 Upload Data

👁 Show Sections

FROGS BIOM to std BIOM Converts a FROGS BIOM in fully compatible BIOM.

FROGSSTAT Phyloseq Structure Visualisation with heatmap plot and ordination plot

FROGS TSV_to_BIOM Converts a TSV file in a BIOM file.

FROGS Affiliation Filters Filters OTUs on several affiliation criteria.

FROGS Tree Reconstruction of phylogenetic tree



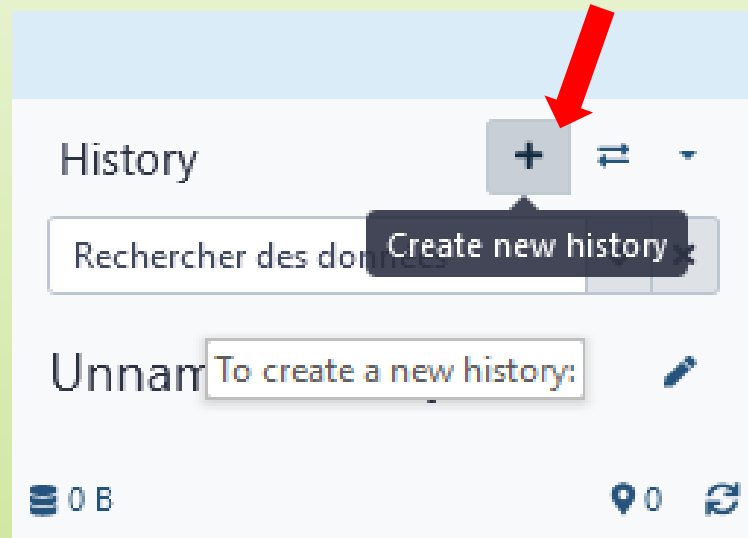
Manipulate Histories

Practice:

CREATE THE NECESSARY HISTORY FOR FROGS TRAINING

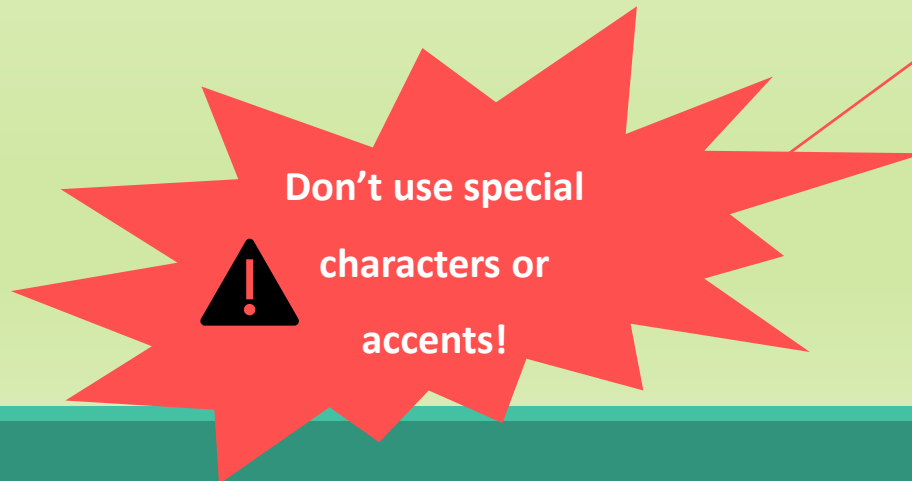
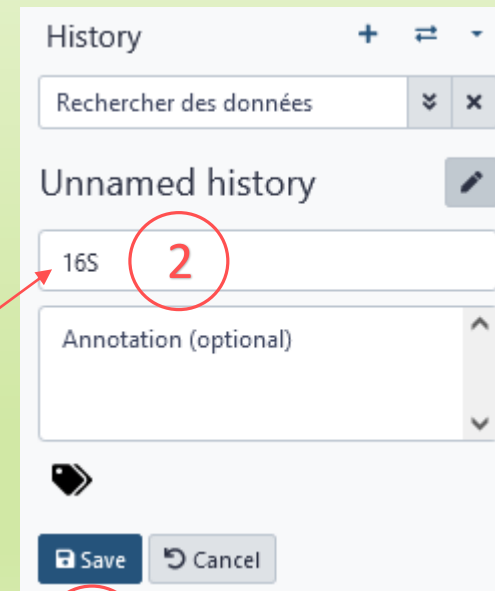
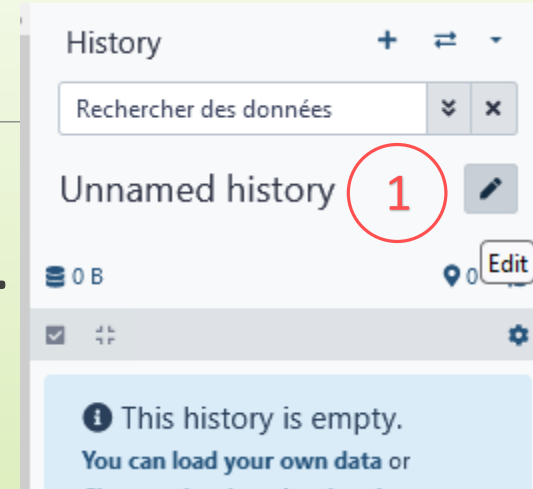
To create a new history:

- Click on the plus.



To rename a history:

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

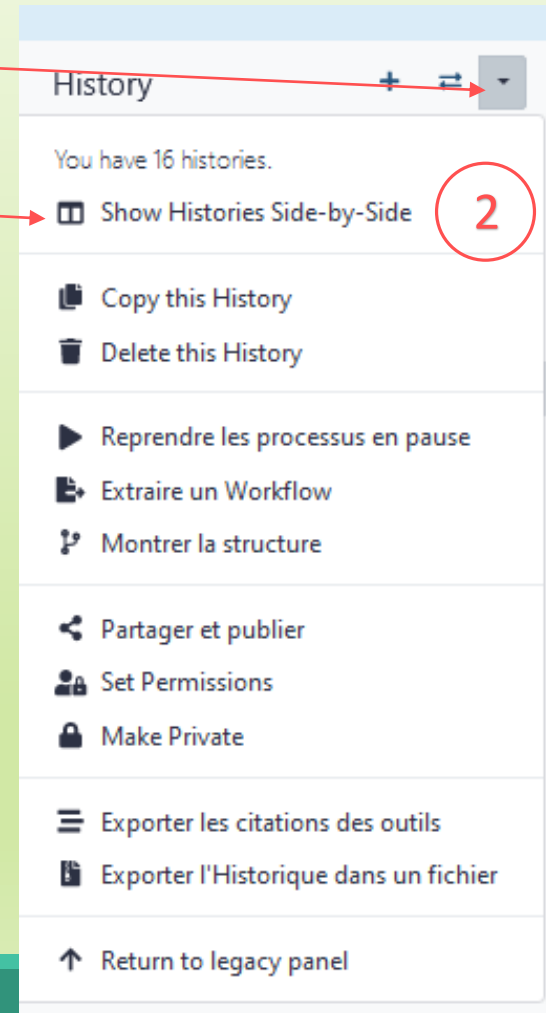


Exercise

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

How to list all histories?

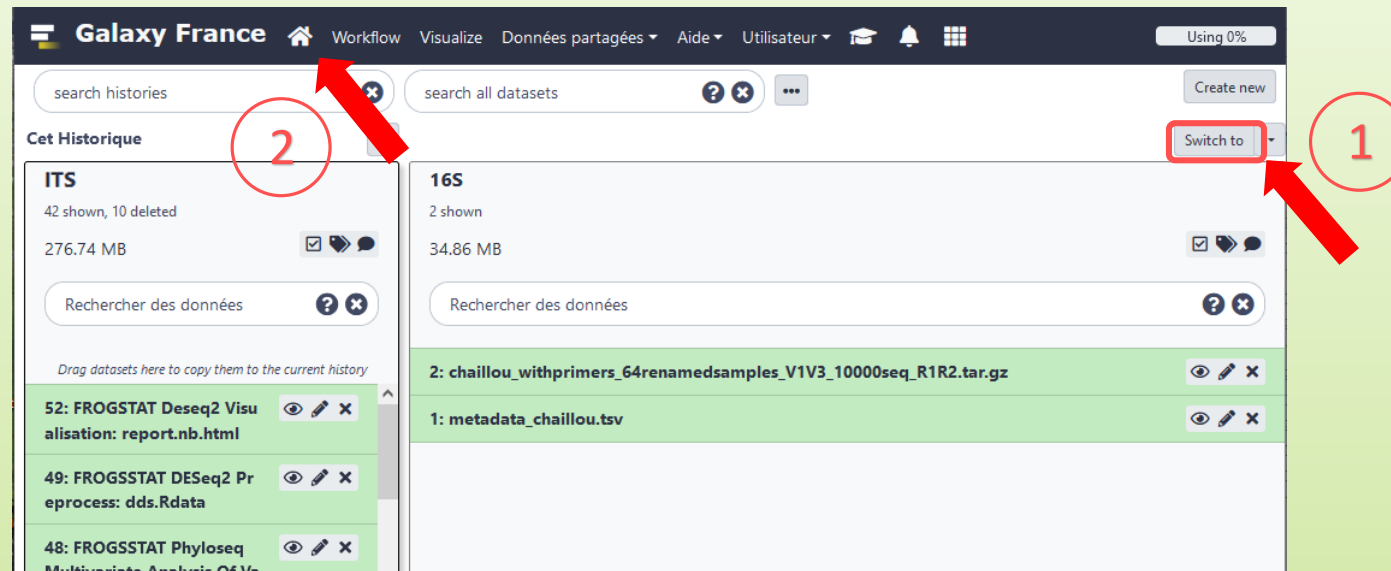
1. To view all histories, click on this arrow
2. And click on



Explore the « View all histories » section

The screenshot displays the Galaxy France web interface. At the top, there is a navigation bar with the logo and menu items: Workflow, Visualize, Données partagées, Aide, Utilisateur, and a grid icon. A status indicator on the right shows 'Using 0%'. Below the navigation bar, there are two search bars: 'search histories' and 'search all datasets'. A 'Create new' button is located to the right of the search bars. The main content area is titled 'Cet Historique' and features a dropdown menu (highlighted with a red box) for switching between histories. To the right of this dropdown is a 'Switch to' button, also highlighted with a red box. The interface shows two history panels. The left panel, titled 'ITS', contains a list of datasets with details like '42 shown, 10 deleted' and '276.74 MB'. The right panel, titled '16S', shows '2 shown' and '34.86 MB'. A context menu (highlighted with a red box) is open over the right panel, offering 'Copy', 'Delete', and 'Purge' actions. Below the search bars, there are instructions to 'Drag datasets here to copy them to the current history'. The dataset list includes items like '52: FROGSTAT Deseq2 Visualisation: report.nb.html', '49: FROGSTAT DESeq2 Preprocess: dds.Rdata', and '48: FROGSTAT Phyloseq Multivariate Analysis Of Va'.

Switch current history



1. When you are several histories, switch to the history named « 16S » as current history.
2. 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		-	
 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

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

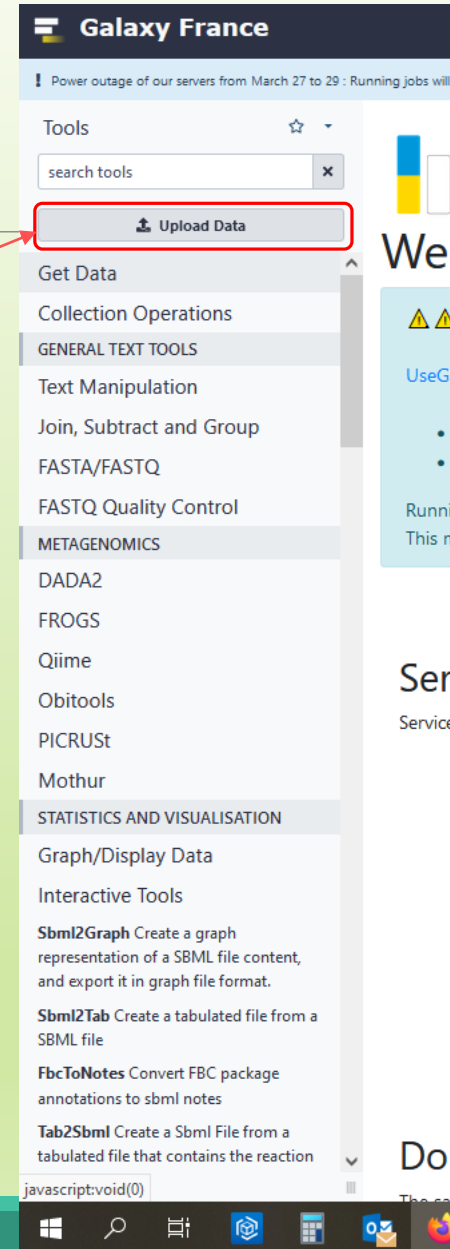


- By URL



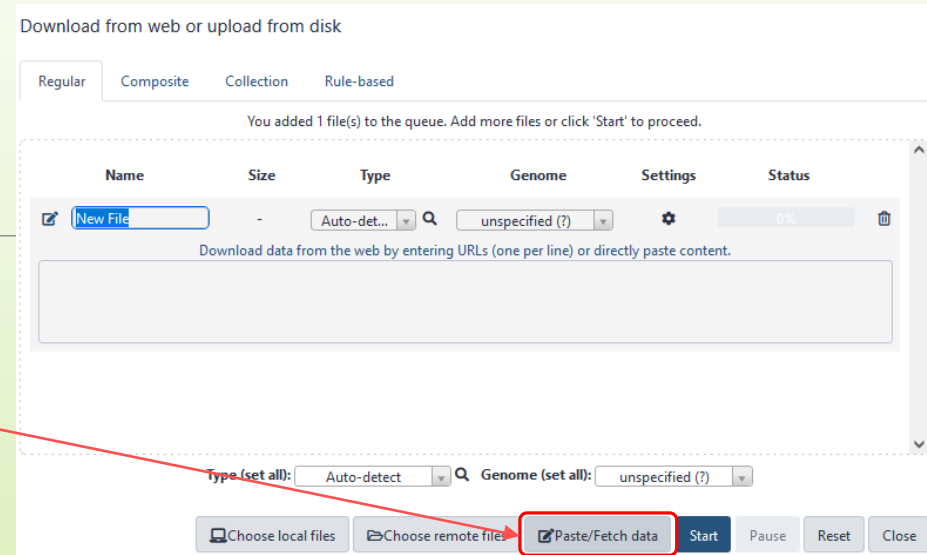
16S history creation

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

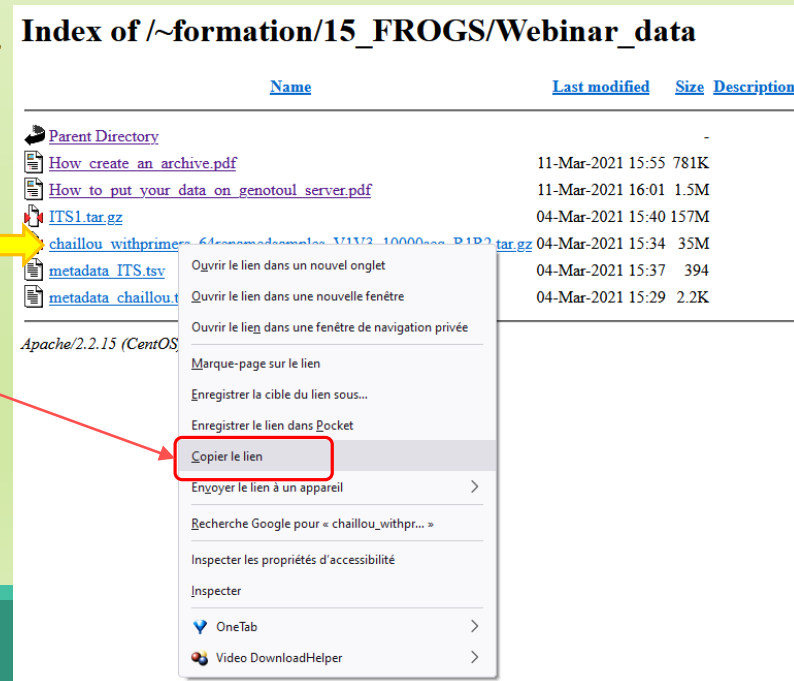
3. Click on **Paste/Fetch Data**




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

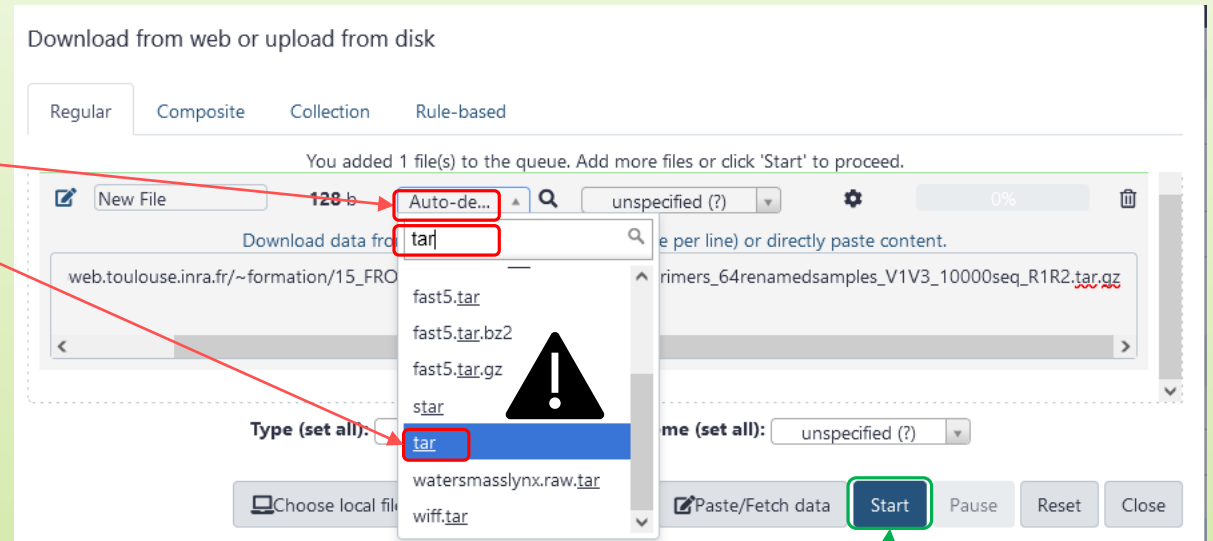


16S history creation

6.  Change the data type!

7. Click on Start

8. 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	394
metadata_chaillou.tsv	04-Mar-2021 15:29	2.2K

Remember to
change your
data type to
TSV

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

History + ⇌ ▾ 1

You have 14 histories.

- Show Histories Side-by-Side
- Copy this History
- Delete this History
- Reprendre les processus en pause
- Extraire un Workflow
- Montrer la structure
- Partager et publier** 2
- Set Permissions
- Make Private
- Exporter les citations des outils
- Exporter l'Historique dans un fichier
- Return to legacy panel

Share a history

The screenshot shows a software interface with a dark navigation bar at the top containing icons for home, workflow, visualize, shared data, help, user, and a grid. A green status bar on the right indicates 'Using 32%'. Below the navigation bar, the main content area is titled 'Share or Publish History `Jeu test pour tuto 16S`'. It features a toggle switch for 'Make History accessible' which is currently turned off. A text block explains that access is restricted to the user and those listed below, and that sharing a history grants access to all its datasets. Under the heading 'Share History with Individual Users', a note states that listed users will be able to view, import, and run the history. An 'Email' section lists three users: vincent.darbot@inrae.fr, maria.bernard@inrae.fr, and olivier.rue@inrae.fr, each with a 'Remove' button. A 'Share with a user' button is highlighted with a red arrow. On the right side, a 'History' panel displays a search bar, the title 'Jeu test pour tuto 16S', and statistics: '47 shown, 13 deleted' and '281.06 MB'. A list of history items follows, each with an eye icon, a pencil icon, and a close icon. Items 59 and 56 are marked with a red 'X' icon, indicating they are deleted. Item 60 is highlighted in green. The items are: 60: FROGSSTAT DESeq2 Preprocess: dds.Rdata; 59: FROGSSTAT DESeq2 Preprocess: dds.Rdata; 56: FROGSSTAT DESeq2 Preprocess: dds.Rdata; 55: FROGSSTAT Phyloseq Multivariate Analysis Of Variance: manova.nb.html; 54: FROGSSTAT Phyloseq Sample Clustering: clustering.nb.html.

Practice: share a history

SHARE 16S HISTORY WITH TRAINERS

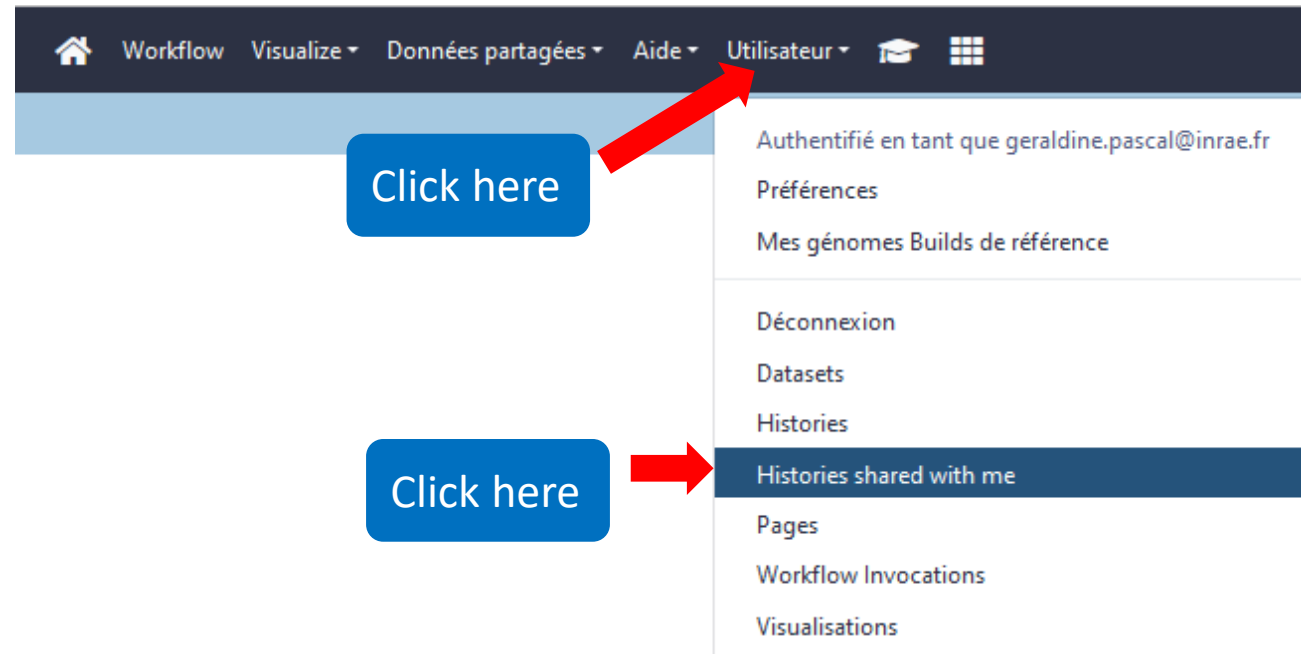
Share your data with us



- 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

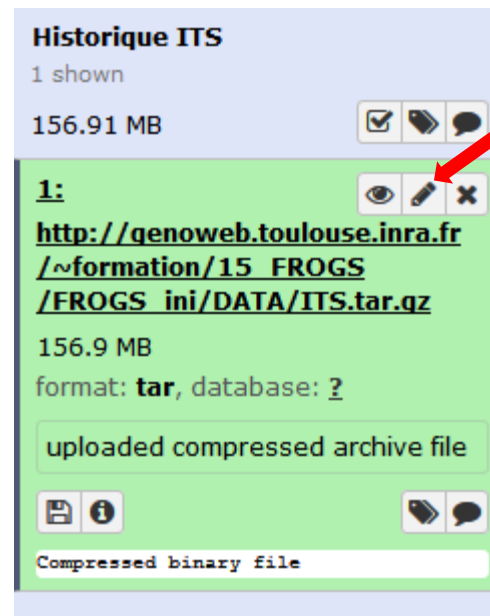
- Click on the name and click on « Copy » to import an history into your Galaxy account.
- To unshare click on unshare

The person who shared the history.

How manipulate datasets

To rename a dataset

- Switch to ITS history



Historique ITS
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

Click here to display attributes and change the name.

To rename a dataset

Change the name here

Edit dataset attributes

≡ Attributs

⚙ Convert

📄 Datatypes

👤 Permissions

Editer les attributs

Convertir en un nouveau
format de données

↺ Auto-detect

💾 Save

Name

FROGS Pre-process: count.tsv

Info

Application

Software: preprocess.py (version: 3.2.3)

Command: /galaxydata/galaxy2021/galaxy01/galaxy/database/dependencies/_conda/envs/mulled-

v1.81e0e76624e2492f20000d51de285ef751406605e2dede0f71ba6e0b0129202/bin/preprocess.py/illumina_output_derep

Annotation

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

Database/Build

----- Additional Species Are Below -----

Number of comment lines

To change the datatype

Edit dataset attributes

Attributes

Convert

Datatypes

Permissions

Changer le format de données

Detect datatype

Changer le format de données










New Type

tsv

This will change the datatype of the existing dataset but not modify its content. Use this if Galaxy has incorrectly guessed the type of your dataset.

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

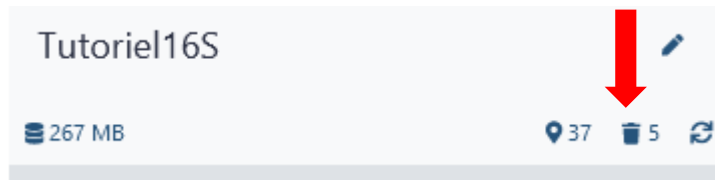
To delete a dataset

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

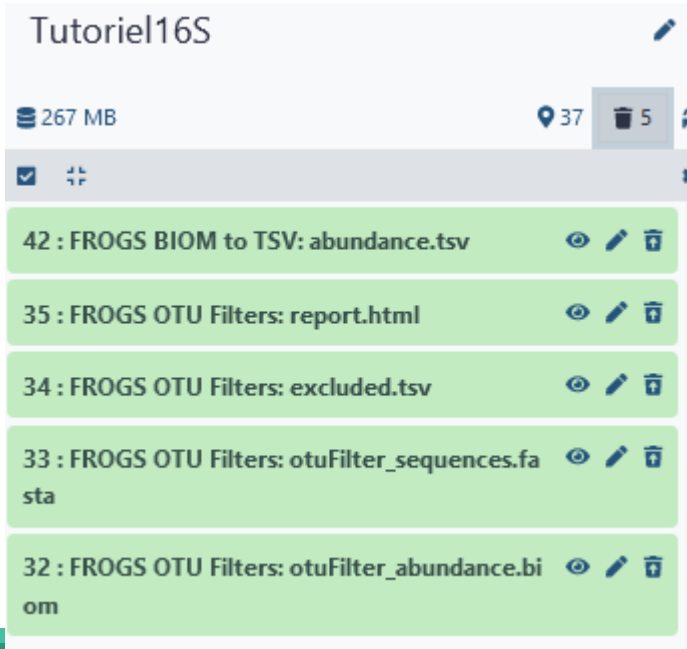


Click here to delete a dataset.

To see deleted dataset

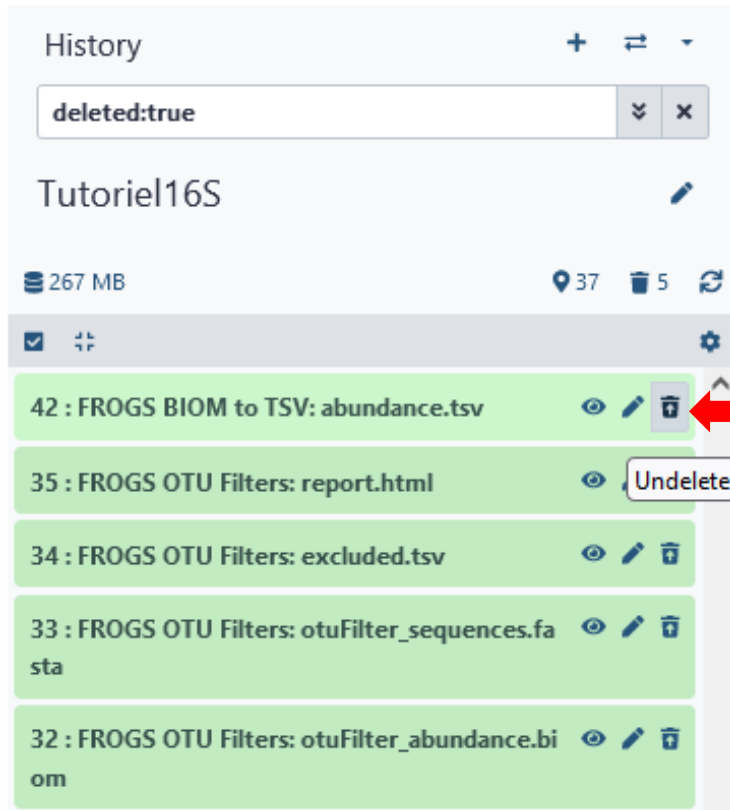


Click here to see deleted dataset.



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

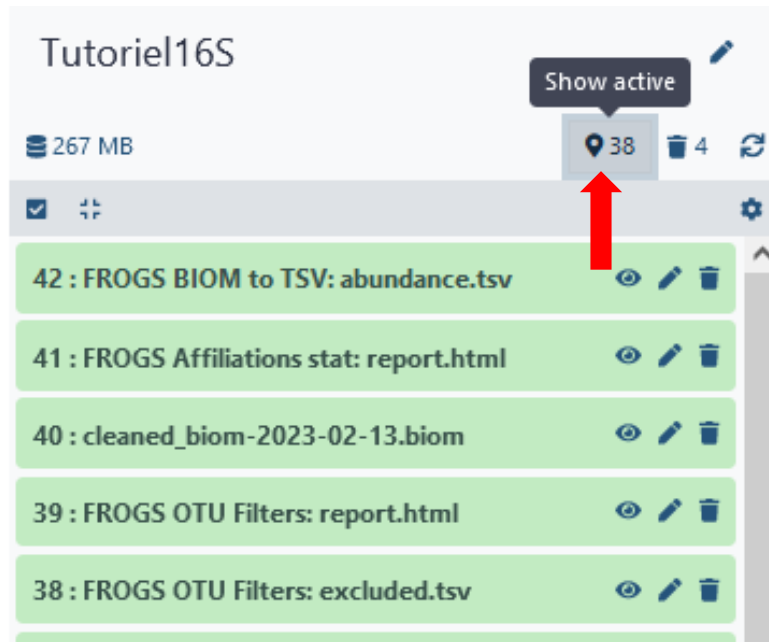
33 : FROGS OTU Filters: otuFilter_sequences.fasta

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 3 April!

YOUR TRAINERS:
GÉRALDINE & LUCAS

