A - Training on Galaxy: Metabarcoding

May 2022 - 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 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

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 biologists.
- 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

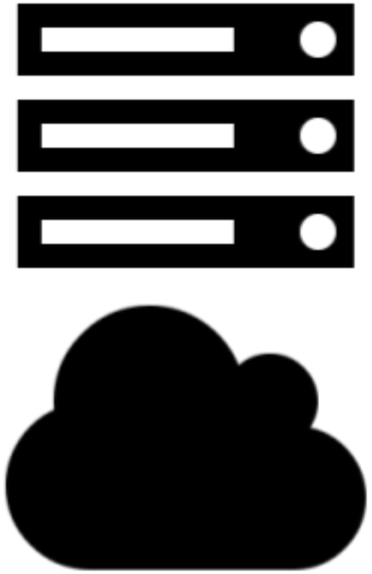
How to cite Galaxy workbench ?

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.

Genotoul Bioinfo

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

migale

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

IFB CORE

- 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

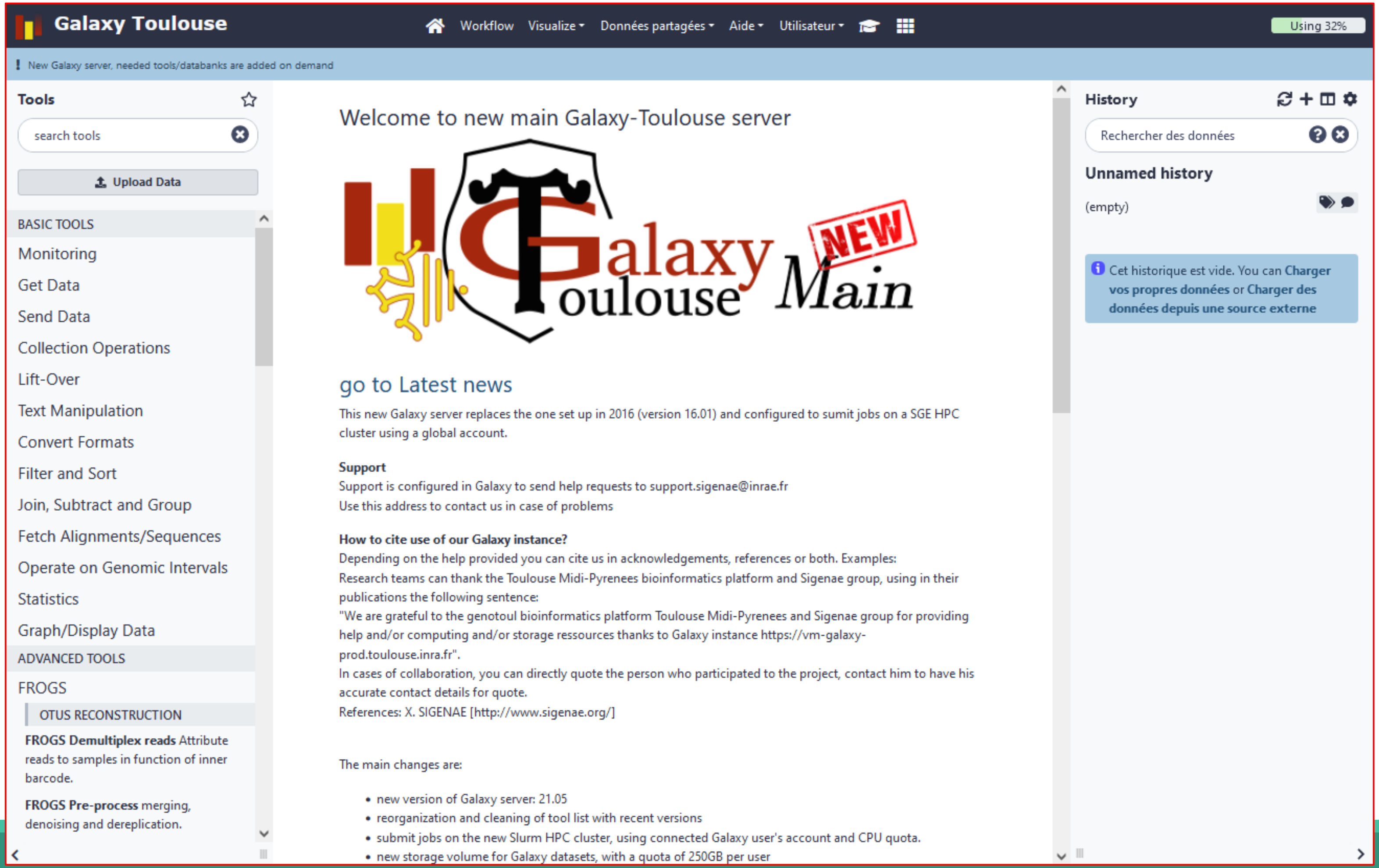
Welcome to the Migale Galaxy instance!

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.

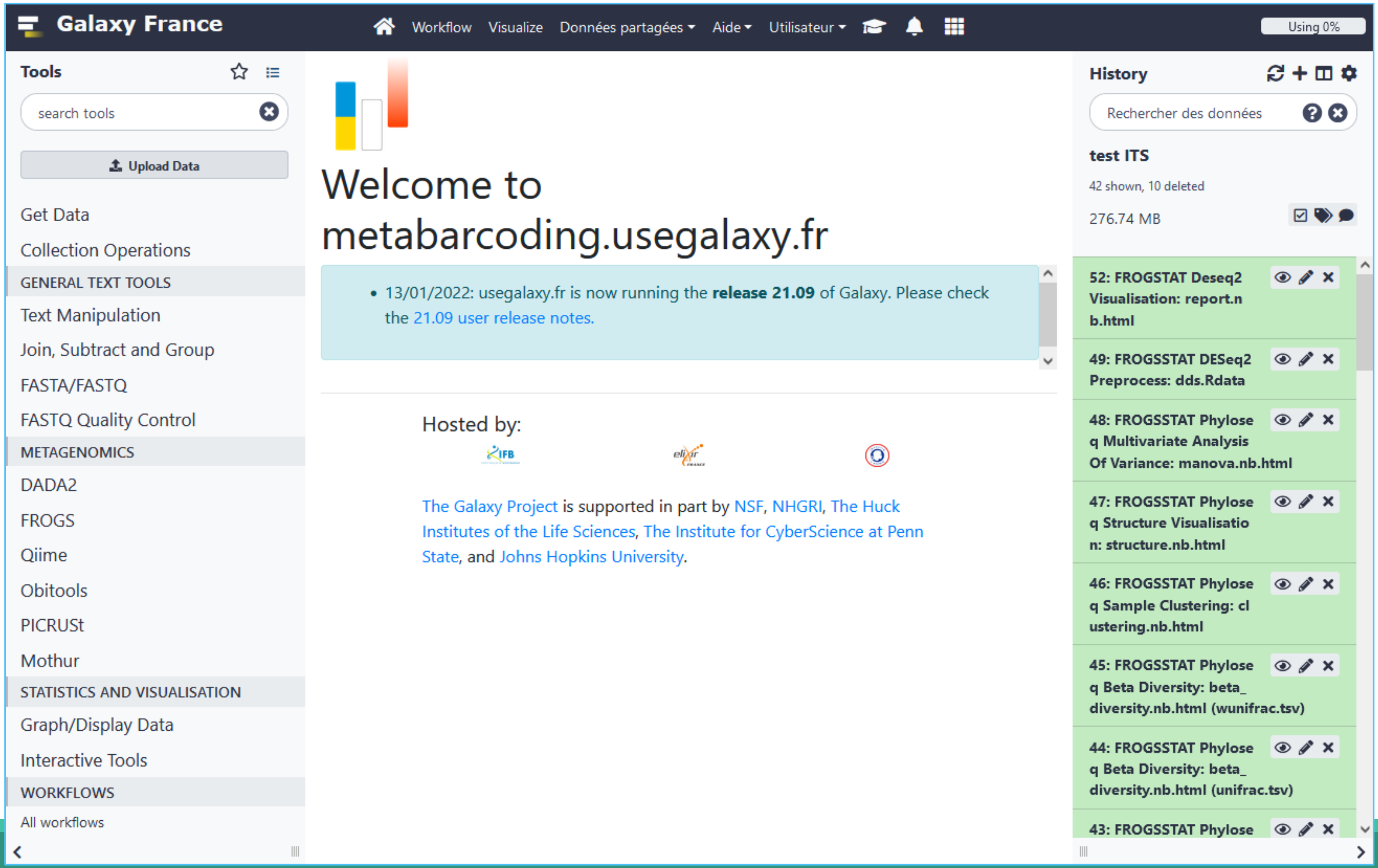
Support
Support is configured in Galaxy to send help requests to support.sigenae@inrae.fr
Use this address to contact us in case of problems

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.

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/ (Genotoul Platform)

Or

https://metabarcoding.usegalaxy.fr/root/login?is_logout_redirect=true

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

Exercise

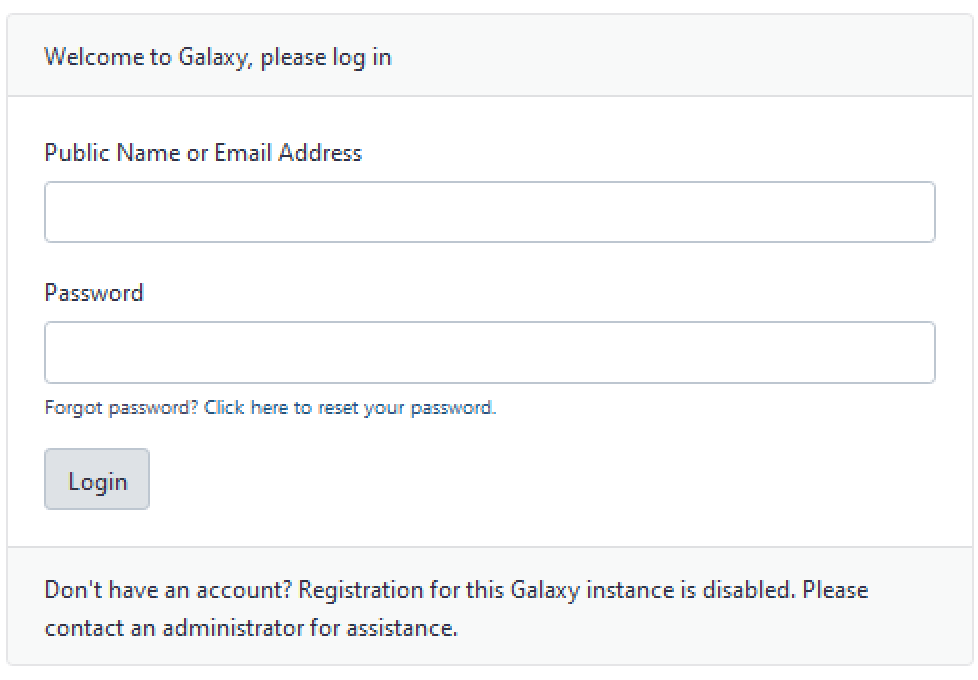
- The Virtual machine during the formation: https://vm-galaxy-prod.toulouse.inra.fr/galaxy_fmt02

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

- Enter your login and your password



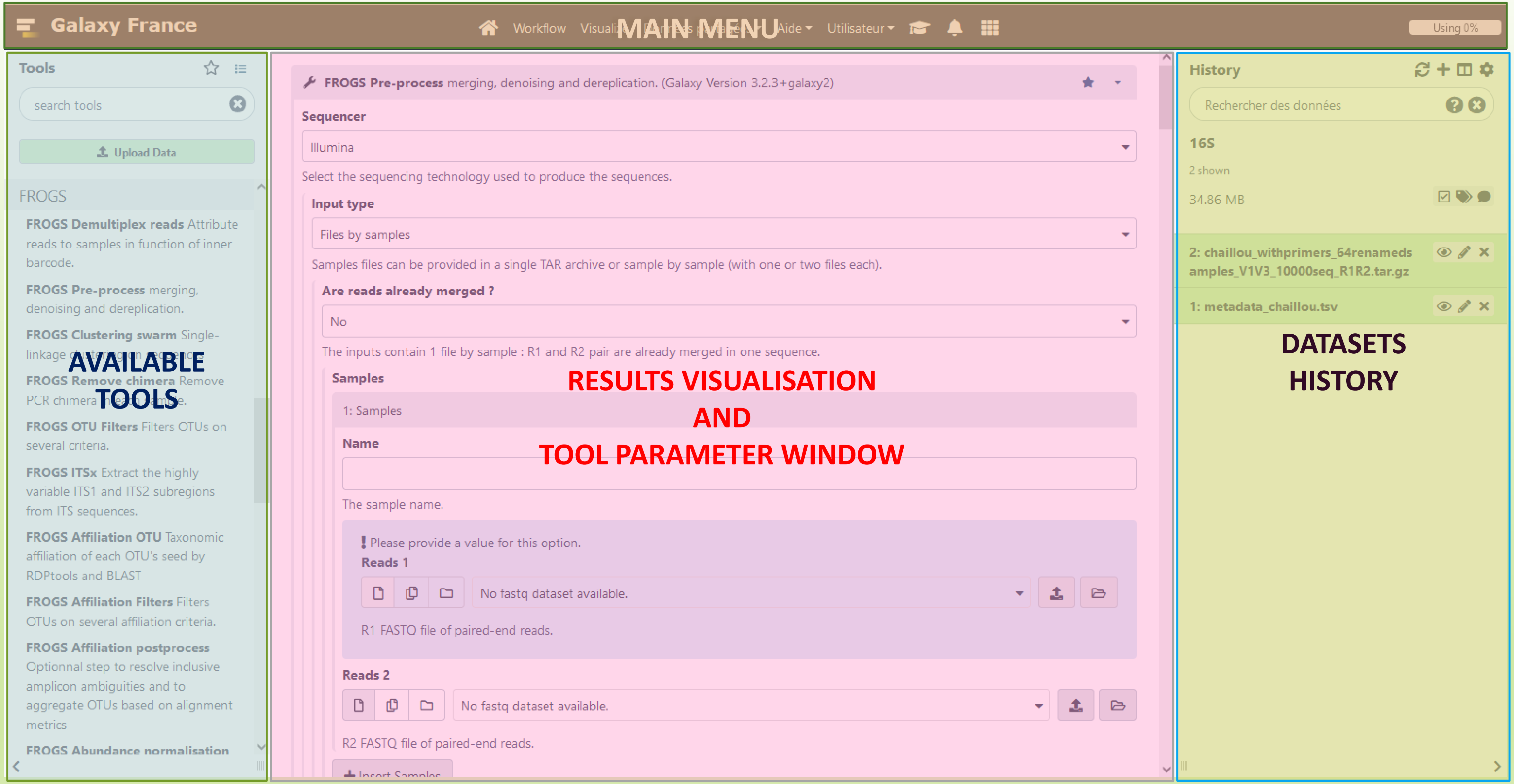
Welcome to Galaxy, please log in

Public Name or Email Address

Password

Forgot password? Click here to reset your password.

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

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

FROGS Affiliation Filters Filters OTUs on several affiliation criteria.

FROGS Affiliation postprocess Optionnal 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.

R1 FASTQ file of paired-end reads.

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:

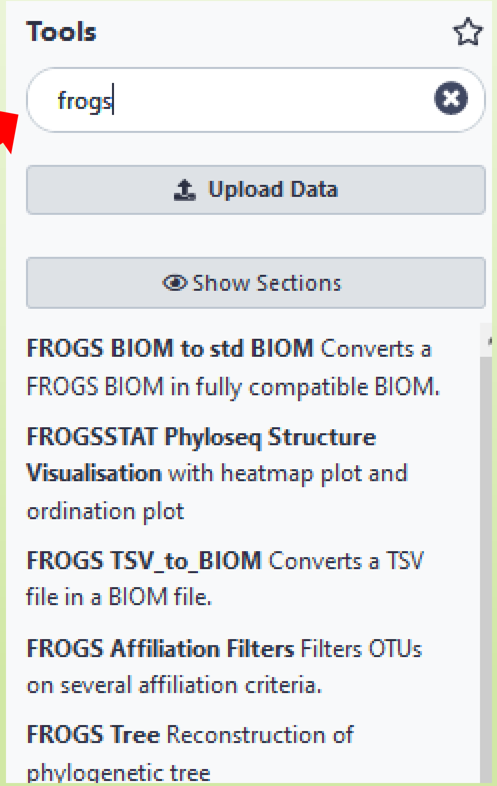
DISCOVER GALAXY

Exercise

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

Exercise

Search a tool by name.



frogs

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

If you search
FROGS, all tools
will be disordered
☹

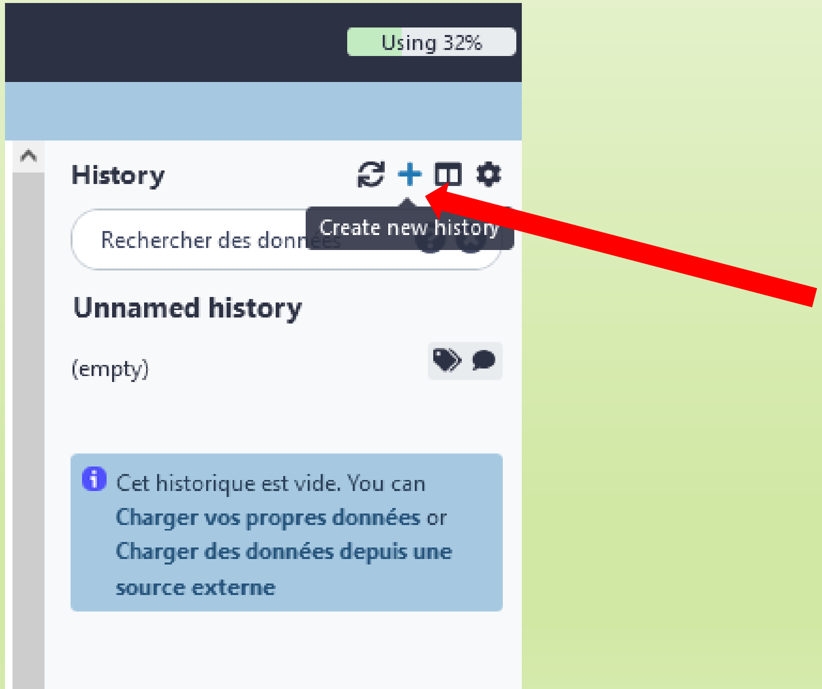
Manipulate Histories

Practice:

CREATE THE 2 HISTORIES NEEDED FOR THE FROGS FORMATION

To create a new history:

- Click on the plus.

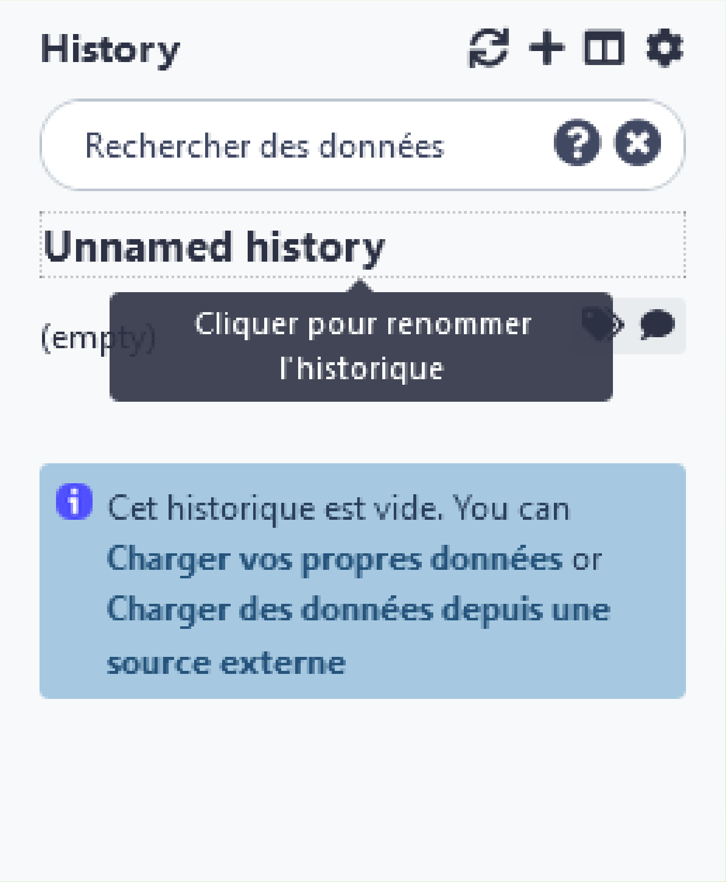


Cet historique est vide. You can Charger vos propres données or Charger des données depuis une source externe

To rename a history:

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

Don’t use special characters or accents!



History

Rechercher des données

Unnamed history

(empty)

Cliquer pour renommer l'historique

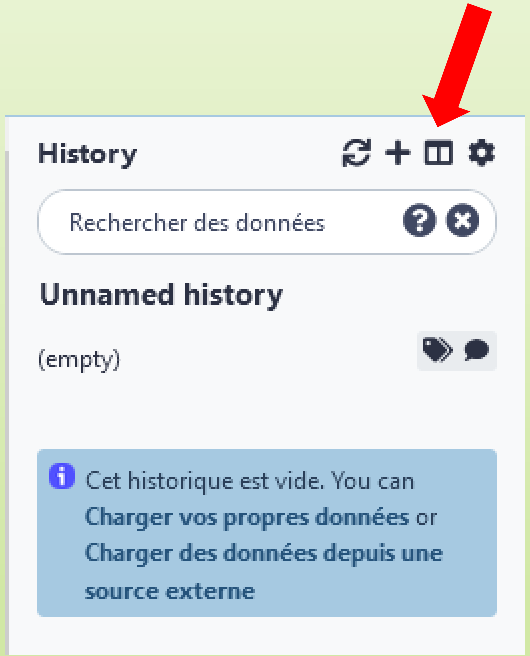
Cet historique est vide. You can Charger vos propres données or Charger des données depuis une source externe

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.

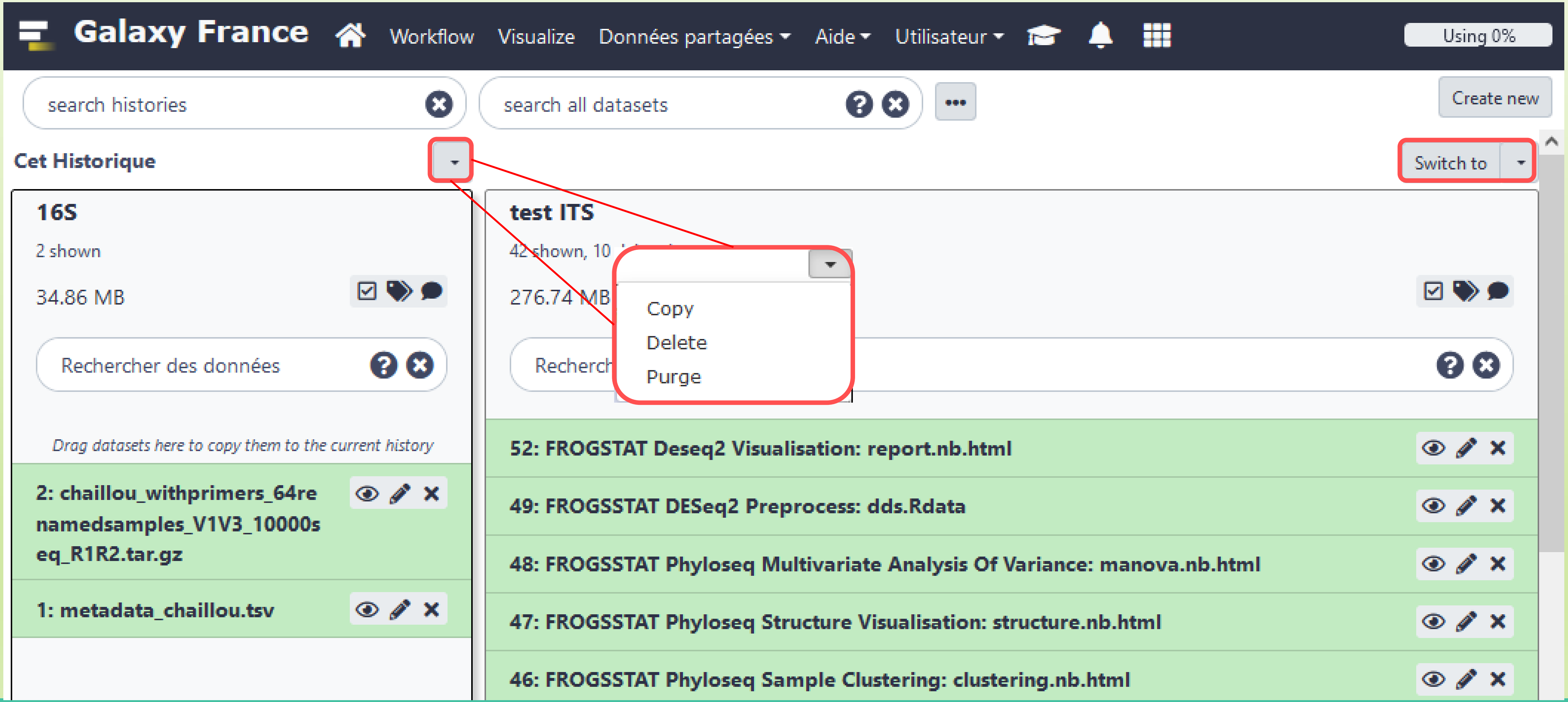


Unnamed history

(empty)

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

Explore the « View all histories » section



16S
2 shown
34.86 MB

Drag datasets here to copy them to the current history

2: chaillou_withprimers_64renamedsamples_V1V3_10000seq_R1R2.tar.gz

1: metadata_chaillou.tsv

test ITS
42 shown, 10
276.74 MB

52: FROGSTAT Deseq2 Visualisation: report.nb.html

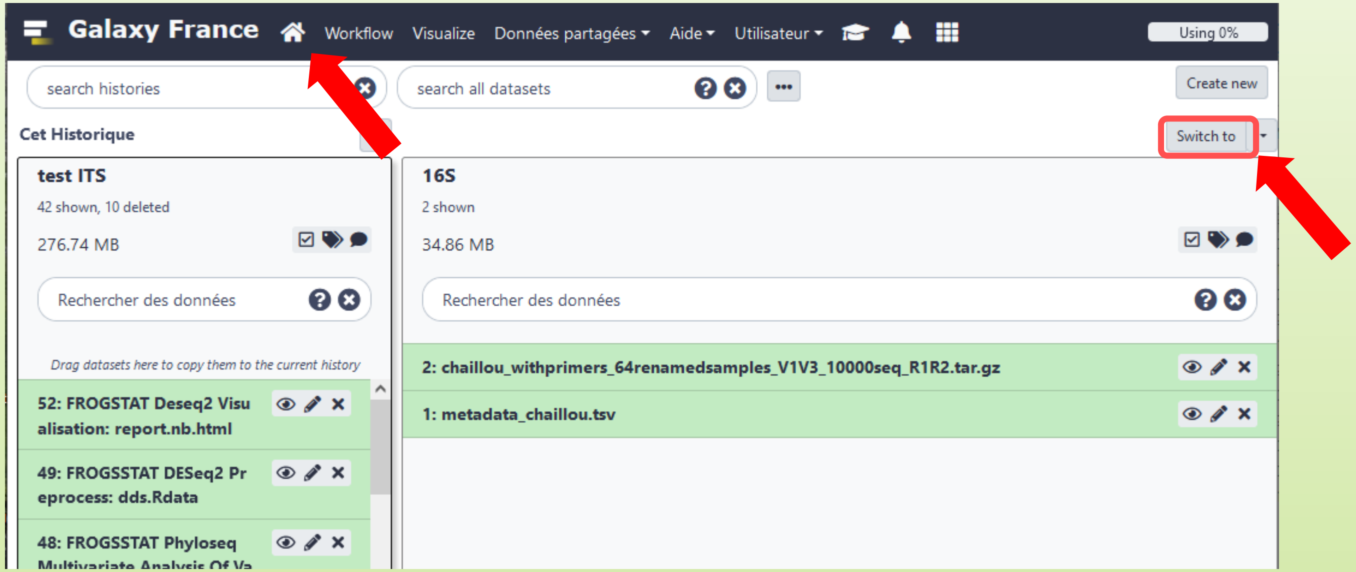
49: FROGSSTAT DESeq2 Preprocess: dds.Rdata

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

Switch current history



- Switch to the history named « 16S » as current history.
- Click on “Home” to go back to the main interface.

Data import

How to import your data to Galaxy ?

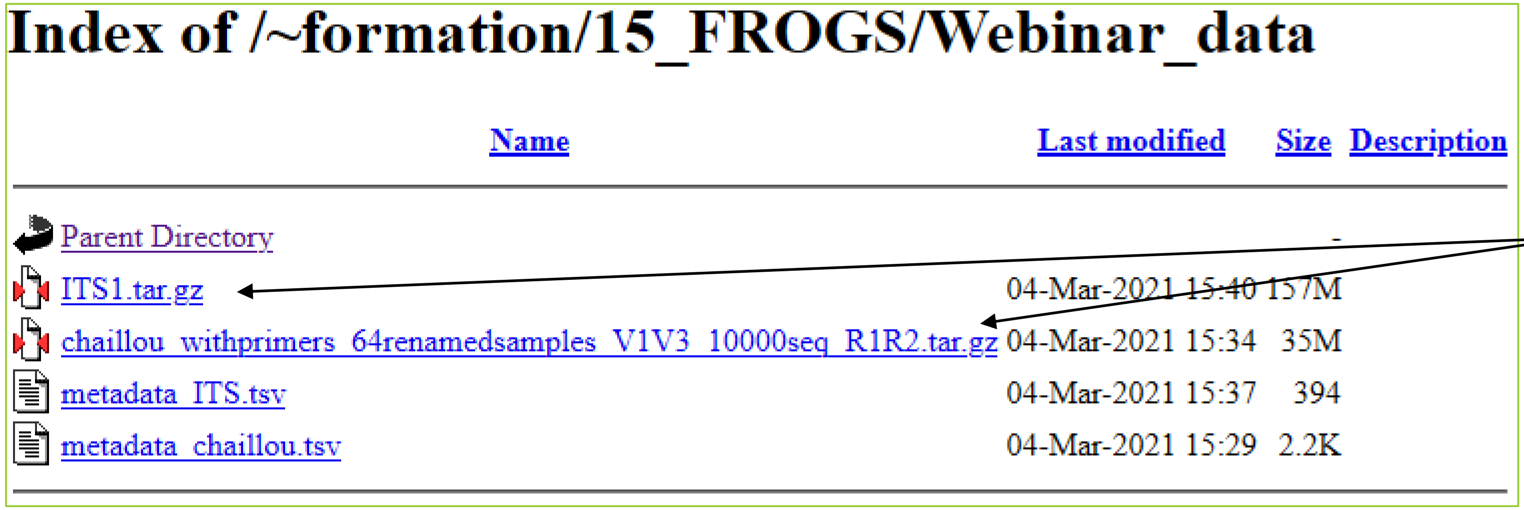
- Several ways to upload your data to Galaxy:
  - From your computer

  - 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

drag-and-drop files

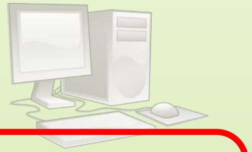
Paste an url

Practice

UPLOAD FILE FROM AN URL

How to import your data to Galaxy ?

- Several ways to upload your data to Galaxy:
  - From your computer

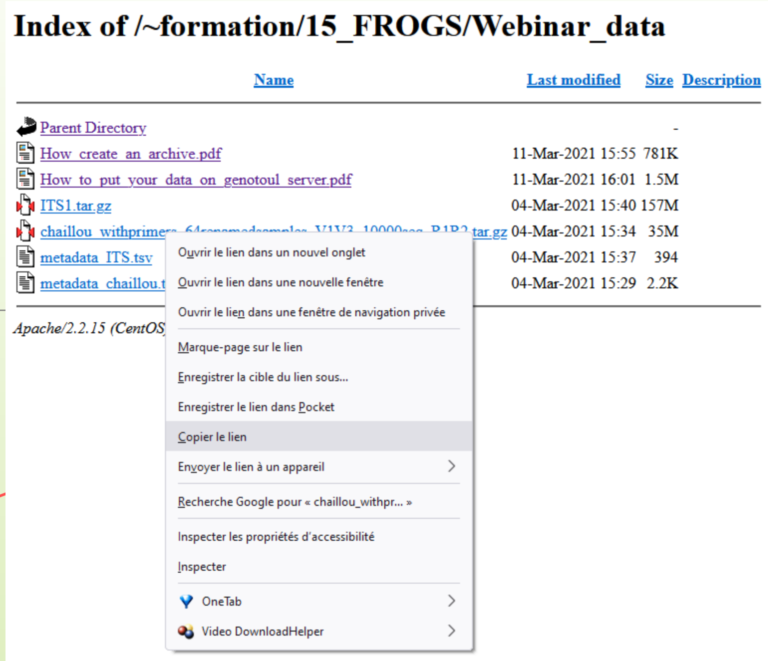


  - By URL

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

16S history creation

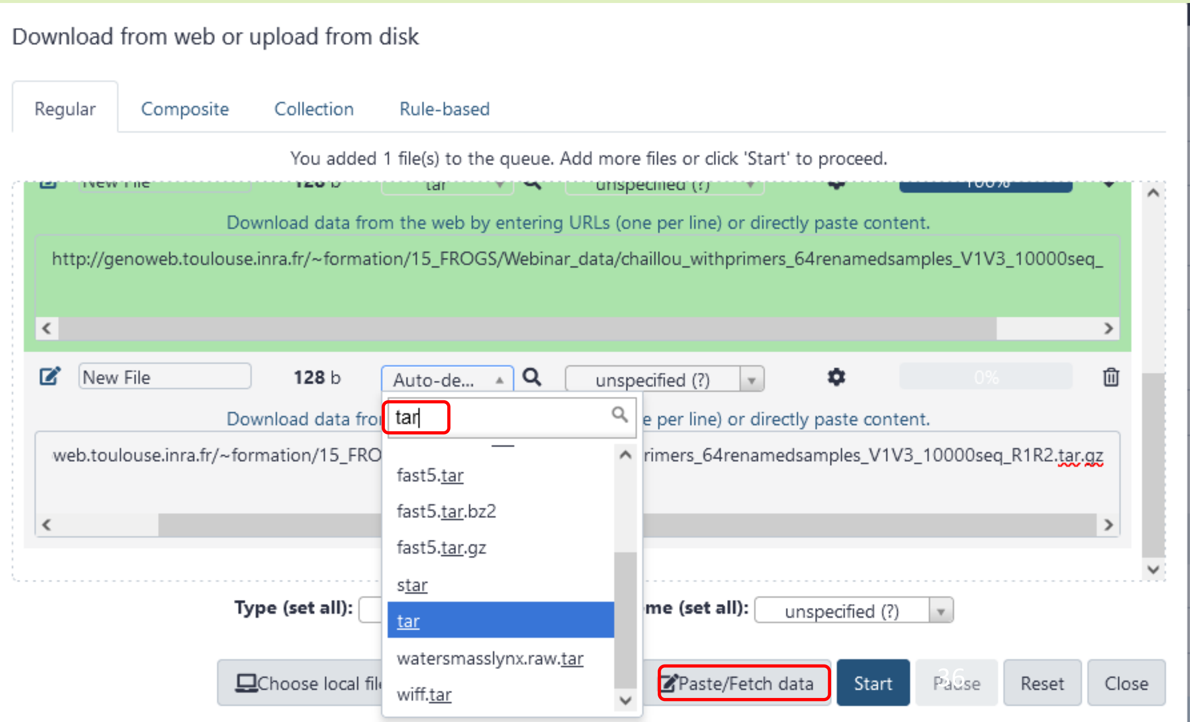
1. In your current history “16S”
2. Go to **Get Data** > **Upload File from your computer**
3. Click on **Paste/Fetch Data**
4. Click right on the name file
5. Copy the address of the link
6. Paste the address
7. Change the data type!
8. Click on Start
9. You can put one address per line for multiple uploads
10. Start again with the second file



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



Download from web or upload from disk

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

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

http://genoweb.toulouse.inra.fr/~formation/15_FROGS/Webinar_data/chaillou_withprimers_64renamedsamples_V1V3_10000seq_

New File 128 b

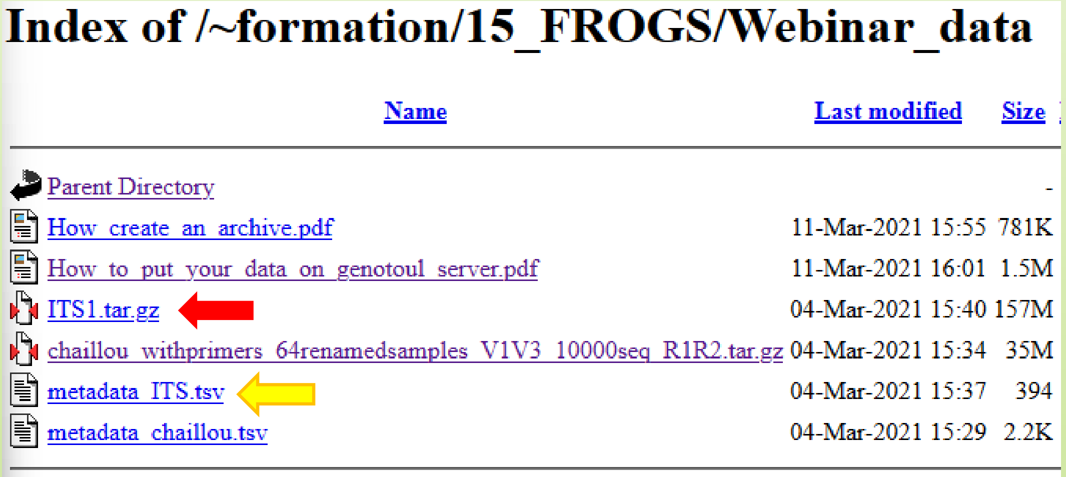
Download data fro e per line) or directly paste content.

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

Type (set all): me (set all): unspecified (?)

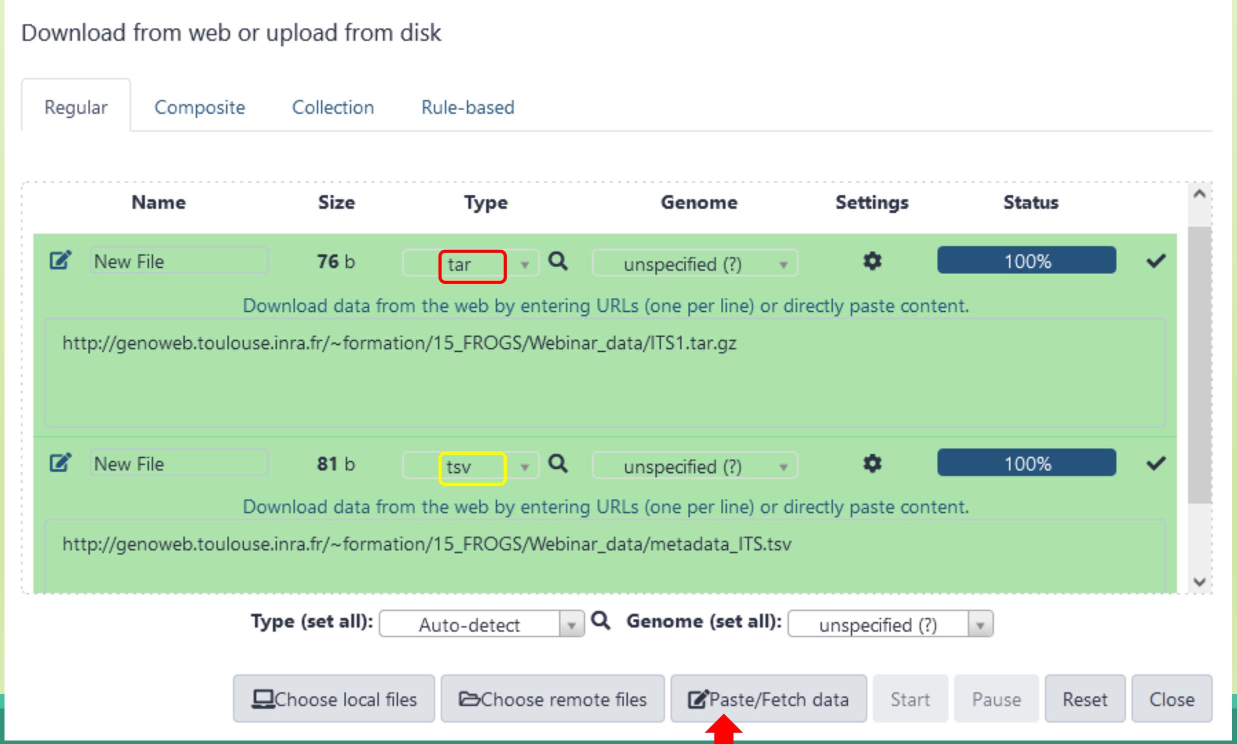
Create and fill **yourself** the ITS history

Get data form here:
http://genoweb.toulouse.inra.fr/~formation/15_FROGS/Webinar_data/



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



Download from web or upload from disk

Name	Size	Type	Genome	Settings	Status
New File	76 b	tar	unspecified (?)		100%

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

http://genoweb.toulouse.inra.fr/~formation/15_FROGS/Webinar_data/ITS1.tar.gz

Name	Size	Type	Genome	Settings	Status
New File	81 b	tsv	unspecified (?)		100%

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

http://genoweb.toulouse.inra.fr/~formation/15_FROGS/Webinar_data/metadata_ITS.tsv

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

Create and fill **yourself** the ITS history

You have to obtain your second history:

54: metadata_ITS.tsv

16 lines, 1 comments

format: **tsv**, génome de référence: **?**

1.	2.Ech	3.Ori
	Ech	Ori
echantillon1-1	1	alim
echantillon1-2	1	alim
echantillon1-3	1	alim
echantillon2-1	2	alim

53: ITS1.tar.gz

156.9 MB

format: **tar**, génome de référence: **?**

```
Compressed binary file
```

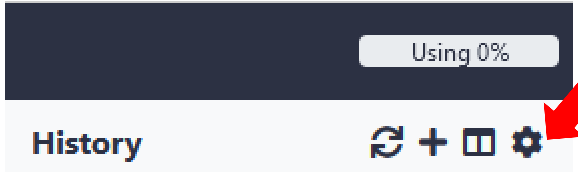
Share a History

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 Actions

- Copy
- Partager et publier
- Montrer la structure
- Extraire un Workflow
- Set Permissions
- Make Private
- Reprendre les processus en pause

Actions sur les jeux de données

- Copier des jeux de données
- Réduire les données étendues
- Afficher les données cachées
- Supprimer les données cachées
- Purger les données supprimées

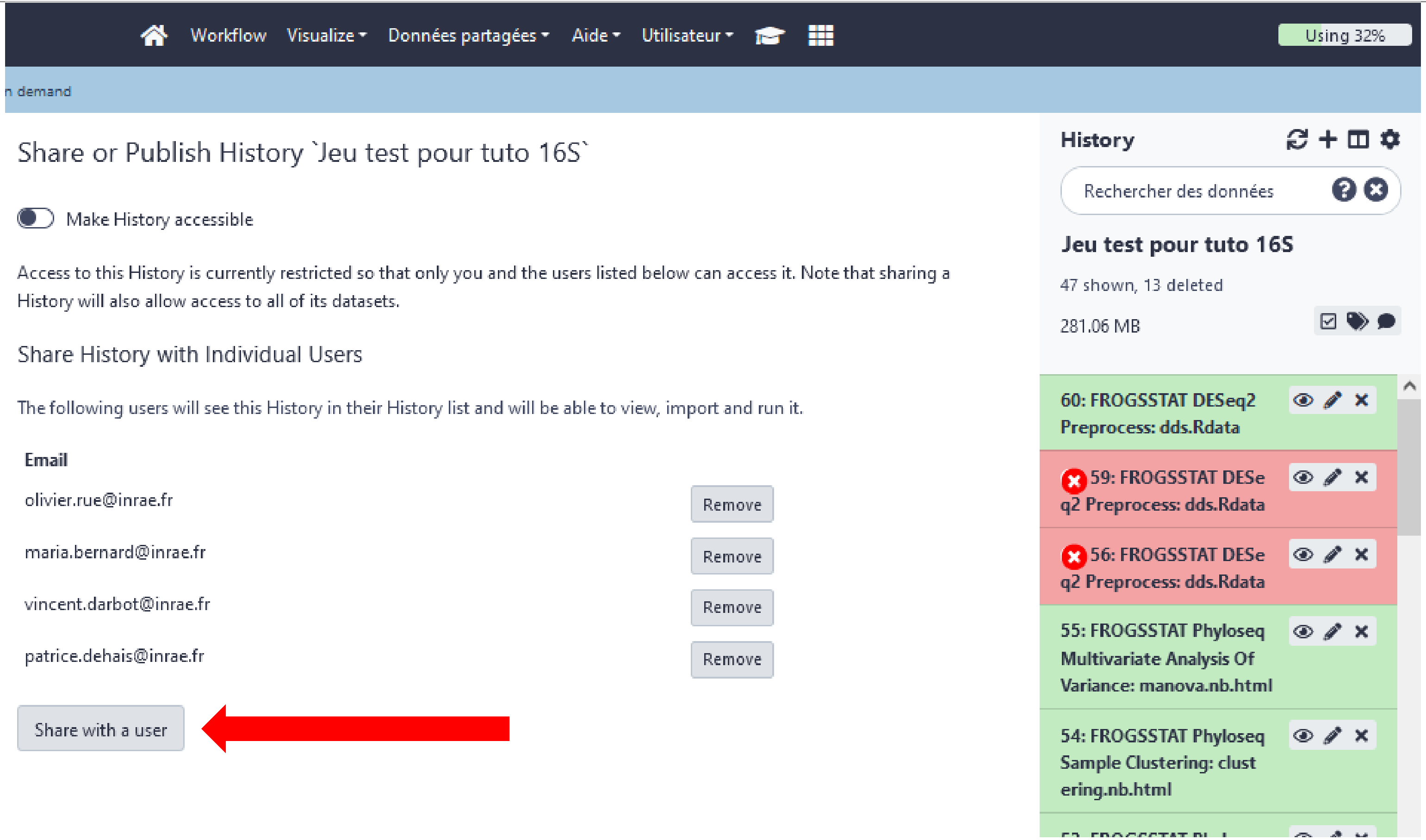
Télécharger

- Exporter les citations des outils
- Exporter l'Historique dans un fichier

Beta Features

- Use Beta History Panel

Share an history



Share or Publish History `Jeu test pour tuto 16S`

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 datasets.

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.

Email	
olivier.rue@inrae.fr	Remove
maria.bernard@inrae.fr	Remove
vincent.darbot@inrae.fr	Remove
patrice.dehais@inrae.fr	Remove

Share with a user

Jeu test pour tuto 16S

47 shown, 13 deleted

281.06 MB

60: FROGSSTAT DESeq2 Preprocess: dds.Rdata

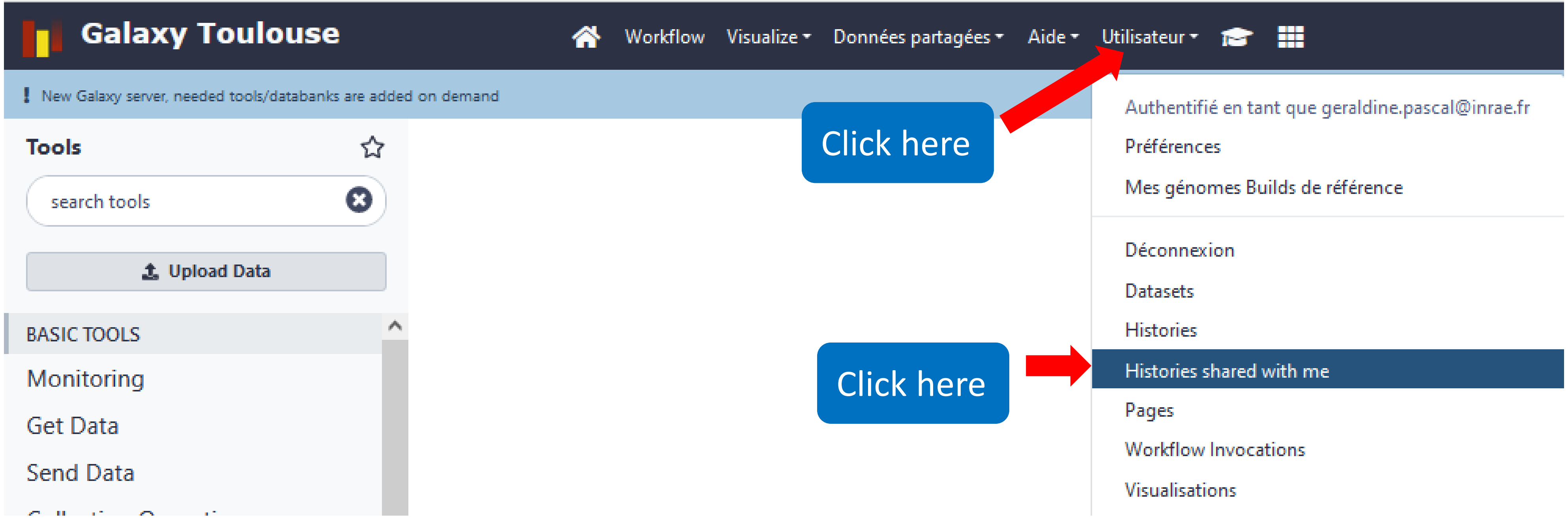
59: FROGSSTAT DESeq2 Preprocess: dds.Rdata

56: FROGSSTAT DESeq2 Preprocess: dds.Rdata

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

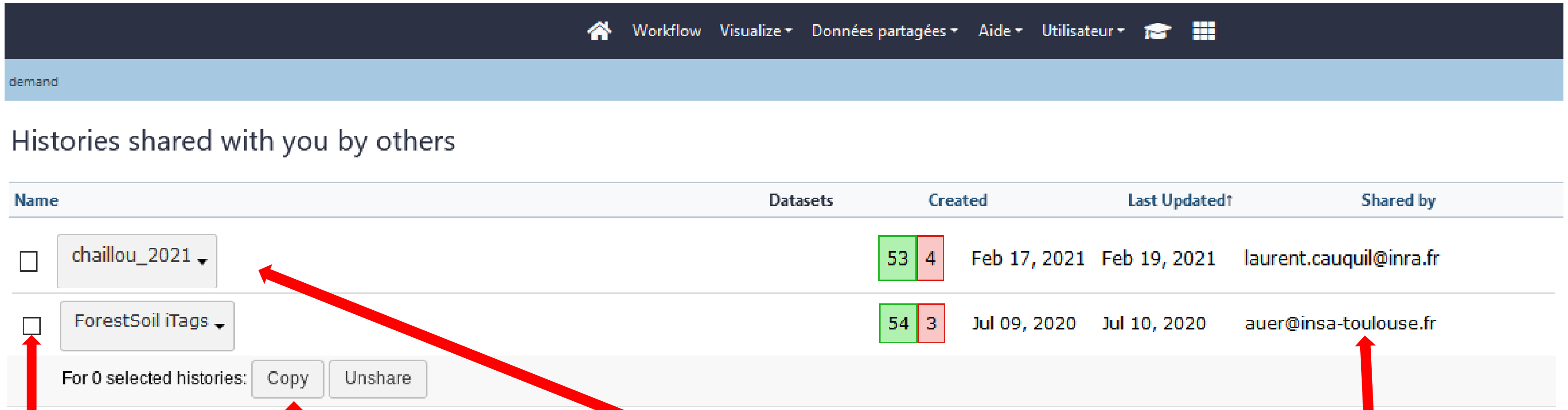
54: FROGSSTAT Phyloseq Sample Clustering: clustering.nb.html

Import a shared history



Click here

Click here



Histories shared with you by others

	Name	Datasets	Created	Last Updated↑	Shared by
☐	chaillou_2021	🟩 53 🟥 4	Feb 17, 2021	Feb 19, 2021	laurent.cauquil@inra.fr
☐	ForestSoil iTags	🟩 54 🟥 3	Jul 09, 2020	Jul 10, 2020	auer@insa-toulouse.fr

For 0 selected histories: Copy Unshare

Name of the history.

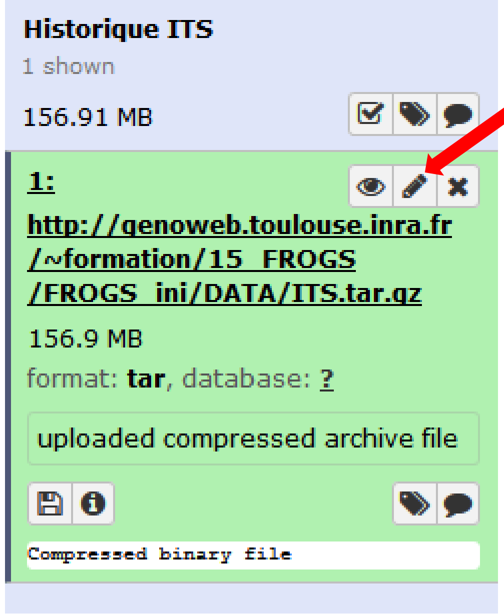
- Check the box and click on « Copy » to import an history into your Galaxy account.

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.gz**

156.9 MB

format: **tar**, database: **?**

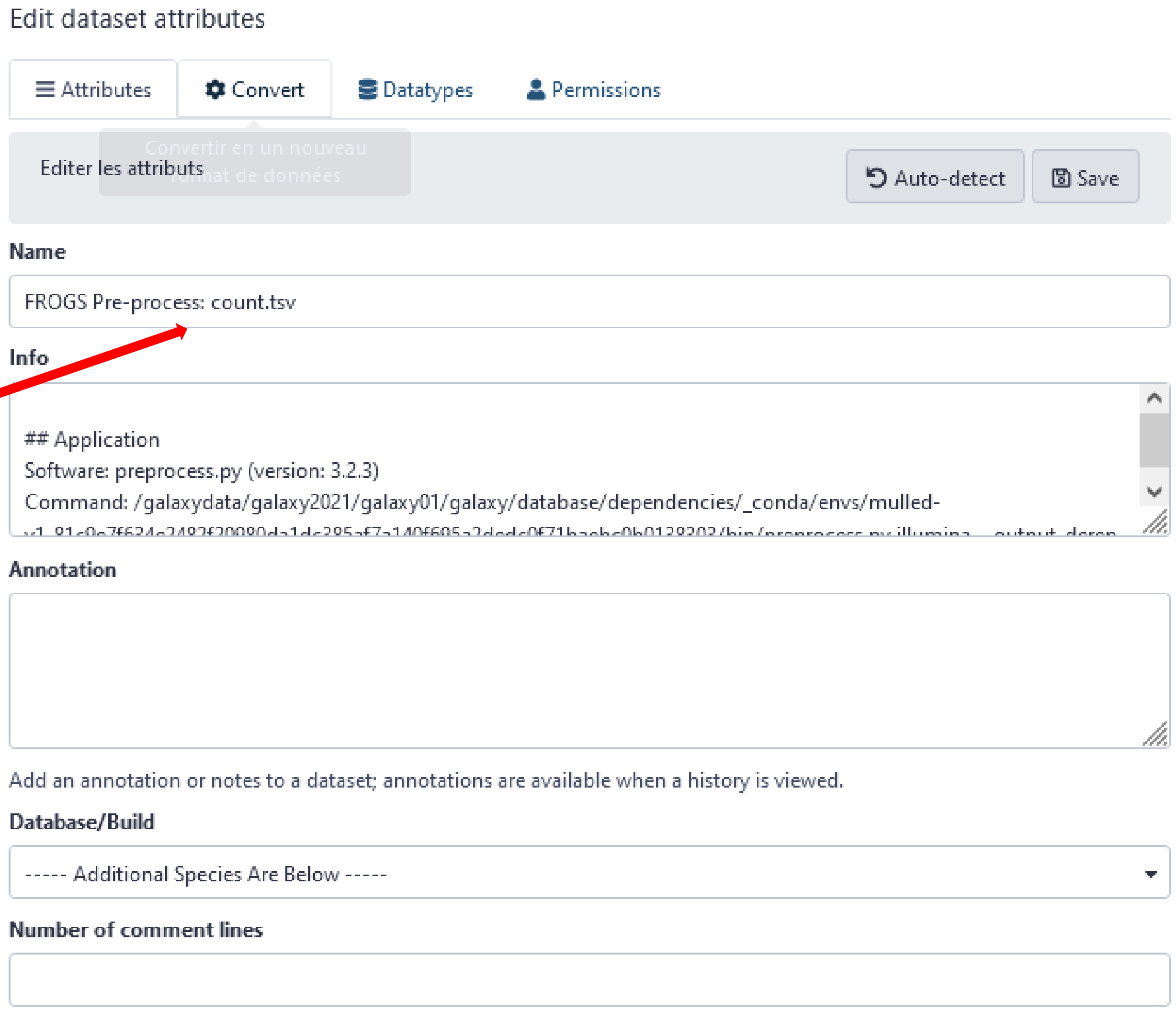
uploaded compressed archive file

```
Compressed binary file
```

Click here to display attributes and change the name.

To rename a dataset

Change the name here



Name

FROGS Pre-process: count.tsv

Info

Application
Software: preprocess.py (version: 3.2.3)
Command: /galaxydata/galaxy2021/galaxy01/galaxy/database/dependencies/_conda/envs/mulled-

Annotation

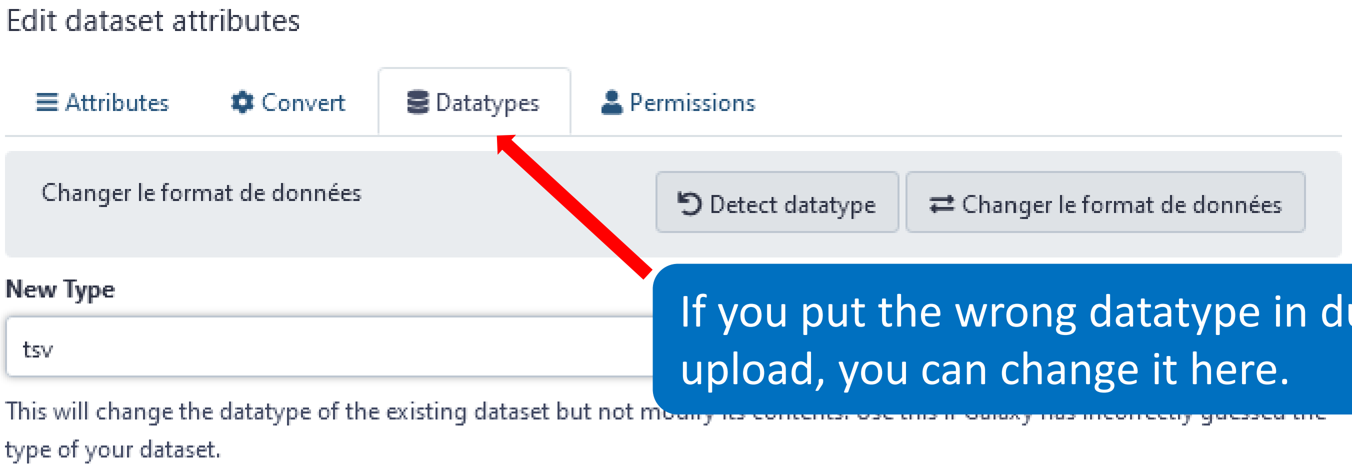
Add an annotation or notes to a dataset; annotations are available when a history is viewed.

Database/Build

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

Number of comment lines

To change the datatype



Edit dataset attributes

Changer le format de données

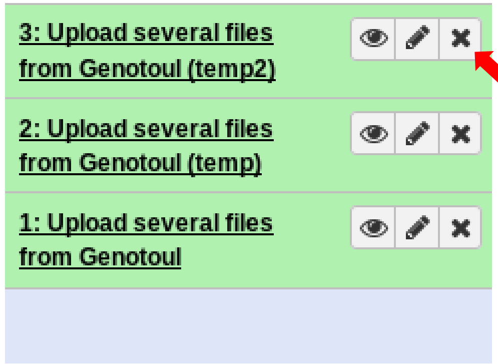
New Type

tsv

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



3: Upload several files from Genotoul (temp2)

2: Upload several files from Genotoul (temp)

1: Upload several files from Genotoul

Click here to delete a dataset.

To delete a dataset

Click here.

History

Rechercher des données

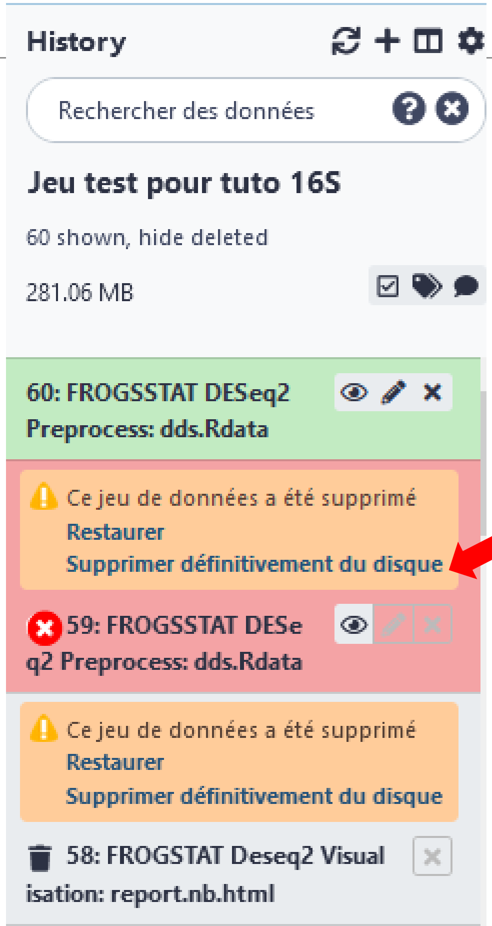
Jeu test pour tuto 16S

46 shown, 14 deleted

281.06 MB

60: FROGSSTAT DESeq2 Preprocess: dds.Rdata

56: FROGSSTAT DESeq2 Preprocess: dds.Rdata



History

Rechercher des données

Jeu test pour tuto 16S

60 shown, hide deleted

281.06 MB

60: FROGSSTAT DESeq2 Preprocess: dds.Rdata

Ce jeu de données a été supprimé
Restaurer
Supprimer définitivement du disque

59: FROGSSTAT DESeq2 Preprocess: dds.Rdata

Ce jeu de données a été supprimé
Restaurer
Supprimer définitivement du disque

58: FROGSTAT Deseq2 Visualisation: report.nb.html

Delete this dataset permanently.

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

Galaxy support

- For Sigenae instance on GenoToul Bioinfo platform - mail: support.sigenae@inrae.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.