FROGS
Find Rapdly OTU with Galaxy Solution

A - Training on Galaxy: Metabarcoding

May/June 2024 - webinar

GALAXY Practice

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

INRAE Genotoul Bioinfo SIGENAE GenPhySE MaIAGE GABI migale IAM INTERACTIONS ARBRES-MICROORGANISMES

RÉPUBLIQUE FRANÇAISE
Liberté
Égalité
Fraternité

Objectives of the 1st part

- Learn the basics of Galaxy

- Prepare the 1 dataset for FROGS formation

What is Galaxy?

Galaxy software framework

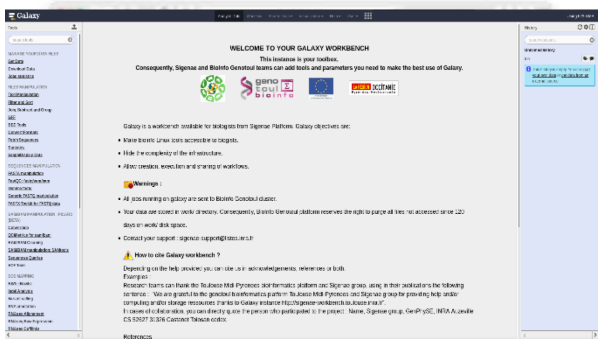
- Galaxy is an overlay that allows researchers without computer expertise to perform computational analyses online, automate them and share the results easily.
- Developed by the Galaxy Team Project.
- Relies on a computational infrastructure (Server).
- Open source

Galaxy software framework

User

You

Web browser



WELCOME TO YOUR GALAXY WORKBENCH

This instance is your toolbox.

Consequently, Sigenae and BioInfo Genotoul teams can add tools and parameters you need to make the best use of Galaxy.

Galaxy is a workbench available for biologists from Sigenae Platform. Galaxy objectives are:

- Make bioinfo Linux tools accessible to biogists.
- Hide the complexity of the infrastructure.
- Allow creation, execution and sharing of workflows.

Warnings :

- All jobs running on galaxy are sent to BioInfo Genotoul cluster.
- Your data are stored in work/ directory. Consequently, BioInfo Genotoul platform reserves the right to purge all files not accessed since 120 days on work/ disk space.
- Contact your support : sigenae-support@listes.inra.fr

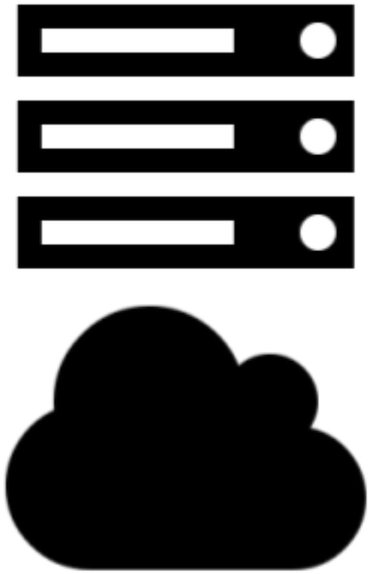
How to cite Galaxy workbench ?

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

What you see

Servers



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.

- Location : **Toulouse**
- Compute (#CPU HT*) : **6224**
- Storage (#TB) : **4400**
- RAM (#GB) : **36500**
- RAM/core (#GB) : **5.56**
- GPU (#Card) : **1**

- Location : **Jouy en Josas**
- Compute (#CPU HT*) : **1016**
- Storage (#TB) : **350**
- RAM (#GB) : **7000**
- RAM/core (#GB) : **6.89**
- GPU (#Card) : -

- Location : **Orsay (IDRIS)**
- Compute (#CPU HT*) : **4300**
- Storage (#TB) : **2400**
- RAM (#GB) : **20008**
- RAM/core (#GB) : **4.65**
- GPU (#Card) : -

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 Plateform: 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

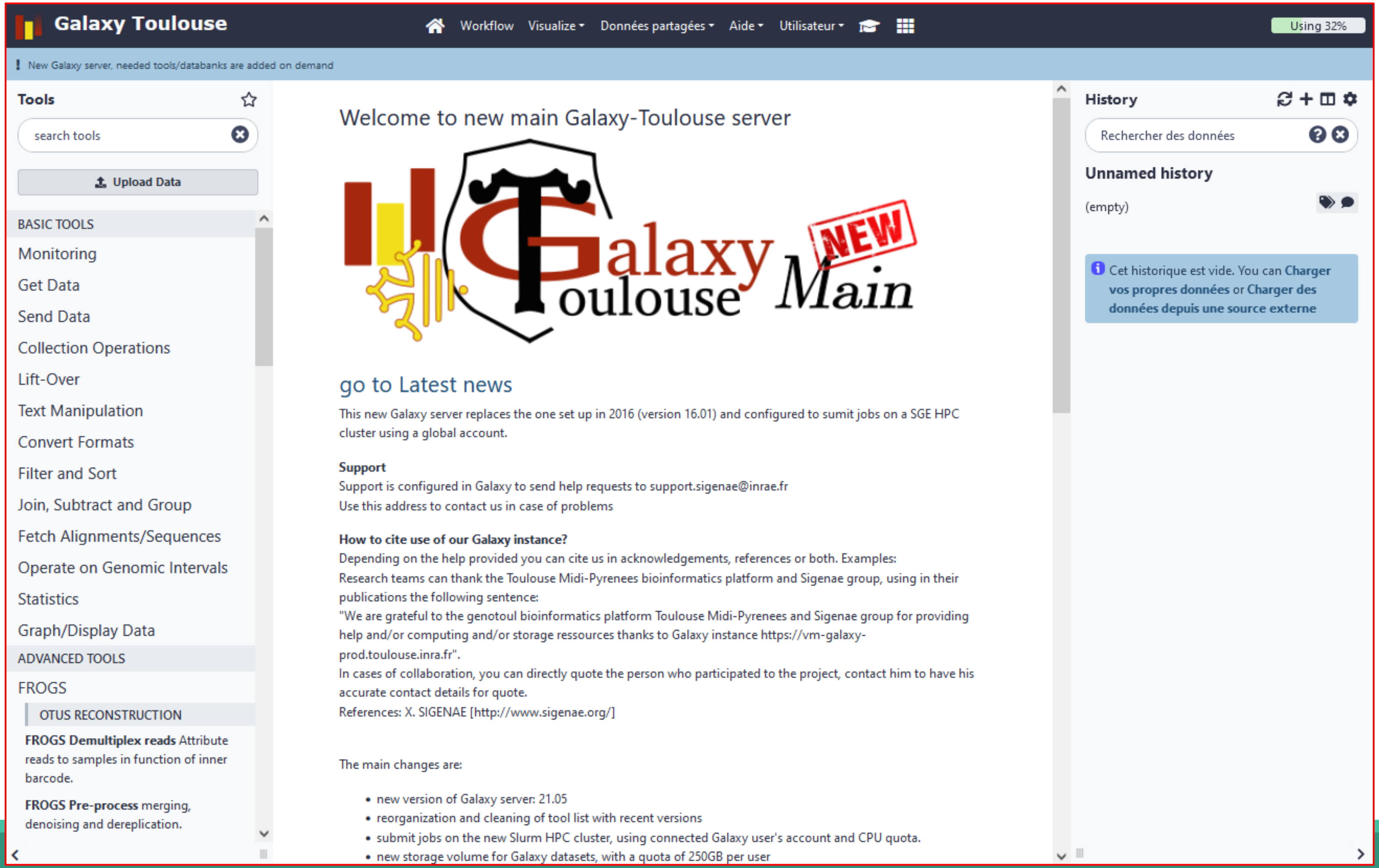
Welcome to the Migale Galaxy instance!

Global load of the nodes dedicated to Galaxy jobs

What's new

- **28th January 2021:** Added Parnsnp in Section *Phylogenyy*.
- **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 staramr 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



! New Galaxy server, needed tools/databanks are added on demand

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

FROGS Pre-process merging, denoising and dereplication.

Cet historique est vide. You can **Charger vos propres données** or **Charger des données depuis une source externe**

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 sumit jobs on a SGE HPC cluster using a global account.

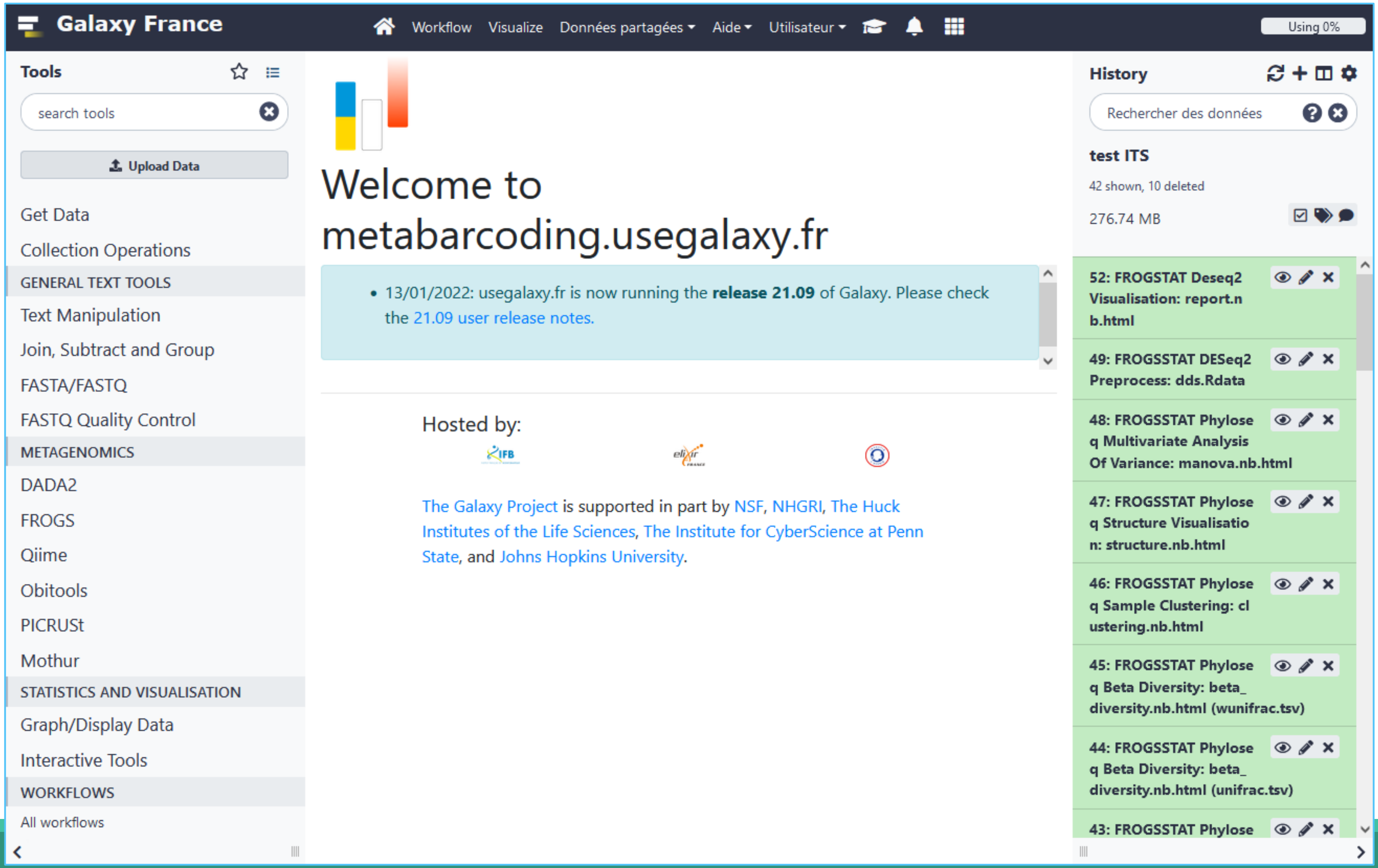
Support
Support is configured in Galaxy to send help requests to support.sigenae@inrae.fr
Use this address to contact us in case of problems

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 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 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.sigenae.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

Exemple of Galaxy platforms



Welcome to metabarcoding.usegalaxy.fr

- 13/01/2022: usegalaxy.fr is now running the **release 21.09** of Galaxy. Please check the 21.09 user release notes.

Hosted by:

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.

test ITS

42 shown, 10 deleted

276.74 MB

52: FROGSTAT Deseq2 Visualisation: report.nb.html

49: FROGSSTAT DESeq2 Preprocess: dds.Rdata

48: FROGSSTAT Phyloseq Multivariate Analysis Of Variance: manova.nb.html

47: FROGSSTAT Phyloseq Structure Visualisation: structure.nb.html

46: FROGSSTAT Phyloseq Sample Clustering: clustering.nb.html

45: FROGSSTAT Phyloseq Beta Diversity: beta_diversity.nb.html (wunifrac.tsv)

44: FROGSSTAT Phyloseq Beta Diversity: beta_diversity.nb.html (unifrac.tsv)

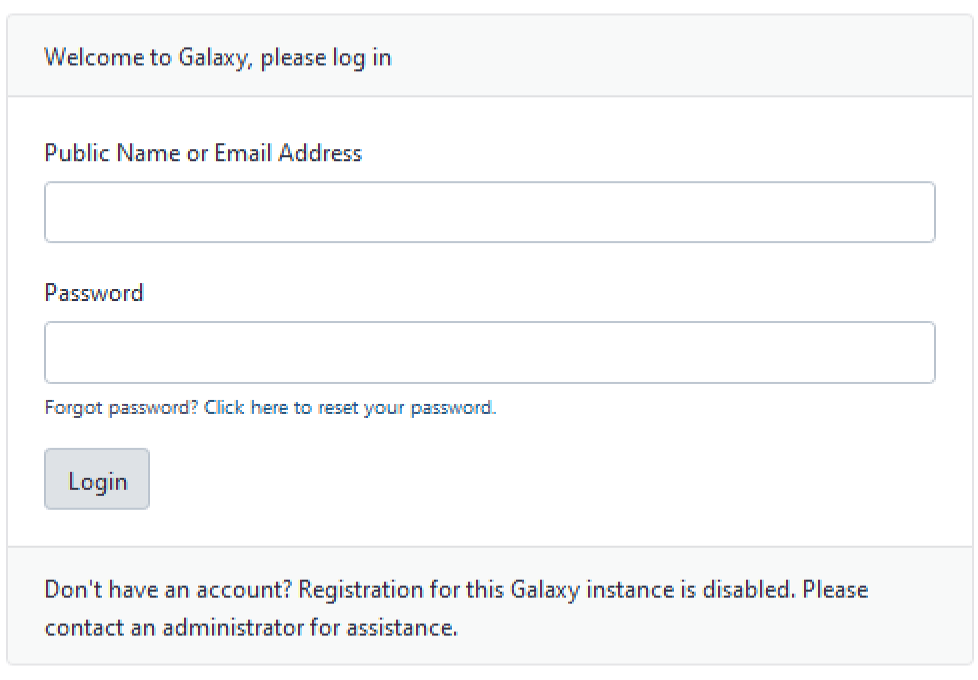
43: FROGSSTAT Phylose

Practice:

CONNECT TO GALAXY WORKBENCH

Exercise

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

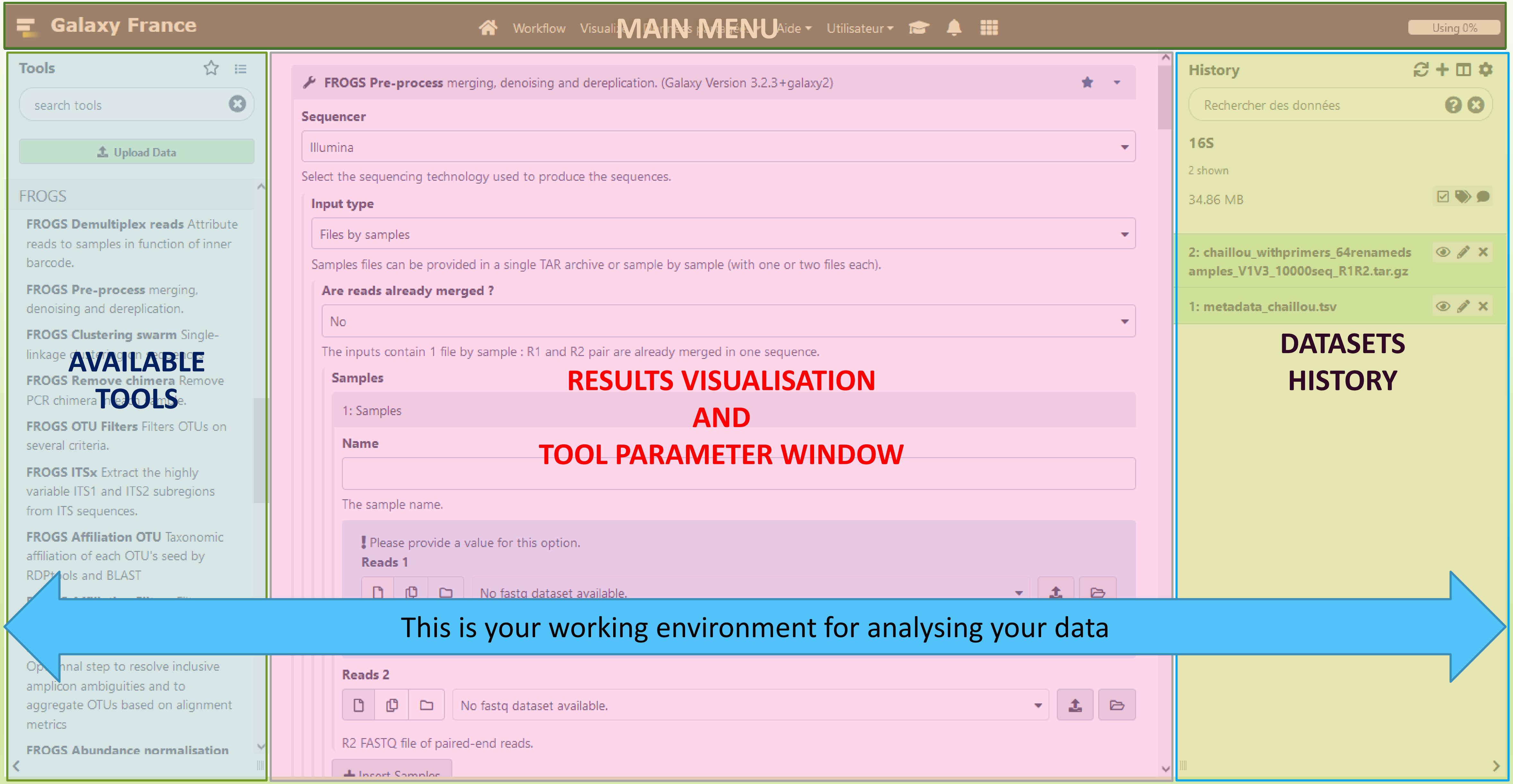


Welcome to Galaxy, please log in

Public Name or Email Address

Password

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



MAIN MENU

AVAILABLE TOOLS

RESULTS VISUALISATION

AND

TOOL PARAMETER WINDOW

DATASETS HISTORY

This is your working environment for analysing your 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 on sequences.

FROGS Remove chimera Remove PCR chimera in each sample.

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 RDPtools 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

16S

2 shown

34.86 MB

2: chaillou_withprimers_64renamedsamples_V1V3_10000seq_R1R2.tar.gz

1: metadata_chaillou.tsv

Vocabulary of Galaxy

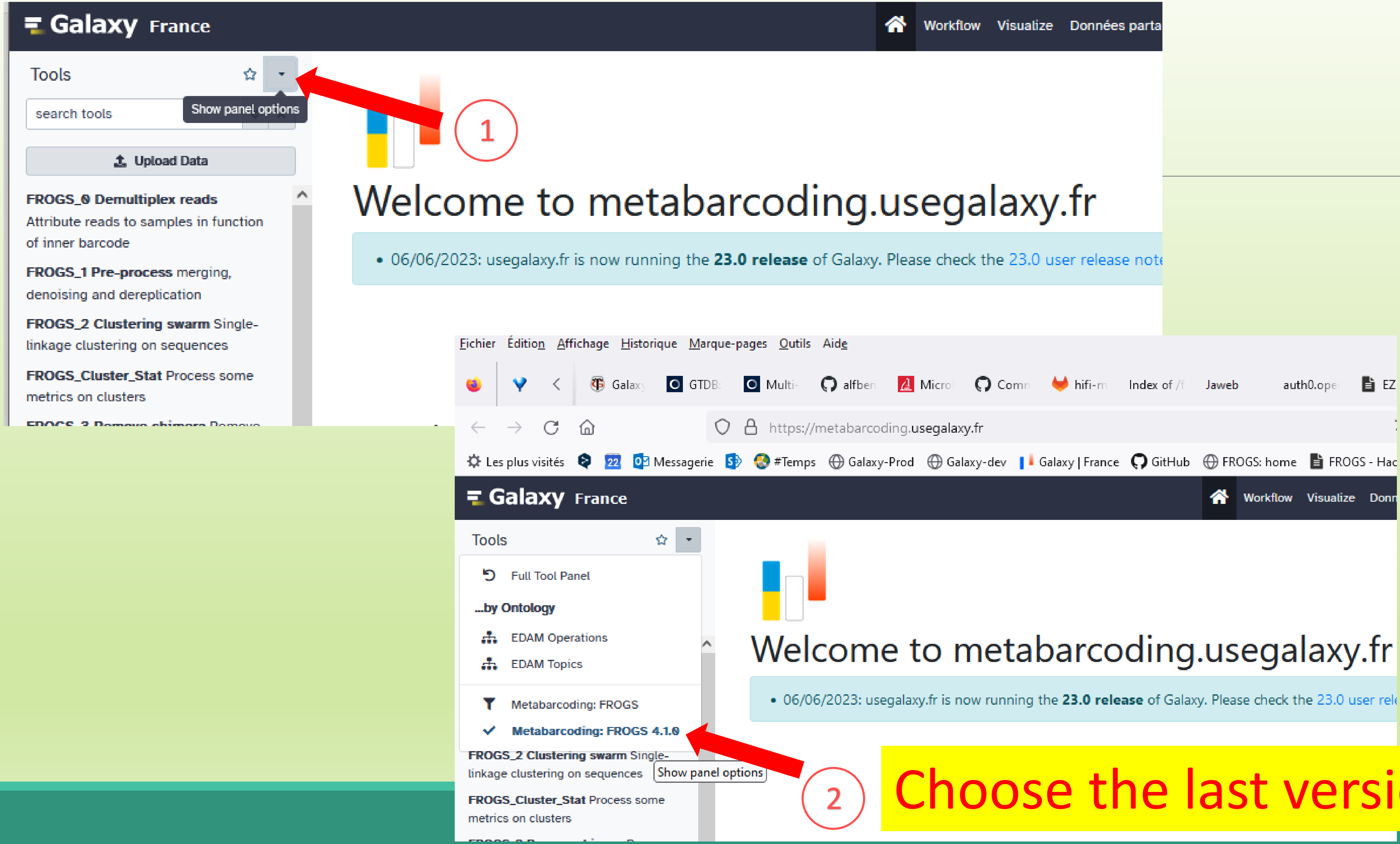
- **Tools:**
  - A tool has a function which is explained when you click on it.
  - Each Galaxy platform has its own tools.
- **Datasets:**
  - A dataset is a file, uploaded to Galaxy by you or produced by a tool.
  - Be careful: a dataset has a datatype.
- **Histories:**
  - A tool generates datasets 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.



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

Choose the last version 4.1.

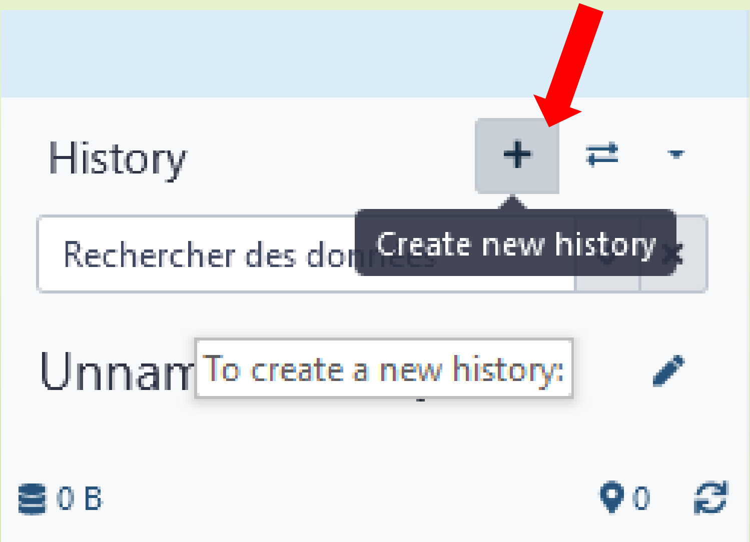
Manipulate Histories

Practice:

CREATE THE NECESSARY HISTORY FOR FROGS TRAINING

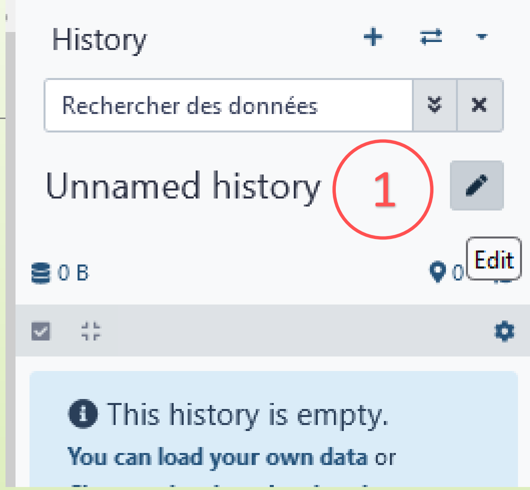
To create a new history:

- Click on the plus.

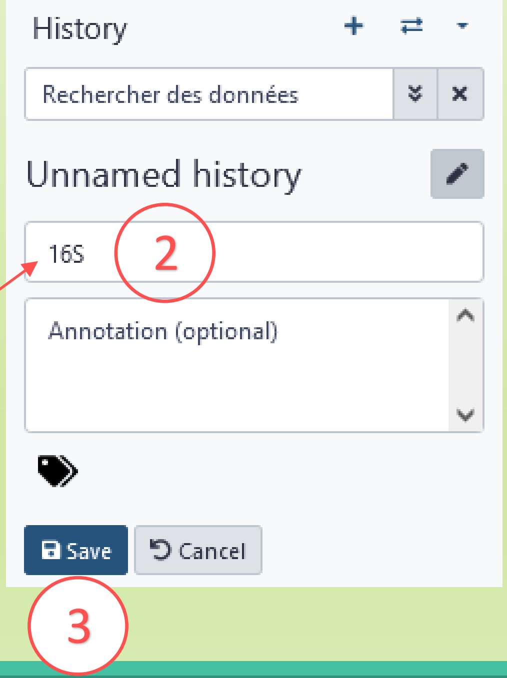


To rename a history:

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



This history is empty.
You can load your own data or



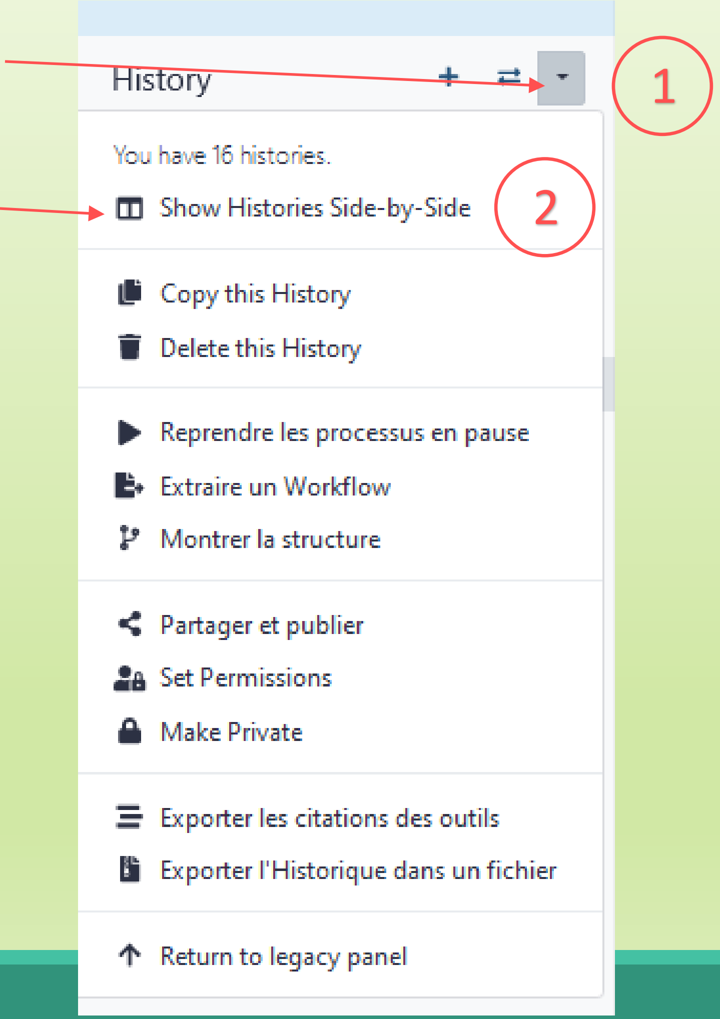
⚠ Don’t use special characters or accents!

Exercise

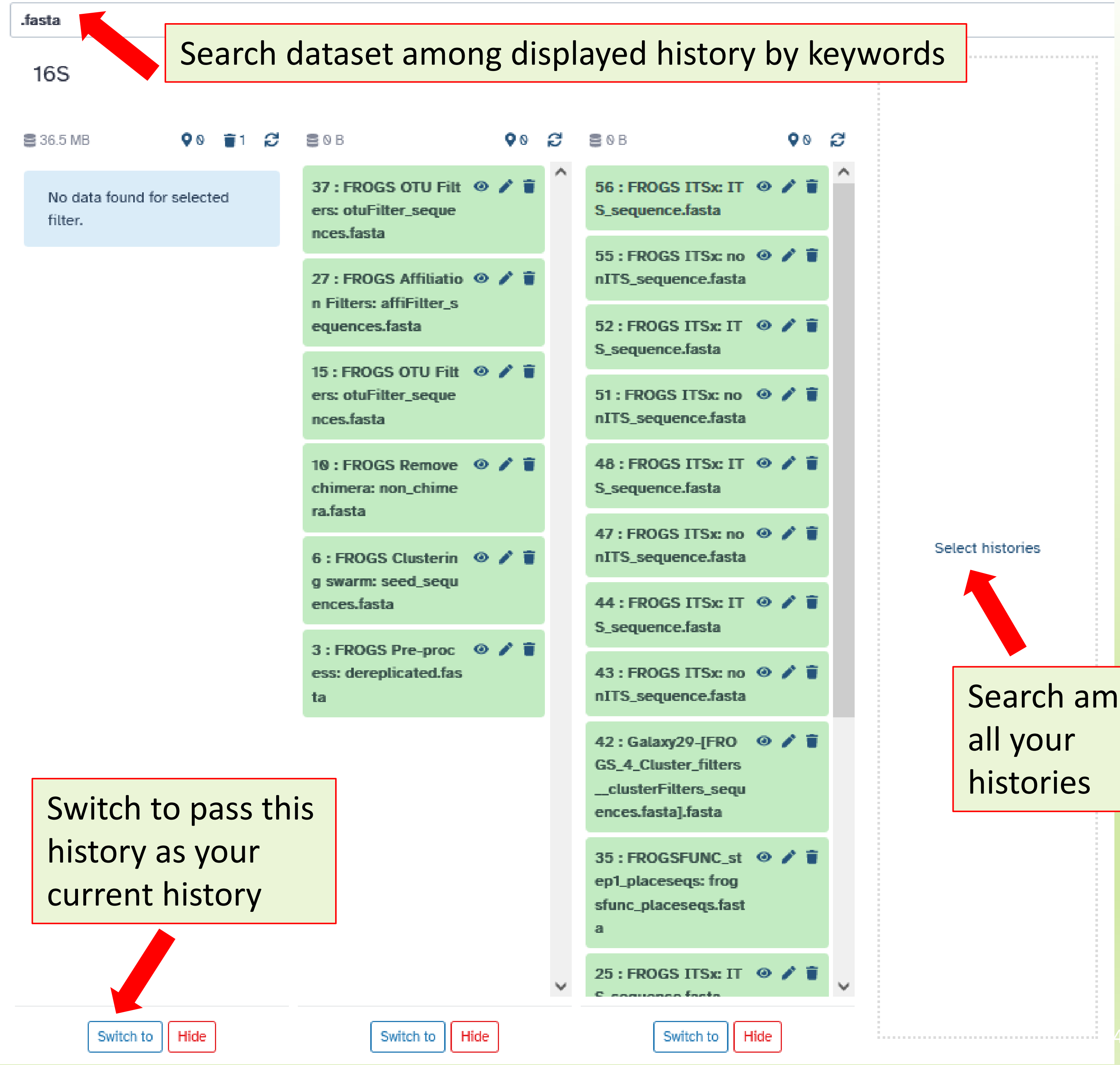
- Create another history as before with 16S. This new history will be named **ITS**

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 0 1

No data found for selected filter.

0 B 0

- 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

0 B 0

- 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: frogsfunc_placeseqs.fasta
- 25 : FROGS ITSx: IT

Select histories

Search dataset among displayed history by keywords

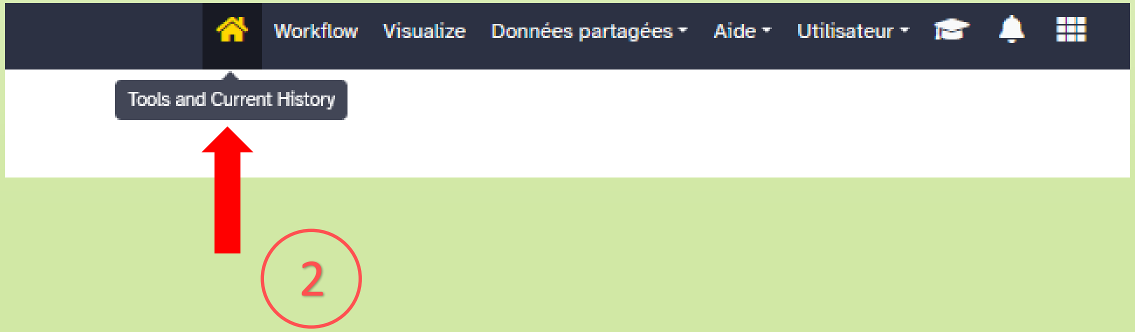
Switch to pass this history as your current history

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.



This history is empty.
You can load your own data or Charger des données depuis une source externe.

Data import

How to import your data to Galaxy ?

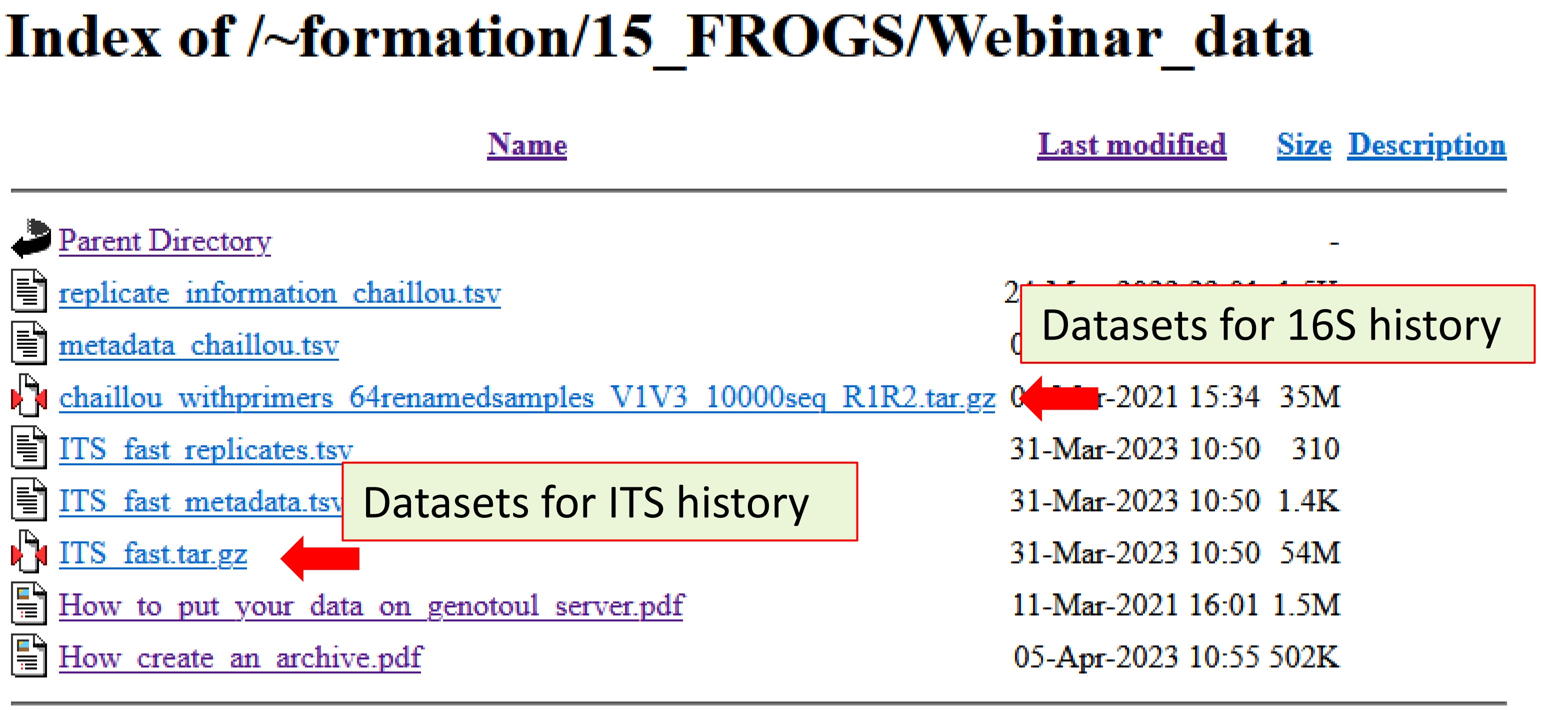
- Several ways to upload your data to Galaxy:
  - From your computer
  - By URL
  - Shared by other users of Galaxy

drag-and-drop files

Paste an url

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		-	
replicate_information_chaillou.tsv	2		
metadata_chaillou.tsv	0		
chaillou_withprimers_64renamedsamples_V1V3_10000seq_R1R2.tar.gz	0 r-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

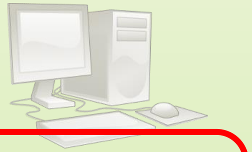
Create an archive is
always possible !
It is our recommendation.

Practice: 16S history

UPLOAD FILE FROM AN URL

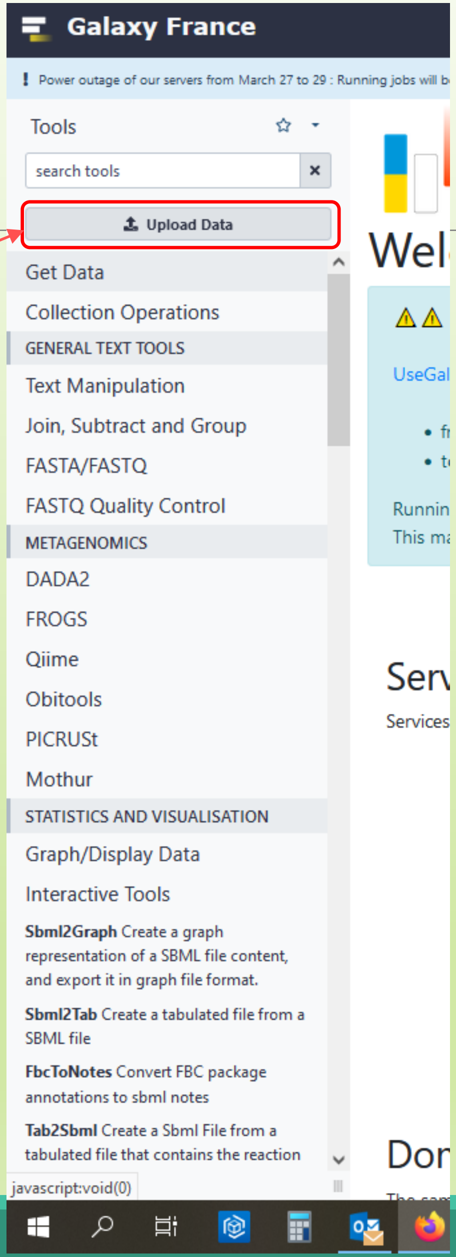
How to import your data to Galaxy ?

- 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



Sbml2Graph Create a graph representation of a SBML file content, and export it in graph file format.

Sbml2Tab Create a tabulated file from a SBML file

FbcToNotes Convert FBC package annotations to sbml notes

Tab2Sbml Create a Sbml File from a tabulated file that contains the reaction

16S history creation

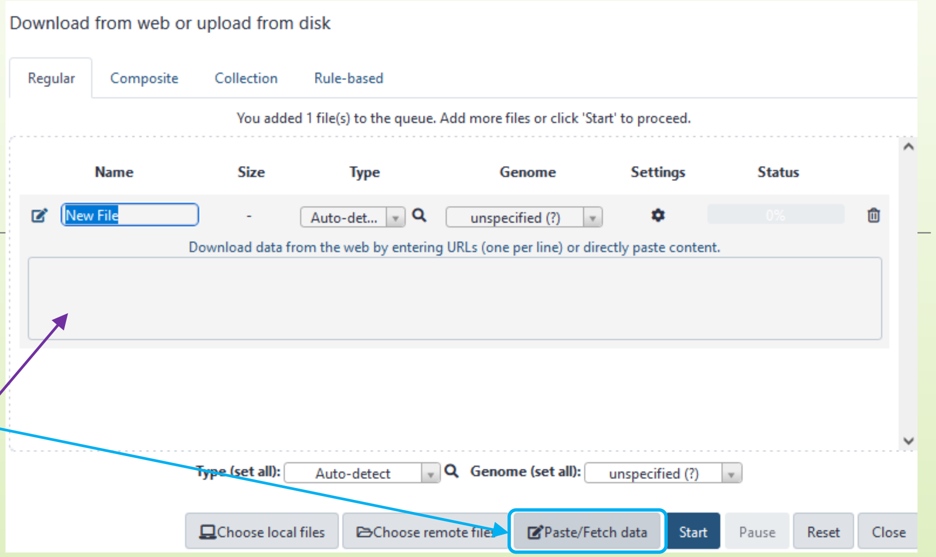
3. Click on **Paste/Fetch Data**

4. Collect your data address on http://genoweb.toulouse.inra.fr/~formation/15_FROGS/Webinar_data/

   Click right on the name file

5. Copy the address of the link

6. Paste the link in the grey window



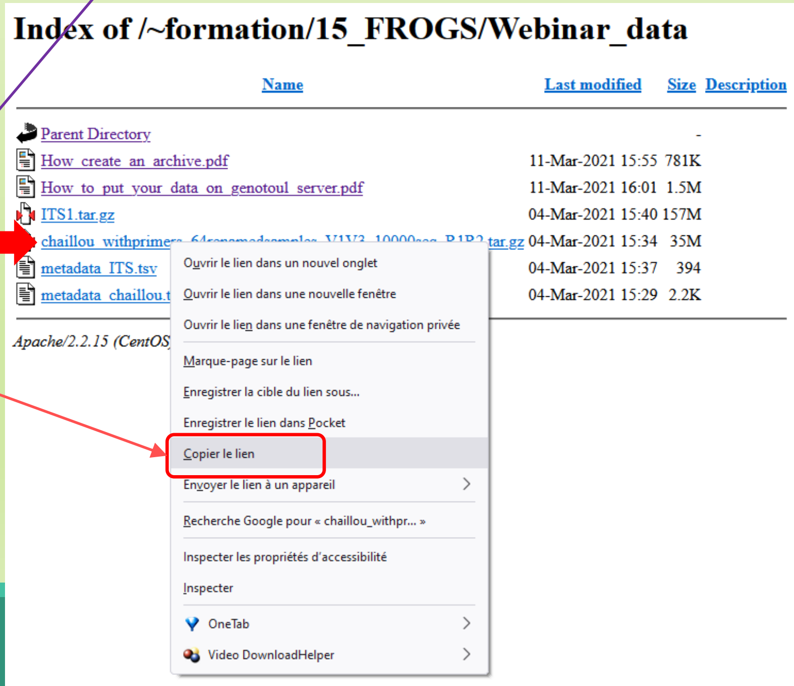
Download from web or upload from disk

You added 1 file(s) to the queue. Add more files or click 'Start' to proceed.

Name	Size	Type	Genome	Settings	Status
New File	-	Auto-det...	unspecified (?)		0%

Download data from the web by entering URLs (one per line) or directly paste content.

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



Index of /~formation/15_FROGS/Webinar_data

Name	Last modified	Size	Description
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.t	04-Mar-2021 15:29	2.2K	

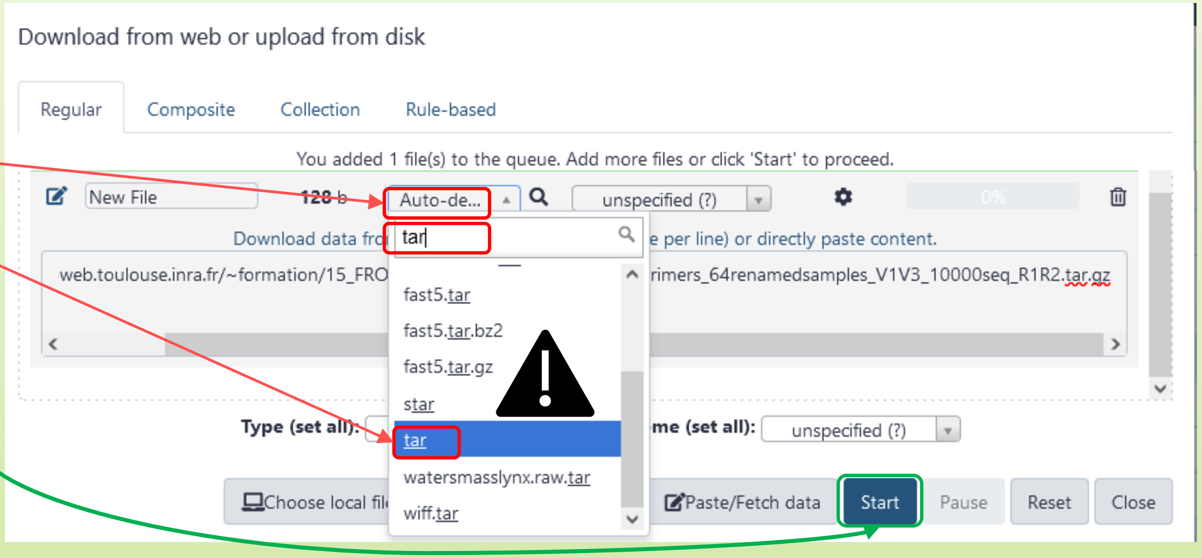
Apache/2.2.15 (CentOS

16S history creation

7. ⚠ Change the data type!

8. Click on Start

9. Start again with the file metadata_chaillou.tsv



Download from web or upload from disk

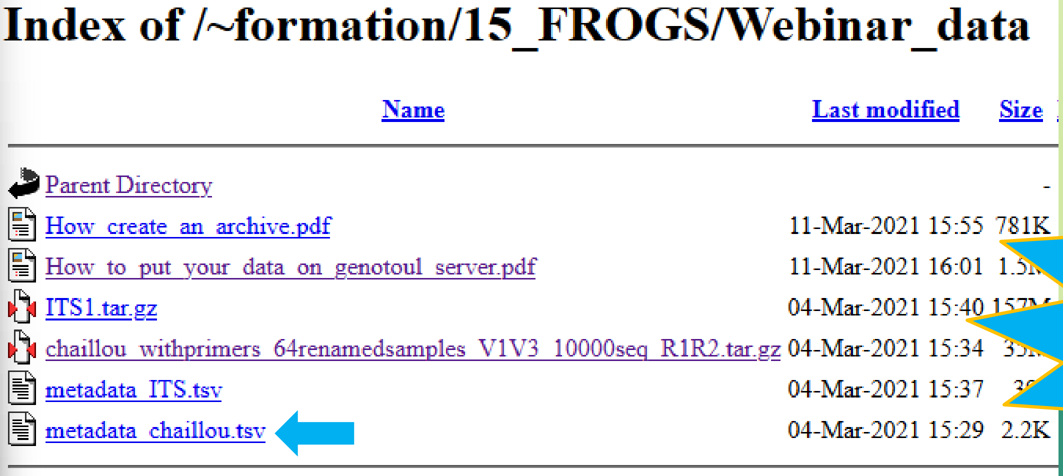
You added 1 file(s) to the queue. Add more files or click 'Start' to proceed.

New File 128 b Auto-de... unspecified (?) 0%

tar

Download data fro... e per line) or directly paste content.

web.toulouse.inra.fr/~formation/15_FRO... rimers_64renamedsamples_V1V3_10000seq_R1R2.tar.gz



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	1.5
ITS1.tar.gz	04-Mar-2021 15:40	157
chaillou_withprimers_64renamedsamples_V1V3_10000seq_R1R2.tar.gz	04-Mar-2021 15:34	35
metadata_ITS.tsv	04-Mar-2021 15:37	3
metadata_chaillou.tsv	04-Mar-2021 15:29	2.2K

Remember to change your data type to TSV

Share a History

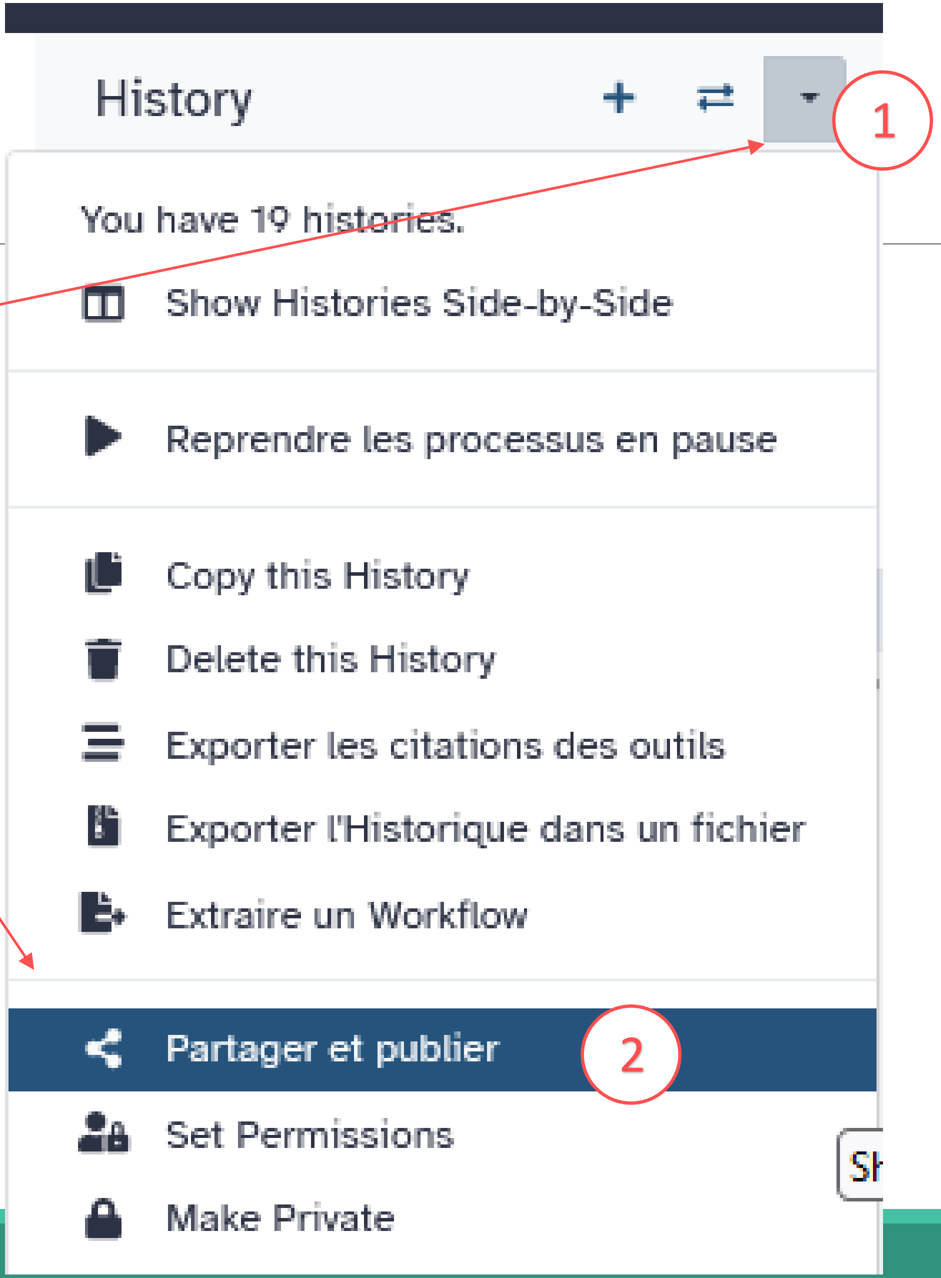
Share a history

- 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

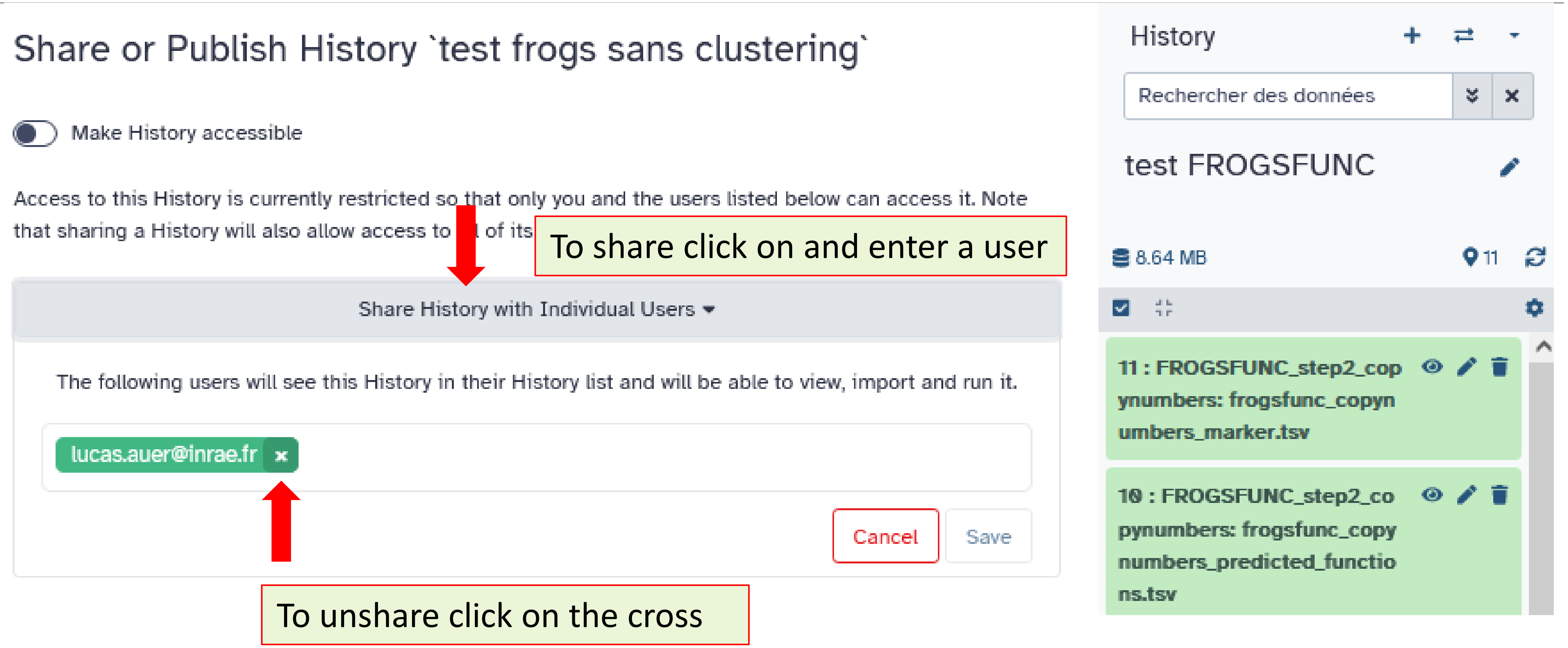
1. Click on this arrow
2. Click on share and publish



History

You have 19 histories.

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 l of its

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

11 : FROGSFUNC_step2_copynumbers: frogsfunc_copynumbers_marker.tsv

10 : FROGSFUNC_step2_copynumbers: frogsfunc_copynumbers_predicted_functions.tsv

To share click on and enter a user

To unshare click on the cross

Share a history

Share or Publish History `test frogs sans clustering`

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

Make History accessible

Make History publicly available in Published Histories

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 AND ITS HISTORIES 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

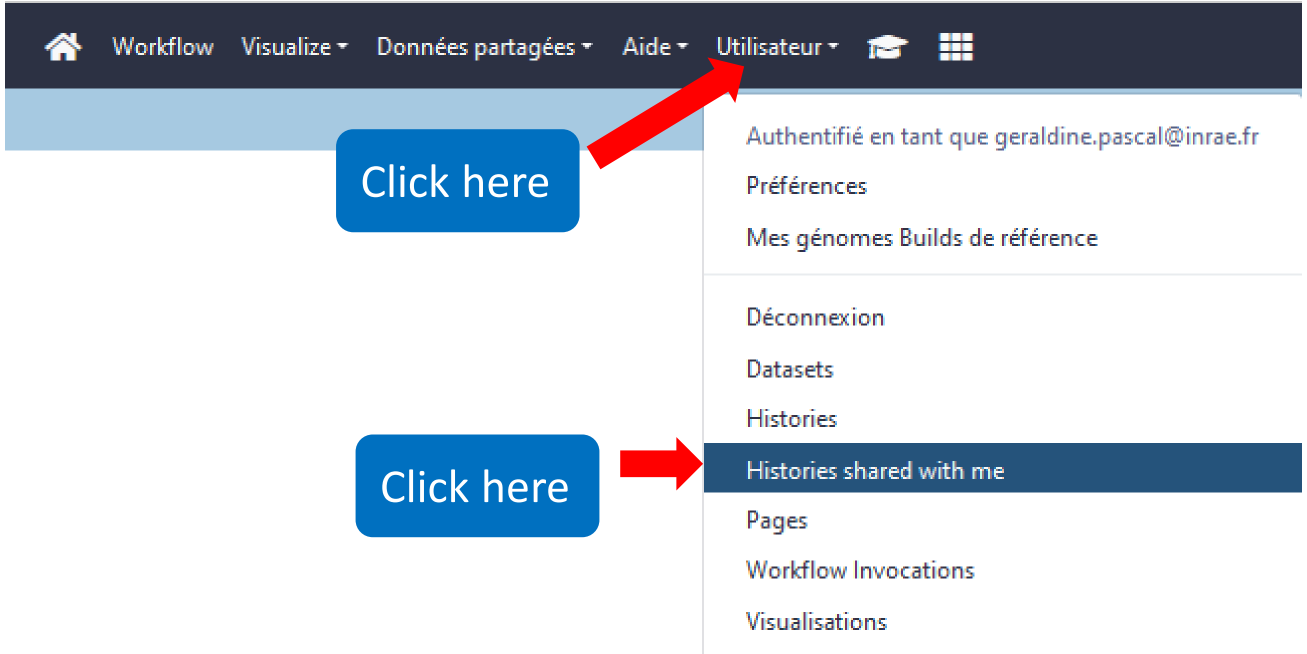
lucas.auer@inrae.fr ×

geraldine.pascal@inrae.fr ×

- Start again with the “ITS” history
- Contact us by email frogs-training@inrae.fr when it is done.

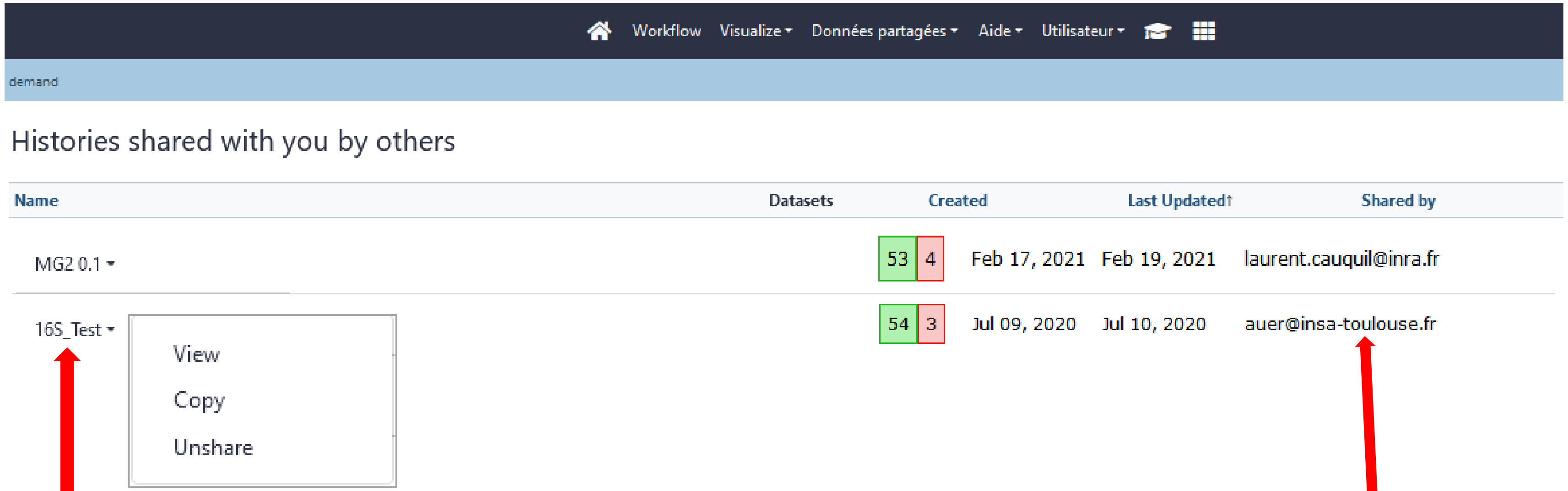
Other useful things

Import a shared history



Click here

Click here



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

- 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

1 shown

156.91 MB

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

156.9 MB

format: **tar**, database: ?

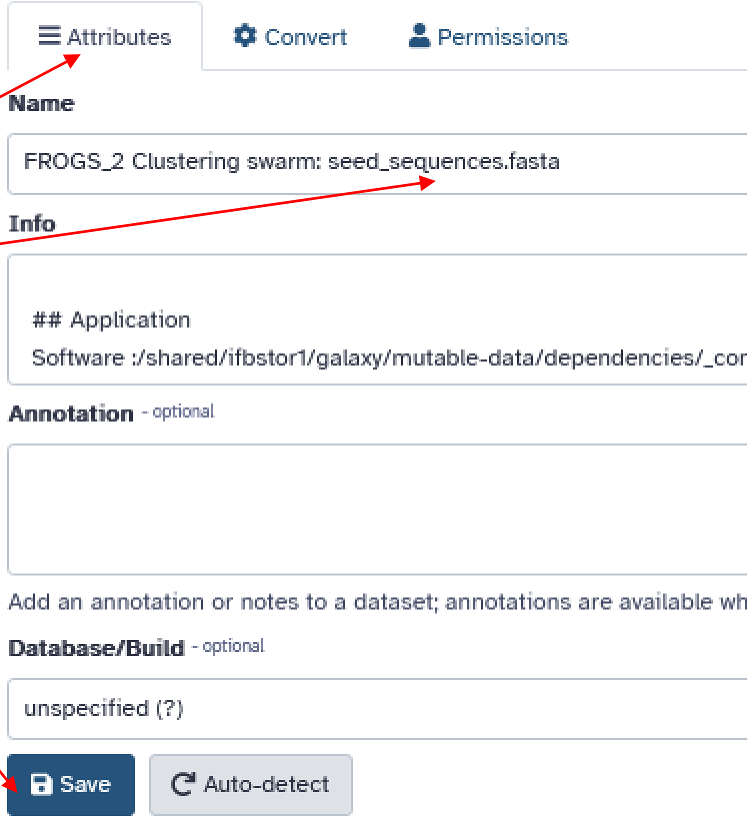
uploaded compressed archive file

```
Compressed binary file
```

Click here to display attributes and change the name.

To rename a dataset

Change the name here



Name

FROGS_2 Clustering swarm: seed_sequences.fasta

Info

```
## Application
Software :/shared/ifbstor1/galaxy/mutable-data/dependencies/_cor
```

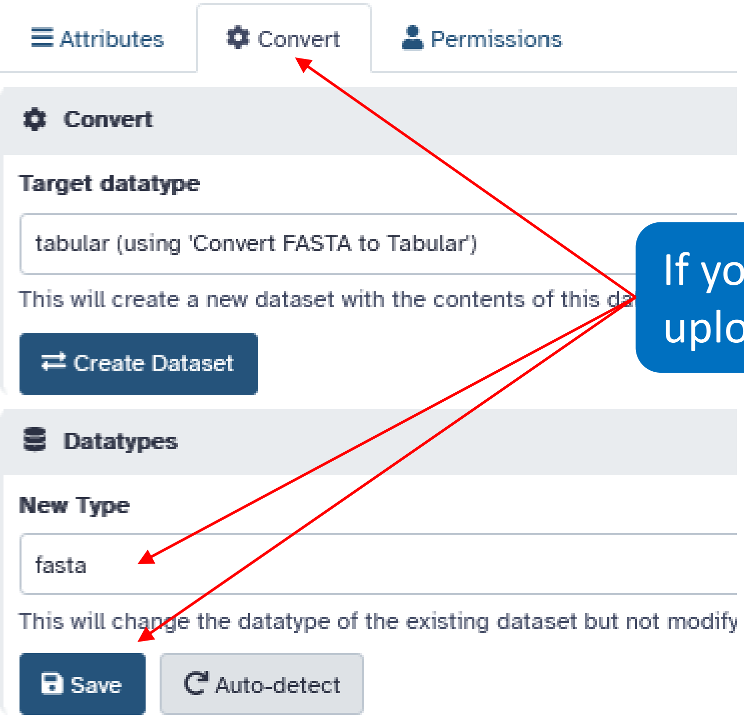
Annotation - optional

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

Database/Build - optional

unspecified (?)

To change the datatype



Convert

Target datatype

tabular (using 'Convert FASTA to Tabular')

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

Datatypes

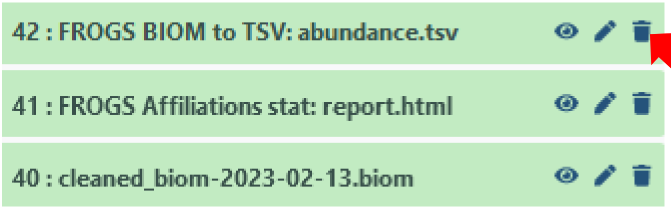
New Type

fasta

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

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

To delete a dataset



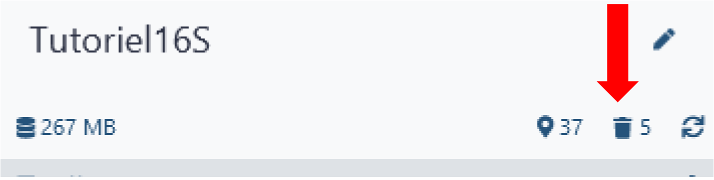
42 : FROGS BIOM to TSV: abundance.tsv

41 : FROGS Affiliations stat: report.html

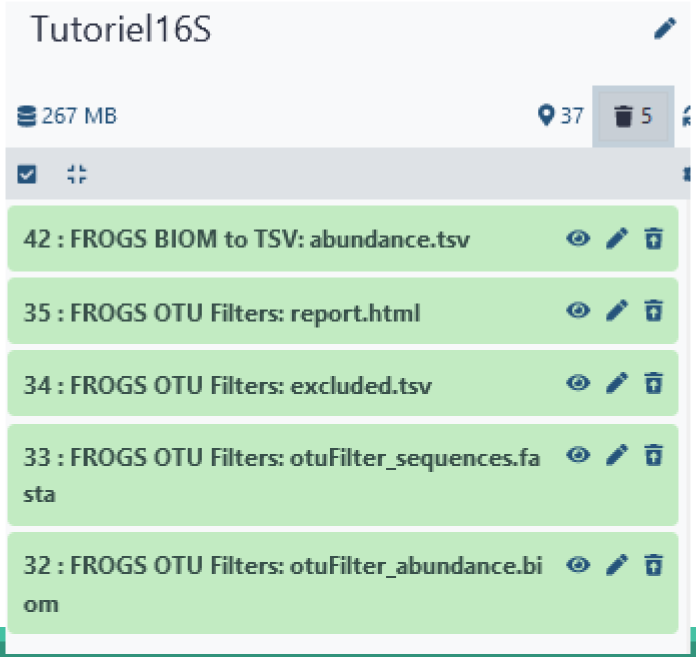
40 : cleaned_biom-2023-02-13.biom

Click here to delete a dataset.

To see deleted dataset



Click here to see deleted dataset.



42 : FROGS BIOM to TSV: abundance.tsv

35 : FROGS OTU Filters: report.html

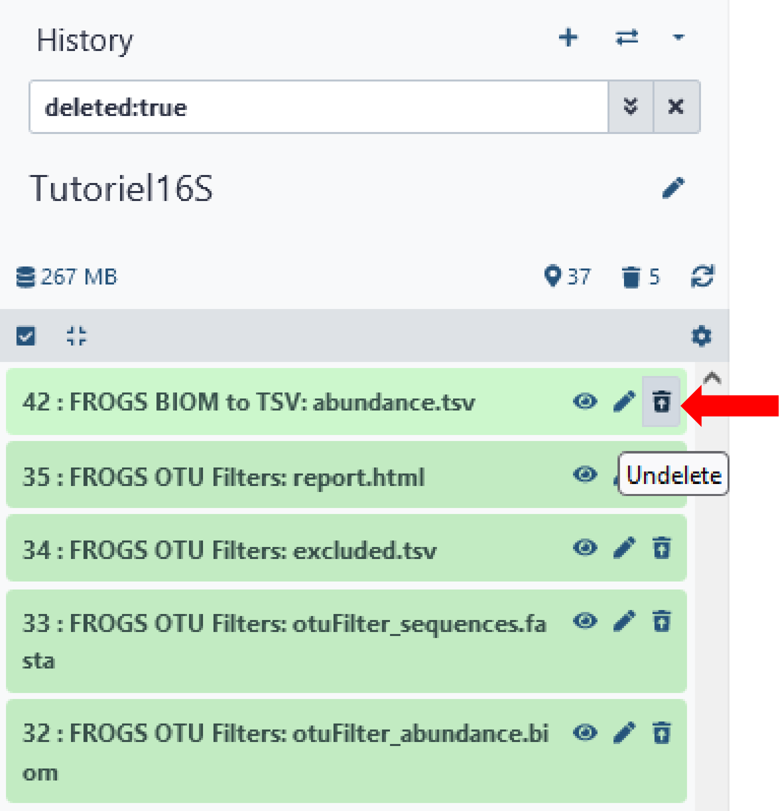
34 : FROGS OTU Filters: excluded.tsv

33 : FROGS OTU Filters: otuFilter_sequences.fasta

32 : FROGS OTU Filters: otuFilter_abundance.biom

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

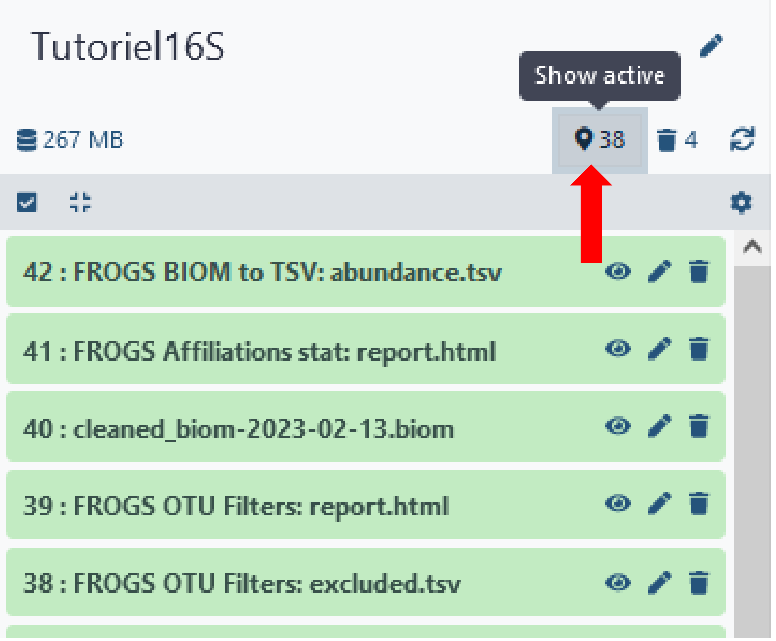
33 : FROGS OTU Filters: otuFilter_sequences.fasta

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



Tutoriel16S

267 MB

38

4

42 : FROGS BIOM to TSV: abundance.tsv

41 : FROGS Affiliations stat: report.html

40 : cleaned_biom-2023-02-13.biom

39 : FROGS OTU Filters: report.html

38 : FROGS OTU Filters: excluded.tsv

Click here to sea 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 27 May!

YOUR TRAINERS:

GÉRALDINE & LUCAS