Formation FROGS juillet 2019

Géraldine Pascal, Laurent Cauquil

Géraldine Pascal, Laurent Cauquil. Formation FROGS juillet 2019. Master. (Trainingon Galaxy : Metagenomics / ITS analysisin FROGS ? / Trainingon Galaxy: Statistics to explore metagenomics), 2019. hal-02790029

HAL Id: hal-02790029
https://hal.inrae.fr/hal-02790029v1

Training on Galaxy: Metagenomics

July 2019

Find, Rapidly, OTUs with Galaxy Solution

FRÉDÉRIC ESCUDIÉ* and LUCAS AUER*, MARIA BERNARD, LAURENT CAUQUIL, SARAH MAMAN, MAHENDRA MARIADASSOU, SYLVIE COMBES, GUILLERMINA HERNANDEZ-RAQUET, GÉRALDINE PASCAL & MALO LE BOULCH

*THESE AUTHORS HAVE CONTRIBUTED EQUALLY TO THE PRESENT WORK.

Introduction to Galaxy platform and preparation of FROGS training

July 2019

MALO LE BOULCH, SARAH MAMAN, GÉRALDINE PASCAL

Chitchat time!

- What is your computer skills level?
- Have you ever heard of or used Galaxy?

Feedback on metagenomics:

- What are your needs in “metagenomics"?
- 454 / MiSeq?
- 16S/ITS ?
- Your background ?

Monday
afternoon

- Frogs
- Coffee break
- Frogs

Tuesday

- Frogs
- Coffee break
- Frogs
- Lunch
- Frogs
- Coffee break
- Frogs

Wednesday

- Frogs
- Coffee break
- Frogs
- Lunch
- Statistics
- Coffee break
- Statistics

Thrusday

- Statistics
- Coffee break
- Statistics
- Lunch
- FROGS & Statistics
- Coffee break
- FROGS & Statistics

9 am to 5 pm

2 short coffee breaks
morning and afternoon

Lunch
12.30 to 2.00 pm

Objectives of the 1st part

- Learn the basics of Galaxy
- Being independent when using it
- Prepare the datasets for FROGS formation

What is Galaxy?

Galaxy software framework

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

- Developed by the Galaxy Team Project.

- Relies on a computational infrastructure (Server).

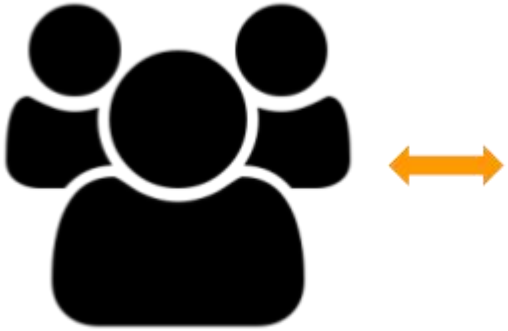
- Open source

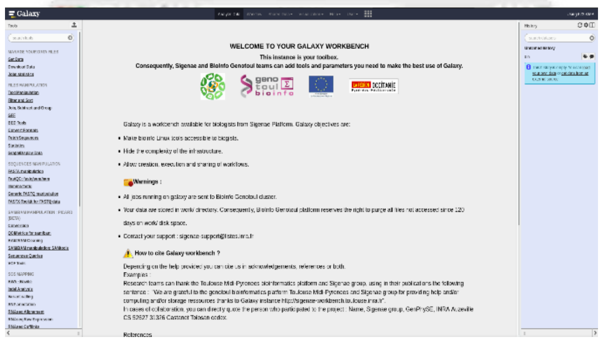
Galaxy software framework

User

Web browser

Servers





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.

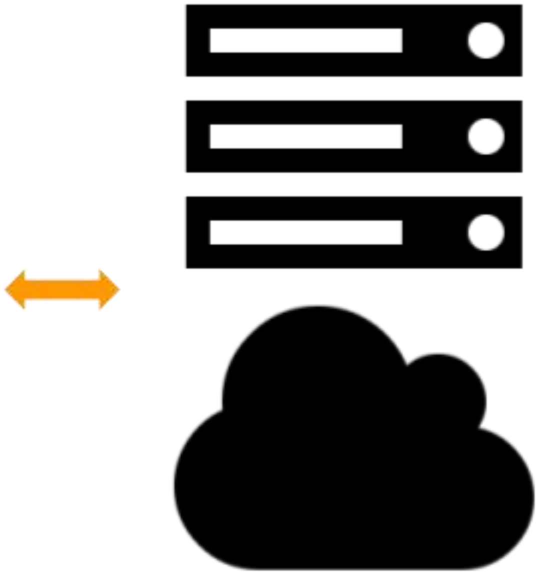
Warnings :

- 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



You

What you see

What you don't see

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

Where to use Galaxy?

- Galaxy software can be installed and run on powerful server farms (Cluster) or on your individual PC.
- Genotoul Bioinfo
  - Cluster: Many computers tightly connected that work together
  - High performance computer:
    - More than 5000 cores
    - 34 TB of RAM
    - More than 1 Peta Byte (1024 TB) of hard drive

Where to use Galaxy?

- Our Galaxy platform is http://sigenae-workbench.toulouse.inra.fr/galaxy/

⚠ The different platforms of Galaxy are not connected together

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

Exemple of 2 INRA Galaxy platforms

✓ **Bienvenue sur le portail Galaxy de la plateforme Migale. Pour tous renseignements, demandes ou remarques, veuillez contacter galaxy-help@jouy.inra.fr**

Plateforme de BioInformatique - INRA Jouy en Josas

Galaxy est une plateforme qui propose une « constellation » d'outils pour analyser, manipuler et visualiser des données génomiques, sans avoir besoin de connaissance en programmation. Elle est développée par The Center for Comparative Genomics and Bioinformatics. L'utilisateur peut réaliser quatre grands types d'opérations :

- **manipulation de fichiers** : ajout ou suppression de colonnes, trier les fichiers, concaténer plusieurs fichiers, comparaison de listes, ...
- **opérations sur les données** : sommer, moyenner, soustraire, calculer la couverture d'une région déterminée, ...
- **analyse de séquences** : calculer des corrélations, utiliser des outils d'EMBOSS, aligner les données de séquençage, ...
- **visualisation des données** : afficher des alignements multiples, générer des graphiques, ...

Contact : galaxy-help@jouy.inra.fr

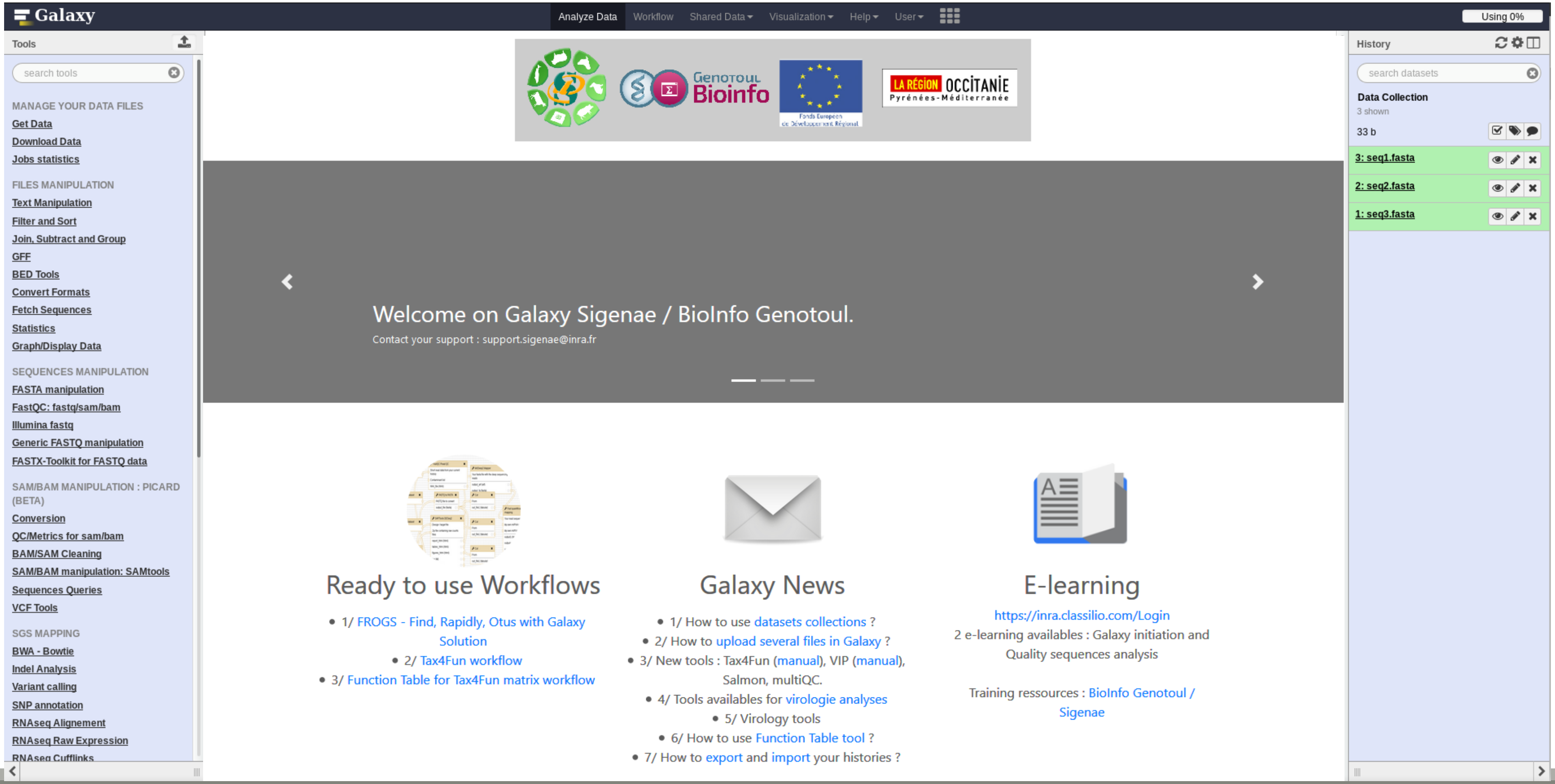
Pour toutes demandes d'intégration de nouvel outil au sein du portail, veuillez remplir le formulaire mis à votre disposition sur le site web de la plateforme Migale.

Une Foire Aux Questions autour de Galaxy a été mise en place sur le site web de la plateforme Migale.

Si vous utilisez le portail Galaxy de la plateforme Migale pour effectuer vos analyses, merci de citer Galaxy ET de remercier la plateforme dans les acknowledgements.

This project is supported in part by NSF, NHGRI, and the Huck Institutes of the Life Sciences.

Exemple of 2 INRA Galaxy platforms



Welcome on Galaxy Sigenae / BioInfo Genotoul.

Contact your support : support.sigenae@inra.fr

Ready to use Workflows

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

Galaxy News

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

E-learning

https://inra.classilio.com/Login

2 e-learning availables : Galaxy initiation and Quality sequences analysis

Training ressources : BioInfo Genotoul / Sigenae

Your Turn!

CONNECT TO OUR GALAXY WORKBENCH

Exercise

During this formation, we will use a remote computer.

Connect to a distant computer by clicking on the remote desktop connection icon in the task bar.

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 formation if you have a personal account, use it.

This account will save all the work done during this week and your future work.

If you don't have a personal account ask us for a temporary account.

You can request a personal account later on : http://bioinfo.genotoul.fr/index.php/ask-for/create-an-account/

All your data will be erased at the end of the week on a temporary account

Exercise

Our Galaxy platform is: http://sigenae-workbench.toulouse.inra.fr/galaxy/

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

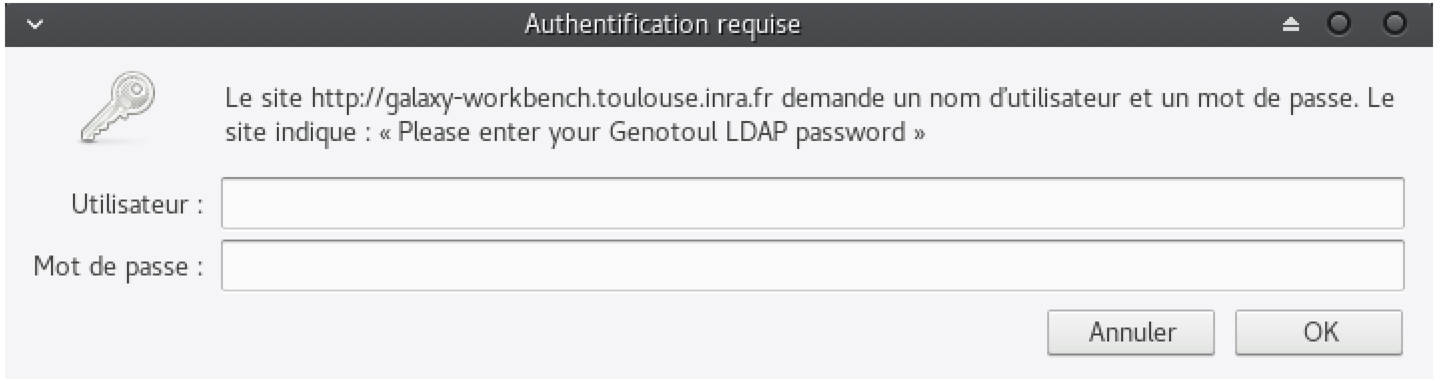
Risk of data duplication

Exercise

Our Galaxy platform is: http://sigenae-workbench.toulouse.inra.fr/galaxy/

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

- The first time in this pop-up window:



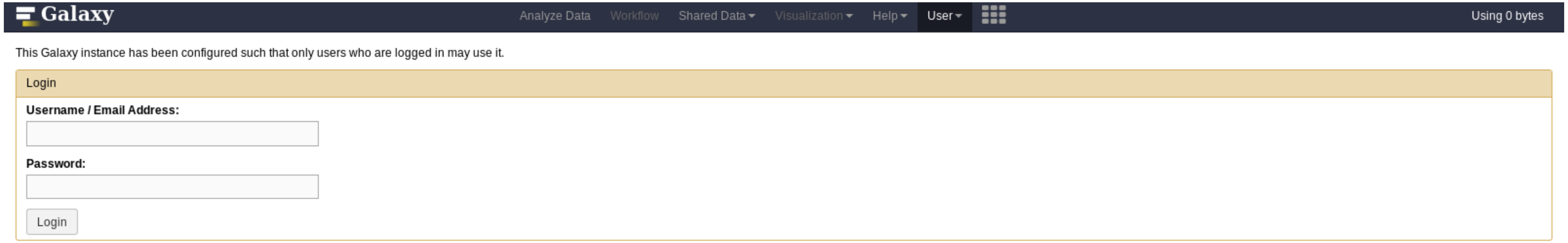
Le site http://galaxy-workbench.toulouse.inra.fr demande un nom d'utilisateur et un mot de passe. Le site indique : « Please enter your Genotoul LDAP password »

Utilisateur :

Mot de passe :

Exercise

- And a second time, in the browser:



This Galaxy instance has been configured such that only users who are logged in may use it.

Login

Username / Email Address:

Password:

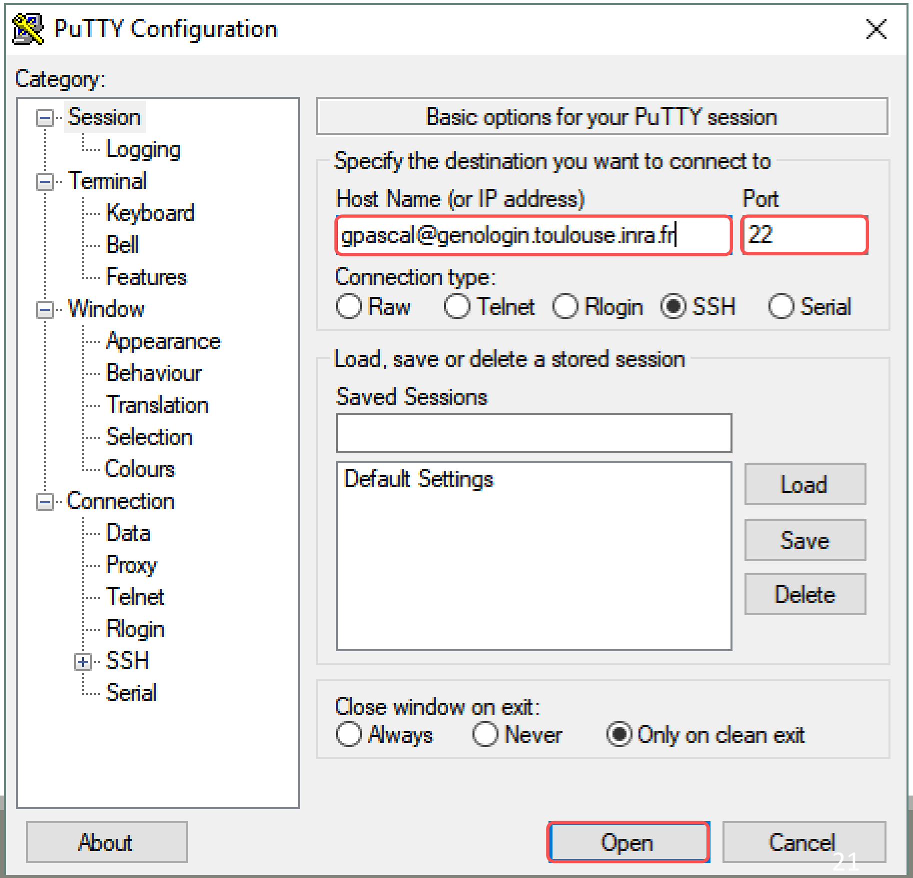
Change your password

- In order to change your password, you need to download PuTTY: http://www.putty.org/
- PuTTY is a terminal emulator, it allows to connect directly to the server in command line.
- You can not change your password via the Galaxy’s interface for the moment.

Change your password

- Launch PuTTY.
- In the following window, you must enter your host name which is:

  YourGenotoulName@genologin.toulouse.inra.fr
- The connection type must be SSH.
- Click on « Open » .



PuTTY Configuration

Category:

- Session
  - Logging
- Terminal
  - Keyboard
  - Bell
  - Features
- Window
  - Appearance
  - Behaviour
  - Translation
  - Selection
  - Colours
- Connection
  - Data
  - Proxy
  - Telnet
  - Rlogin
  - SSH
  - Serial

Basic options for your PuTTY session

Specify the destination you want to connect to

Host Name (or IP address): gpascal@genologin.toulouse.inra.fr

Port: 22

Connection type:
○ Raw ○ Telnet ○ Rlogin ● SSH ○ Serial

Load, save or delete a stored session

Saved Sessions

Default Settings

Load

Save

Delete

Close window on exit:
○ Always ○ Never ● Only on clean exit

About

Open

Cancel

Change your password

- A new window appear, click on « Yes ».
- In the following window, enter your password and hit « Enter ».

```
Using username "gpascal".
gpascal@genologin.toulouse.inra.fr's password:
```

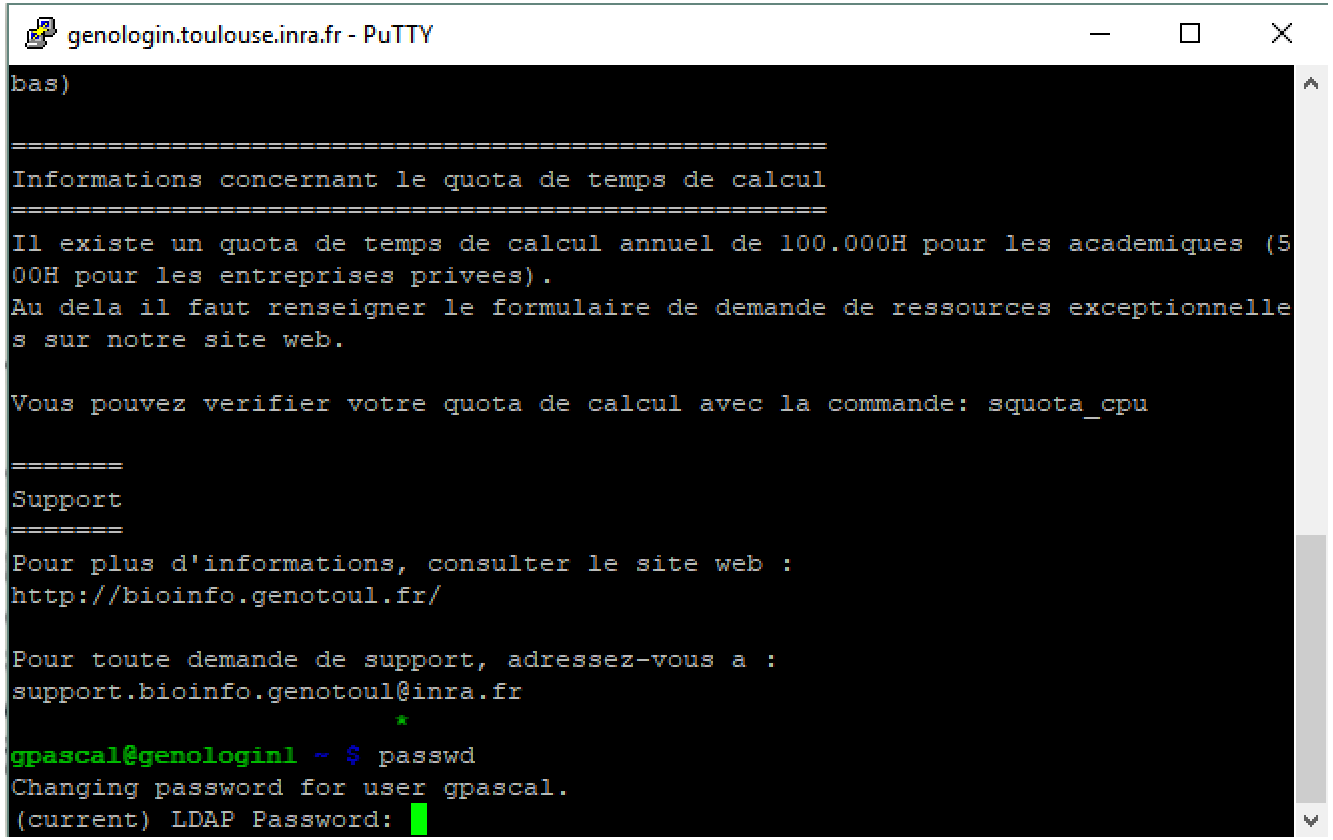
Change your password

- Type « passwd » and hit « Enter ».

```
e cluster).
Pour obtenir plus de memoire, veuillez consulter la FAQ de notre site web (cf +
bas)

====================================================
Informations concernant le quota de temps de calcul
====================================================
Il existe un quota de temps de calcul annuel de 100.000H pour les academiques (5
00H pour les entreprises privees).
Au dela il faut renseigner le formulaire de demande de ressources exceptionnelle
s sur notre site web.

Vous pouvez verifier votre quota de calcul avec la commande: squota_cpu

=======
Support
=======
Pour plus d'informations, consulter le site web :
http://bioinfo.genotoul.fr/

Pour toute demande de support, adressez-vous a :
support.bioinfo.genotoul@inra.fr
                          *
gpascal@genologin1 ~ $ passwd
```

Change your password

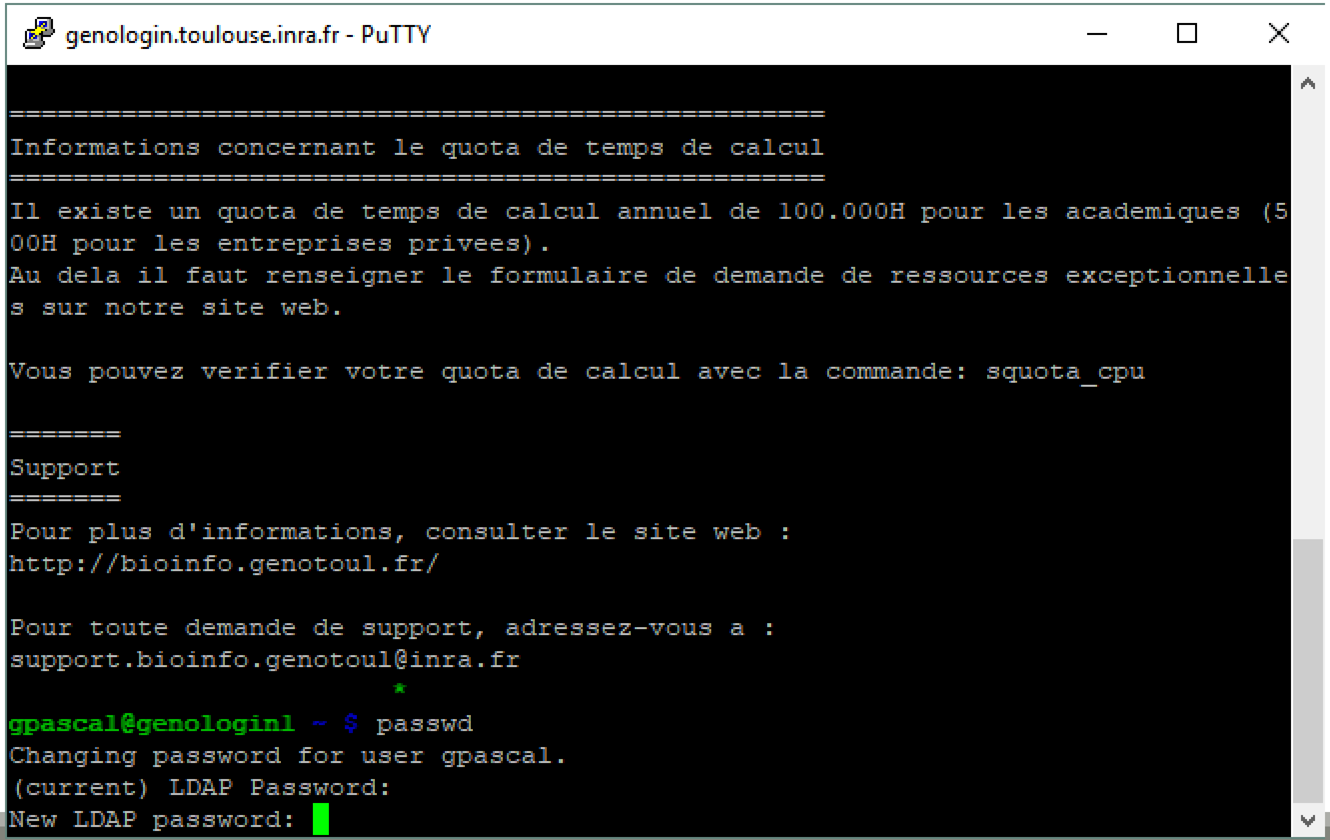
- Enter your current password and hit « Enter ».



```
bas)

===================================================
Informations concernant le quota de temps de calcul
===================================================
Il existe un quota de temps de calcul annuel de 100.000H pour les academiques (5
00H pour les entreprises privees).
Au dela il faut renseigner le formulaire de demande de ressources exceptionnelle
s sur notre site web.

Vous pouvez verifier votre quota de calcul avec la commande: squota_cpu

=======
Support
=======
Pour plus d'informations, consulter le site web :
http://bioinfo.genotoul.fr/

Pour toute demande de support, adressez-vous a :
support.bioinfo.genotoul@inra.fr
                        *
gpascal@genologin1 ~ $ passwd
Changing password for user gpascal.
(current) LDAP Password:
```

Change your password

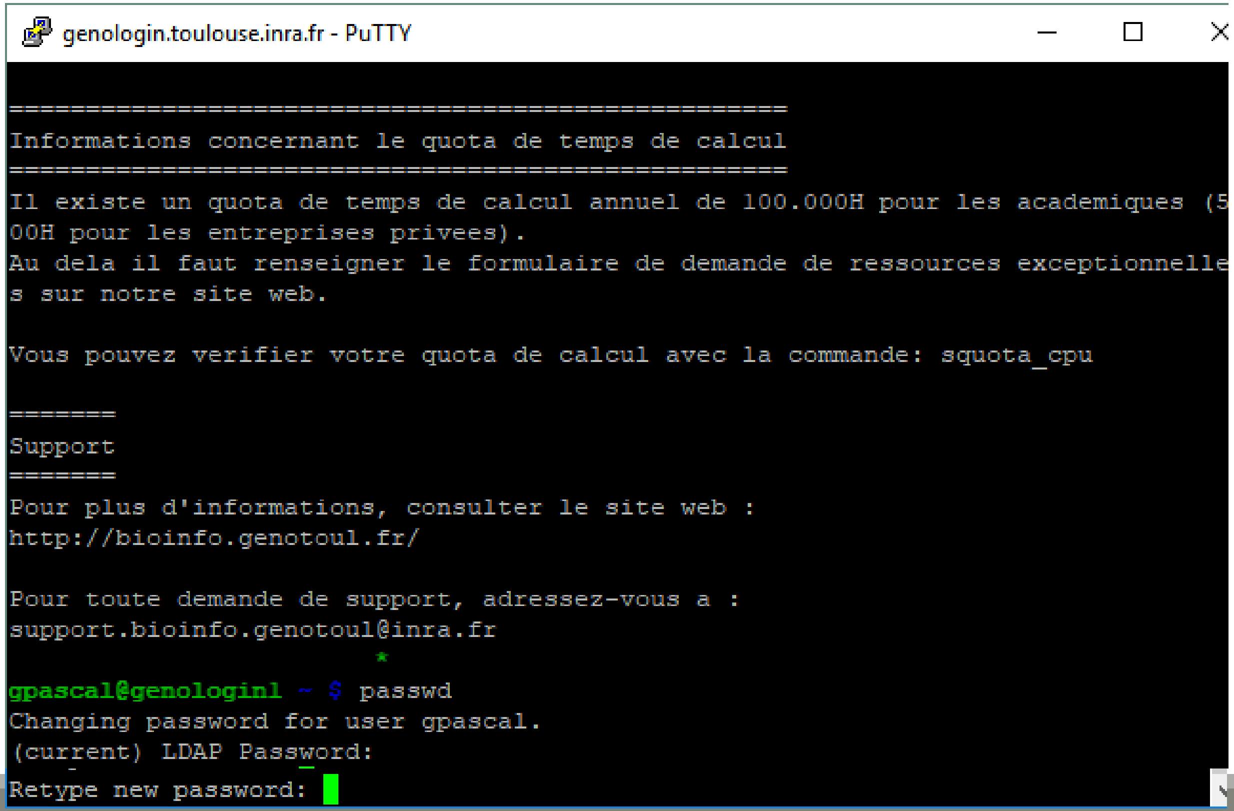
- Enter your new password (with an upper case, a number and a special character in it) and hit « Enter ».



```
====================================================
Informations concernant le quota de temps de calcul
====================================================
Il existe un quota de temps de calcul annuel de 100.000H pour les academiques (5
00H pour les entreprises privees).
Au dela il faut renseigner le formulaire de demande de ressources exceptionnelle
s sur notre site web.

Vous pouvez verifier votre quota de calcul avec la commande: squota_cpu

=======
Support
=======
Pour plus d'informations, consulter le site web :
http://bioinfo.genotoul.fr/

Pour toute demande de support, adressez-vous a :
support.bioinfo.genotoul@inra.fr
                         *
gpascal@genologin1 ~ $ passwd
Changing password for user gpascal.
(current) LDAP Password:
New LDAP password:
```

Change your password

- Enter a second time your new password and hit « Enter ». You have changed your Genotoul password, you can close PuTTY.



```
===================================================
Informations concernant le quota de temps de calcul
===================================================
Il existe un quota de temps de calcul annuel de 100.000H pour les academiques (5
00H pour les entreprises privees).
Au dela il faut renseigner le formulaire de demande de ressources exceptionnelle
s sur notre site web.

Vous pouvez verifier votre quota de calcul avec la commande: squota_cpu

=======
Support
=======
Pour plus d'informations, consulter le site web :
http://bioinfo.genotoul.fr/

Pour toute demande de support, adressez-vous a :
support.bioinfo.genotoul@inra.fr
                        *
gpascal@genologin1 ~ $ passwd
Changing password for user gpascal.
(current) LDAP Password:
Retype new password:
```

MAIN MENU

How to cite Galaxy workbench ?

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

Examples :

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

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

References

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

Ready to use Workflows

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

Galaxy News

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

E-learning

https://inra.classilio.com/Login

2 e-learning availables : Galaxy initiation and Quality sequences analysis

AVAILABLE TOOLS

RESULTS VISUALISATION AND TOOL PARAMETER WINDOW

DATASETS HISTORY

Vocabulary of Galaxy

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

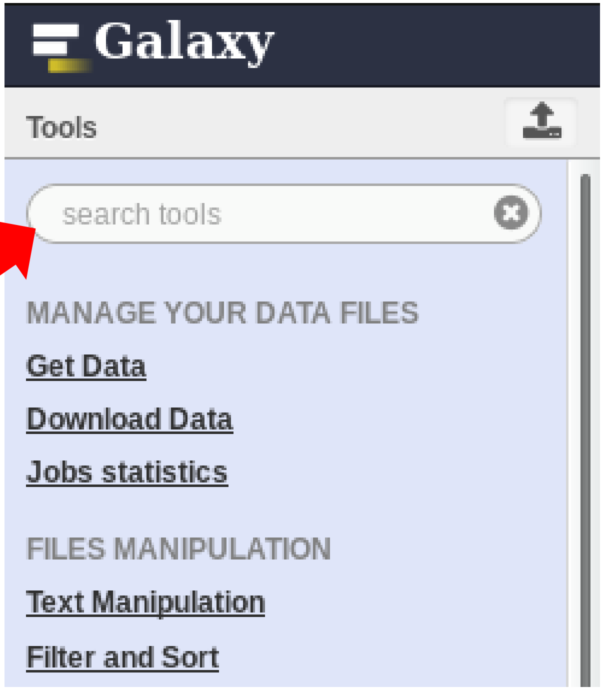
Your Turn!

DISCOVER GALAXY

Exercise

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

Exercise



Search a tool by name.

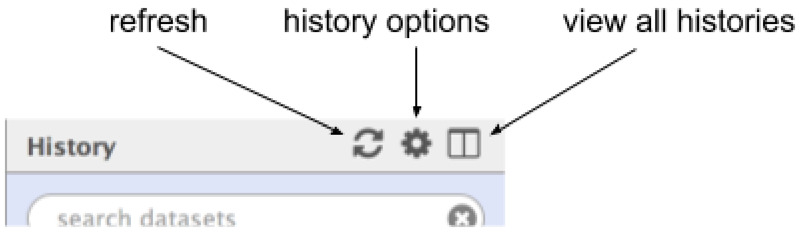
Manipulate Histories

Your Turn!

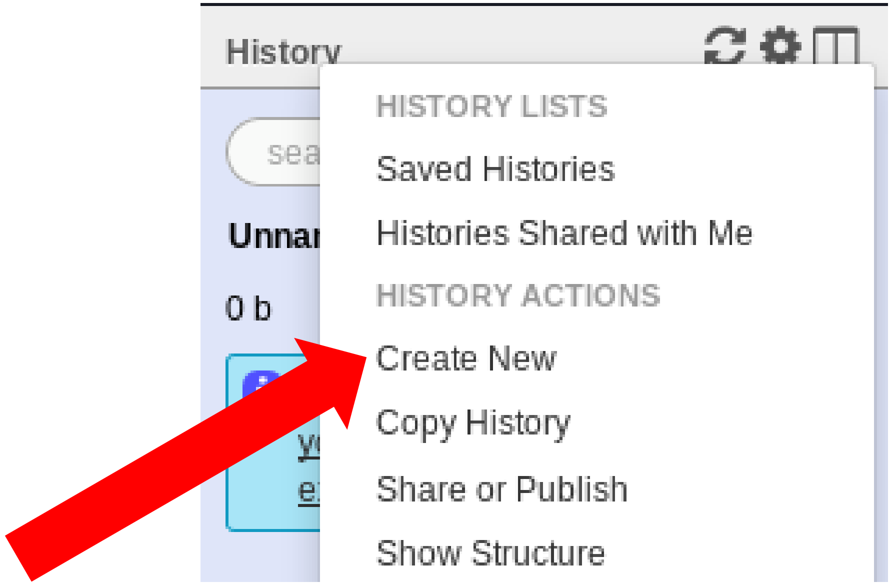
CREATE THE 4 HISTORIES NEEDED FOR THE FROGS FORMATION

To create a new history:

- Click on the wheel.



- Click on « create new ».



To rename a history:

- Click on the history name (at the top).
- Enter « multiplex ».
- Hit « Enter » to validate.

This history is empty. You can load your own data or get data from an external source

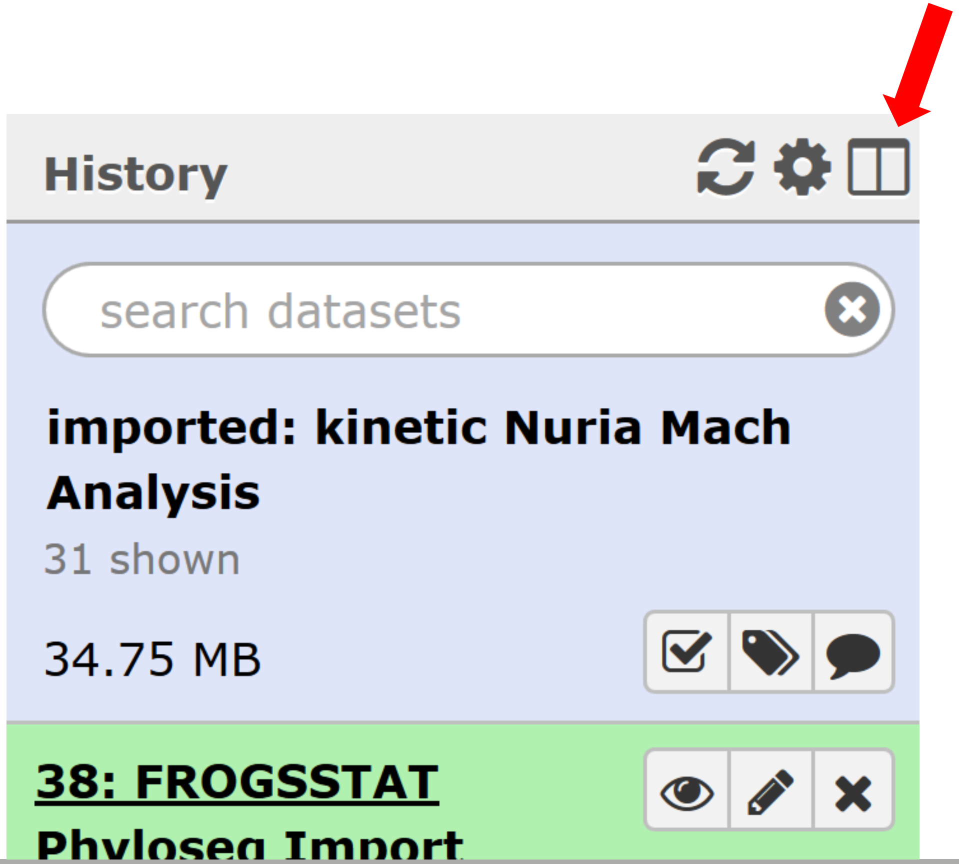
Don't use special characters or accents!

Exercise

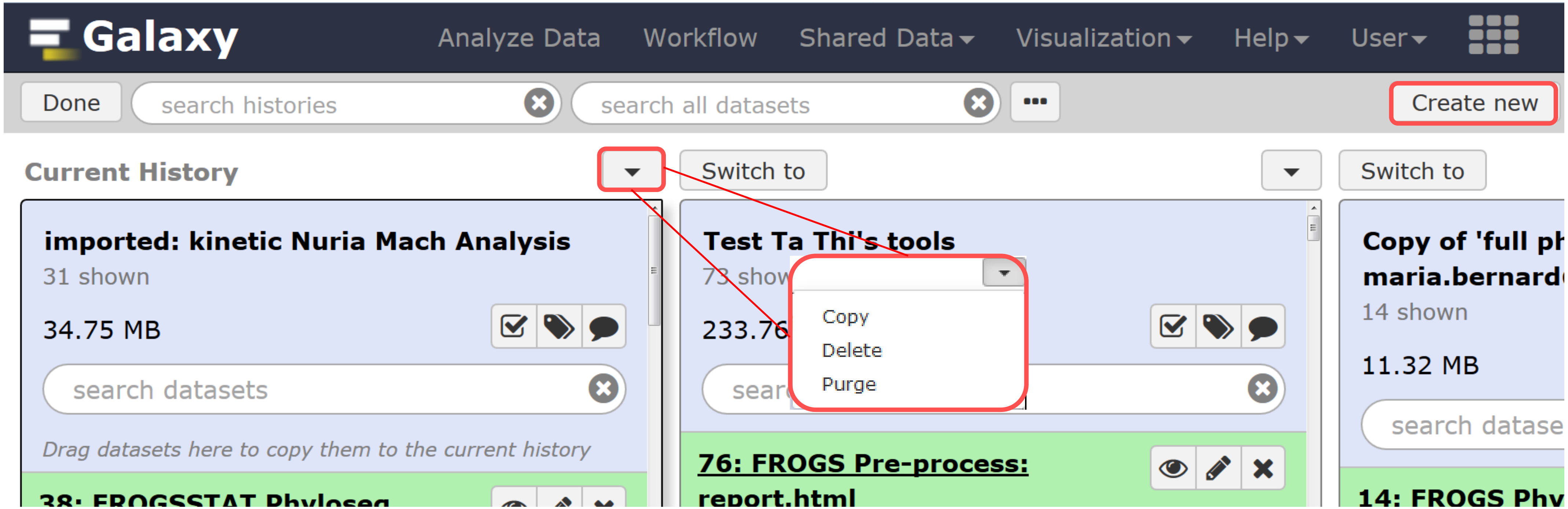
- Create histories named:
  - 454 or ITS
  - merged
  - temp
- Switch to the history named « multiplex » as current history.
- Go back to the main interface.

How to list all histories?

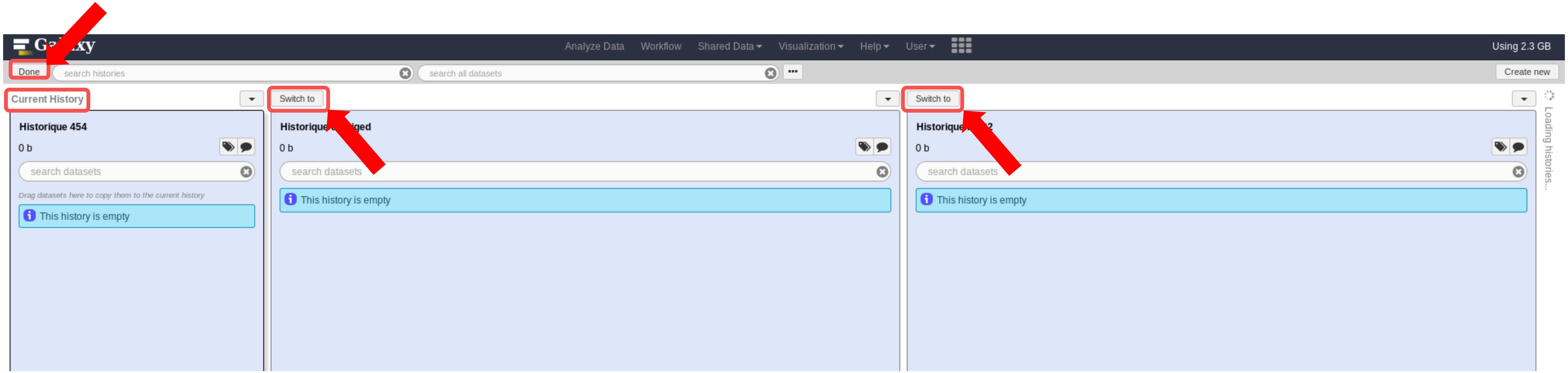
- To view all histories, click on this icon.



Explore the « View all histories » section



Switch current history



Drag datasets here to copy them to the current history

This history is empty

This history is empty

This history is empty

- Switch to the history named « multiplex » as current history.
- Click on “Done” to go back to the main interface.

Data import

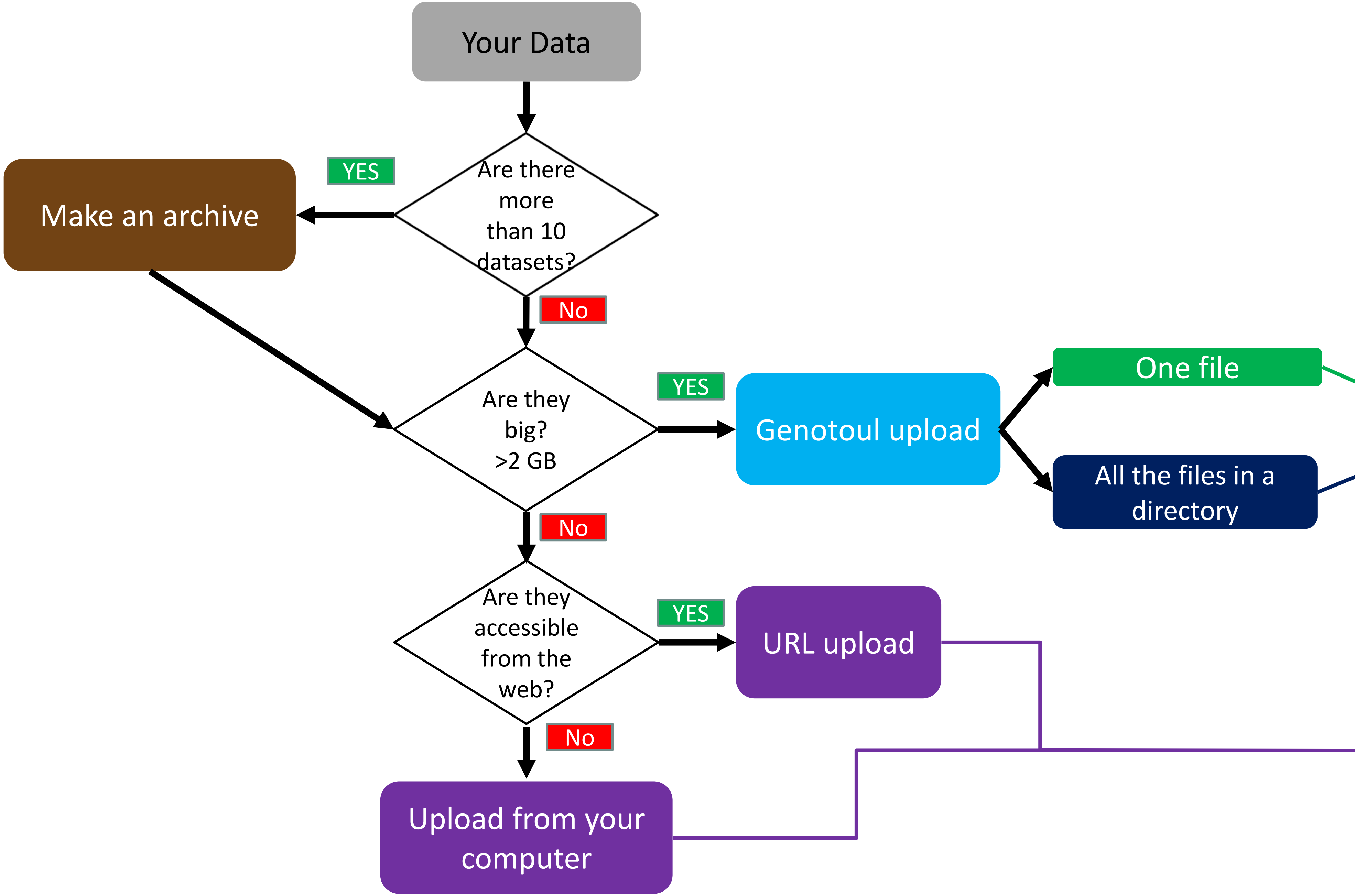
How to import your data to Galaxy

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

  - By URL

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

How to choose your upload method?



Your Data

Are there more than 10 datasets?

YES

Make an archive

No

Are they big? >2 GB

YES

Genotoul upload

One file

All the files in a directory

No

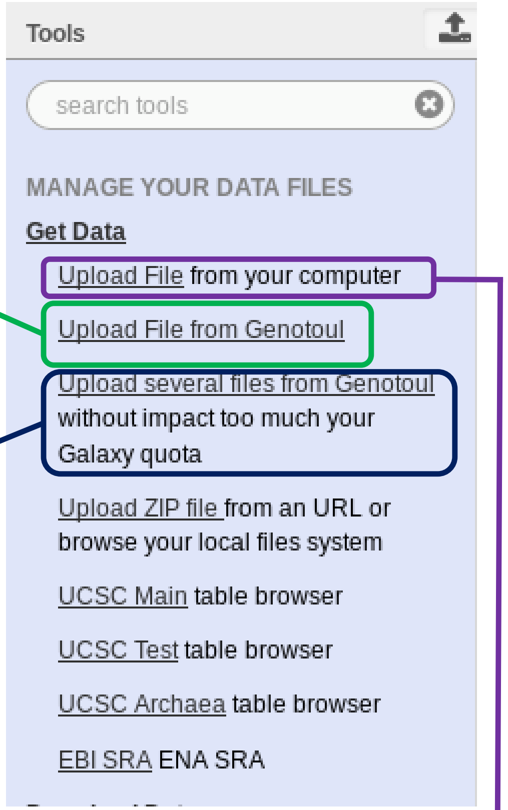
Are they accessible from the web?

YES

URL upload

No

Upload from your computer



MANAGE YOUR DATA FILES

Get Data

Upload File from your computer

Upload File from Genotoul

Upload several files from Genotoul without impact too much your Galaxy quota

Upload ZIP file from an URL or browse your local files system

UCSC Main table browser

UCSC Test table browser

UCSC Archaea table browser

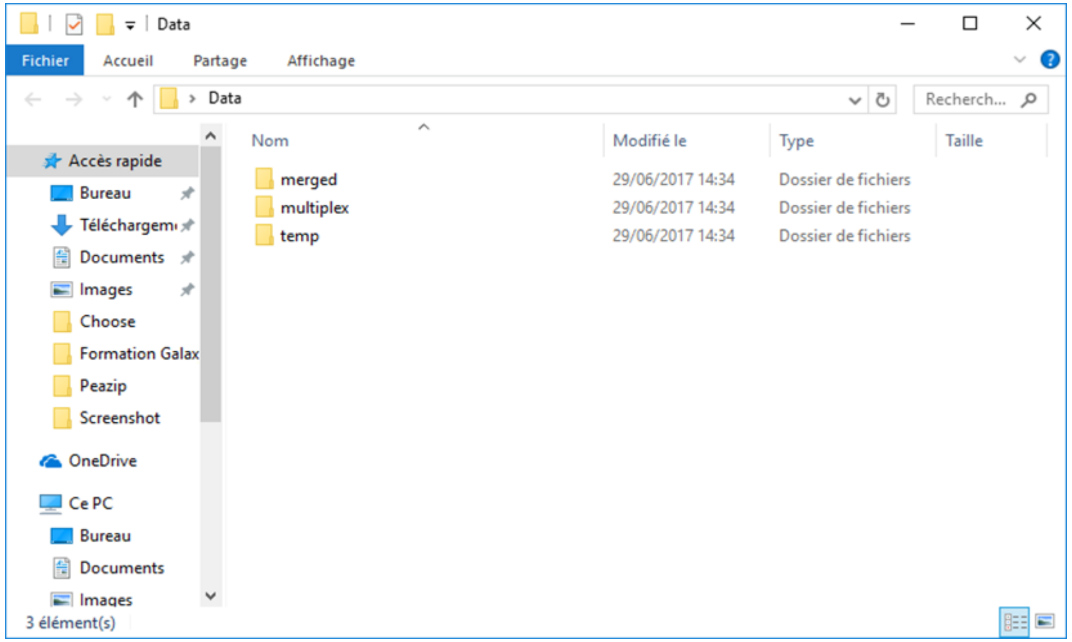
EBI SRA ENA SRA

Your Turn!

PREPARE FILES

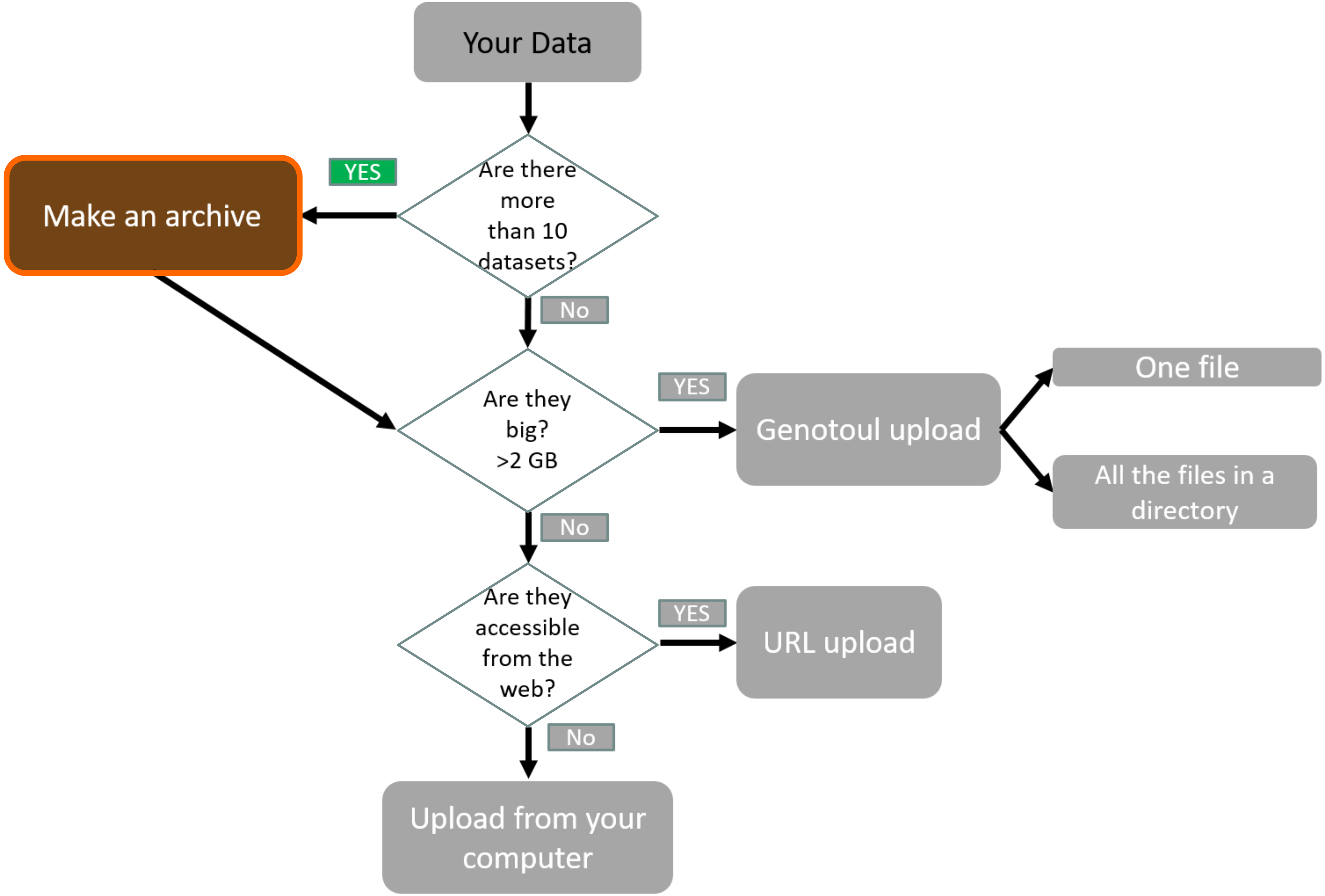
Formation data

- All the formation data are in the Data folder on your desktop.



Nom	Modifié le	Type	Taille
merged	29/06/2017 14:34	Dossier de fichiers	
multiplex	29/06/2017 14:34	Dossier de fichiers	
temp	29/06/2017 14:34	Dossier de fichiers	

- If you don't have this directory on your desktop, ask us.
- You can find all the formation data online on: http://genoweb.toulouse.inra.fr/~formation/15_FROGS/FROGS_ini/DATA/Dataset/



Your Data

Are there more than 10 datasets?

YES

Make an archive

No

Are they big? >2 GB

YES

Genotoul upload

One file

All the files in a directory

No

Are they accessible from the web?

YES

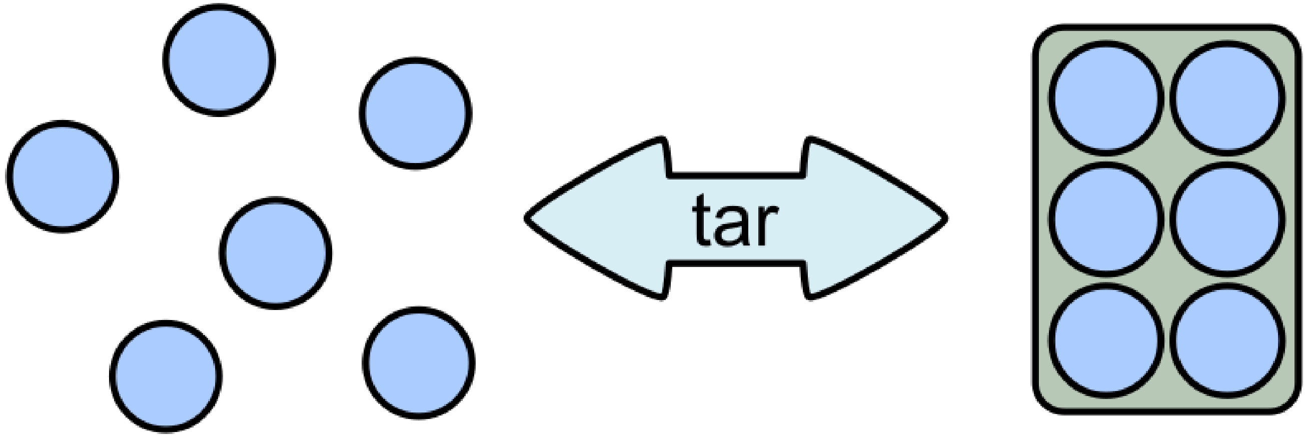
URL upload

No

Upload from your computer

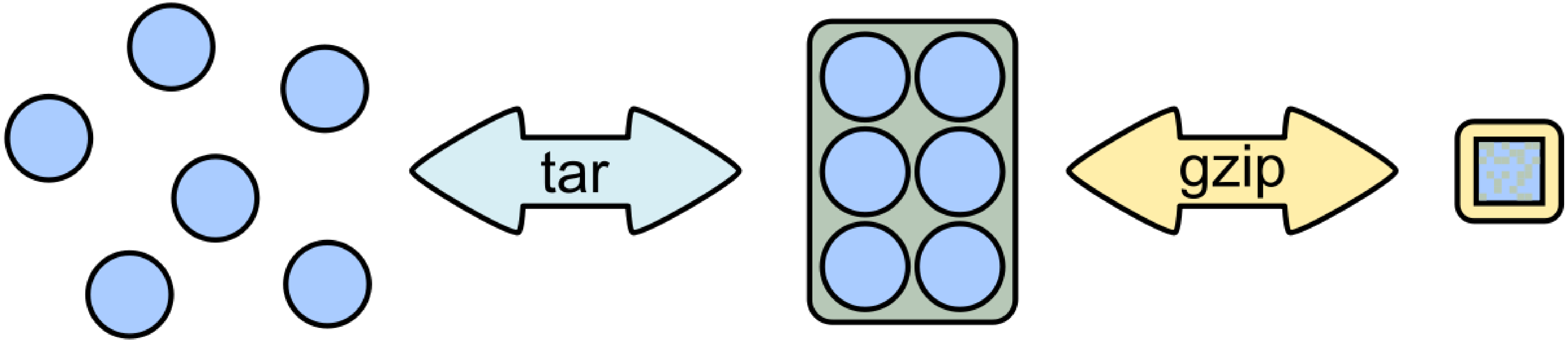
Create a Tar file

- Uploading multiple local files is time consuming.
- Solution: put all your files in an archive!
- What is a tar file?



Create a Tar.gz

- Moreover, we can compress the archive to free up space.



Your Turn!

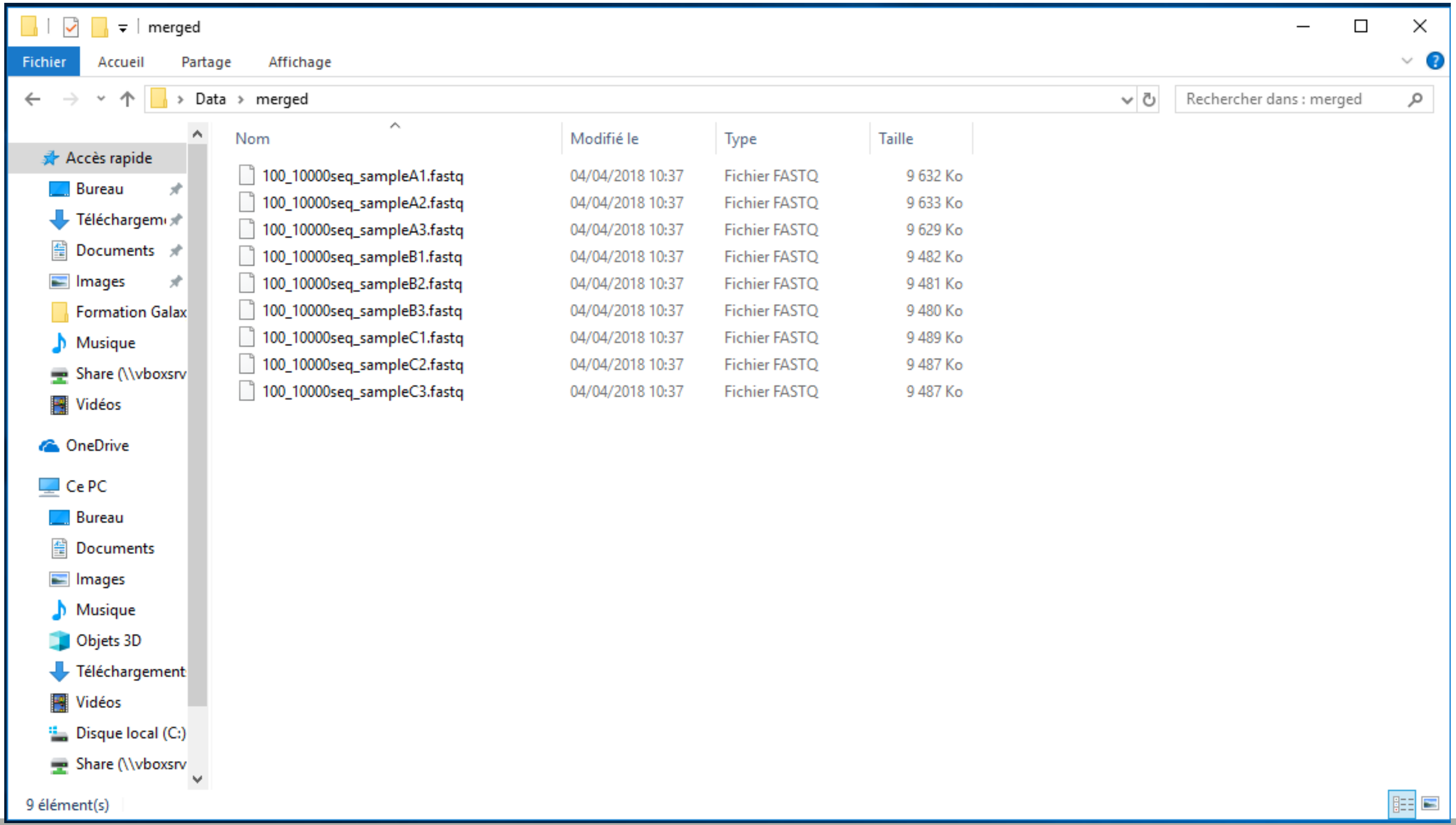
CREATE AN ARCHIVE WITH THE FILES IN MERGED

7-Zip

- 7-Zip is a software called a file archiver.
- Can archive and compress files.
- Open source
- At the lab, you could download it at: http://www.7-zip.org/

Create a Tar file with 7-Zip

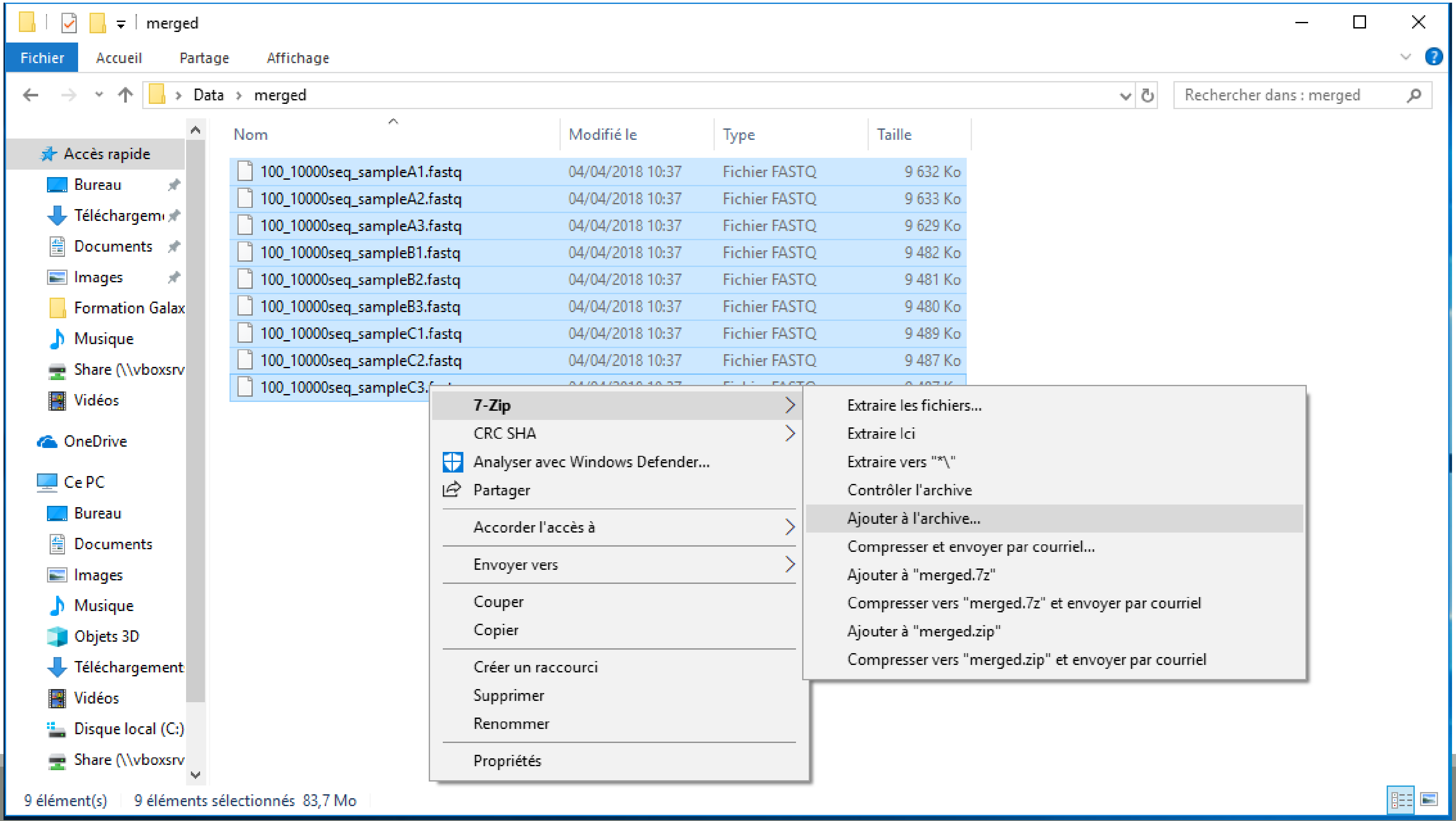
- Browse into the Data directory and next into the merged directory.



Nom	Modifié le	Type	Taille
100_10000seq_sampleA1.fastq	04/04/2018 10:37	Fichier FASTQ	9 632 Ko
100_10000seq_sampleA2.fastq	04/04/2018 10:37	Fichier FASTQ	9 633 Ko
100_10000seq_sampleA3.fastq	04/04/2018 10:37	Fichier FASTQ	9 629 Ko
100_10000seq_sampleB1.fastq	04/04/2018 10:37	Fichier FASTQ	9 482 Ko
100_10000seq_sampleB2.fastq	04/04/2018 10:37	Fichier FASTQ	9 481 Ko
100_10000seq_sampleB3.fastq	04/04/2018 10:37	Fichier FASTQ	9 480 Ko
100_10000seq_sampleC1.fastq	04/04/2018 10:37	Fichier FASTQ	9 489 Ko
100_10000seq_sampleC2.fastq	04/04/2018 10:37	Fichier FASTQ	9 487 Ko
100_10000seq_sampleC3.fastq	04/04/2018 10:37	Fichier FASTQ	9 487 Ko

Create a Tar file with 7-Zip

- Select all files and right click on it > 7-Zip > Add to archive . . .



Nom	Modifié le	Type	Taille
100_10000seq_sampleA1.fastq	04/04/2018 10:37	Fichier FASTQ	9 632 Ko
100_10000seq_sampleA2.fastq	04/04/2018 10:37	Fichier FASTQ	9 633 Ko
100_10000seq_sampleA3.fastq	04/04/2018 10:37	Fichier FASTQ	9 629 Ko
100_10000seq_sampleB1.fastq	04/04/2018 10:37	Fichier FASTQ	9 482 Ko
100_10000seq_sampleB2.fastq	04/04/2018 10:37	Fichier FASTQ	9 481 Ko
100_10000seq_sampleB3.fastq	04/04/2018 10:37	Fichier FASTQ	9 480 Ko
100_10000seq_sampleC1.fastq	04/04/2018 10:37	Fichier FASTQ	9 489 Ko
100_10000seq_sampleC2.fastq	04/04/2018 10:37	Fichier FASTQ	9 487 Ko
100_10000seq_sampleC3.f			

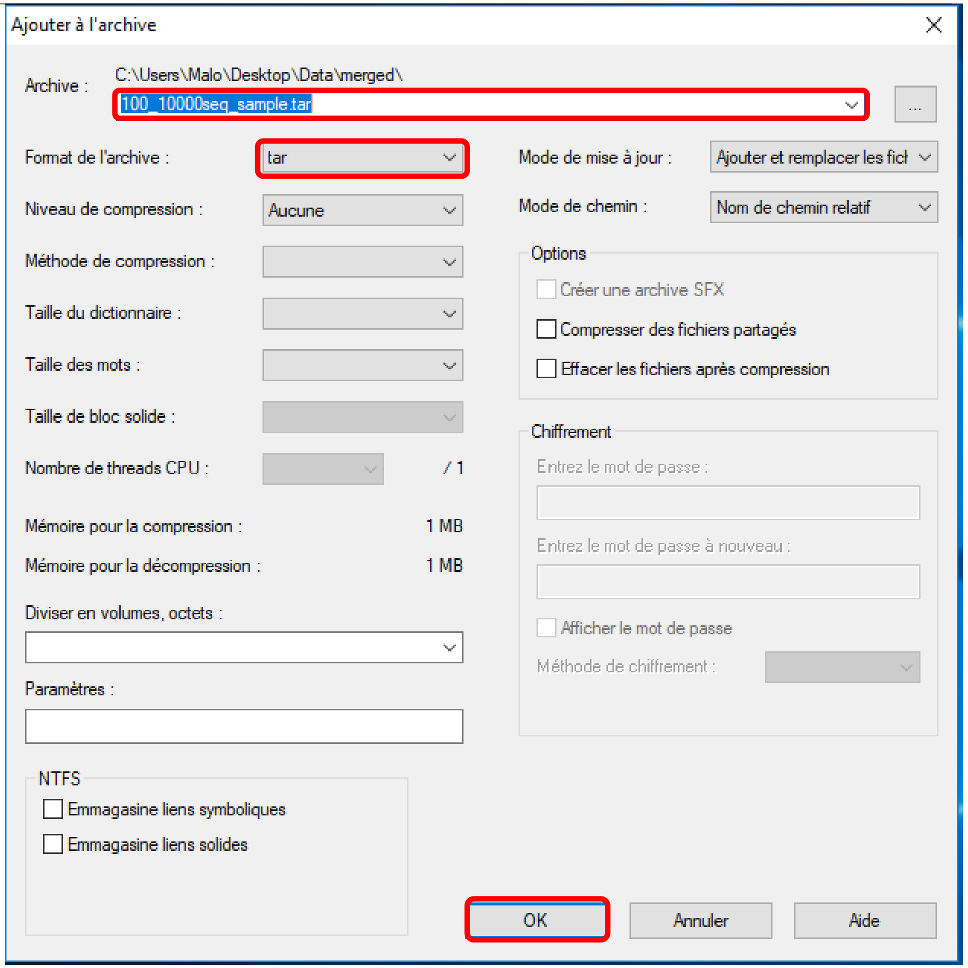
9 élément(s) 9 éléments sélectionnés 83,7 Mo

Create a Tar file with 7-Zip

- Change the name of the archive to:

« 100_10000seq_sample.tar »

- Choose « tar » as archive format .
- Click on OK .



Ajouter à l'archive

Archive : C:\Users\Malo\Desktop\Data\merged\
100_10000seq_sample.tar

Format de l'archive : tar

Niveau de compression : Aucune

Méthode de compression :

Taille du dictionnaire :

Taille des mots :

Taille de bloc solide :

Nombre de threads CPU : / 1

Mémoire pour la compression : 1 MB

Mémoire pour la décompression : 1 MB

Diviser en volumes, octets :

Paramètres :

Mode de mise à jour : Ajouter et remplacer les fich

Mode de chemin : Nom de chemin relatif

Options

☐ Créer une archive SFX

☐ Compresser des fichiers partagés

☐ Effacer les fichiers après compression

Chiffrement

Entrez le mot de passe :

Entrez le mot de passe à nouveau :

☐ Afficher le mot de passe

Méthode de chiffrement :

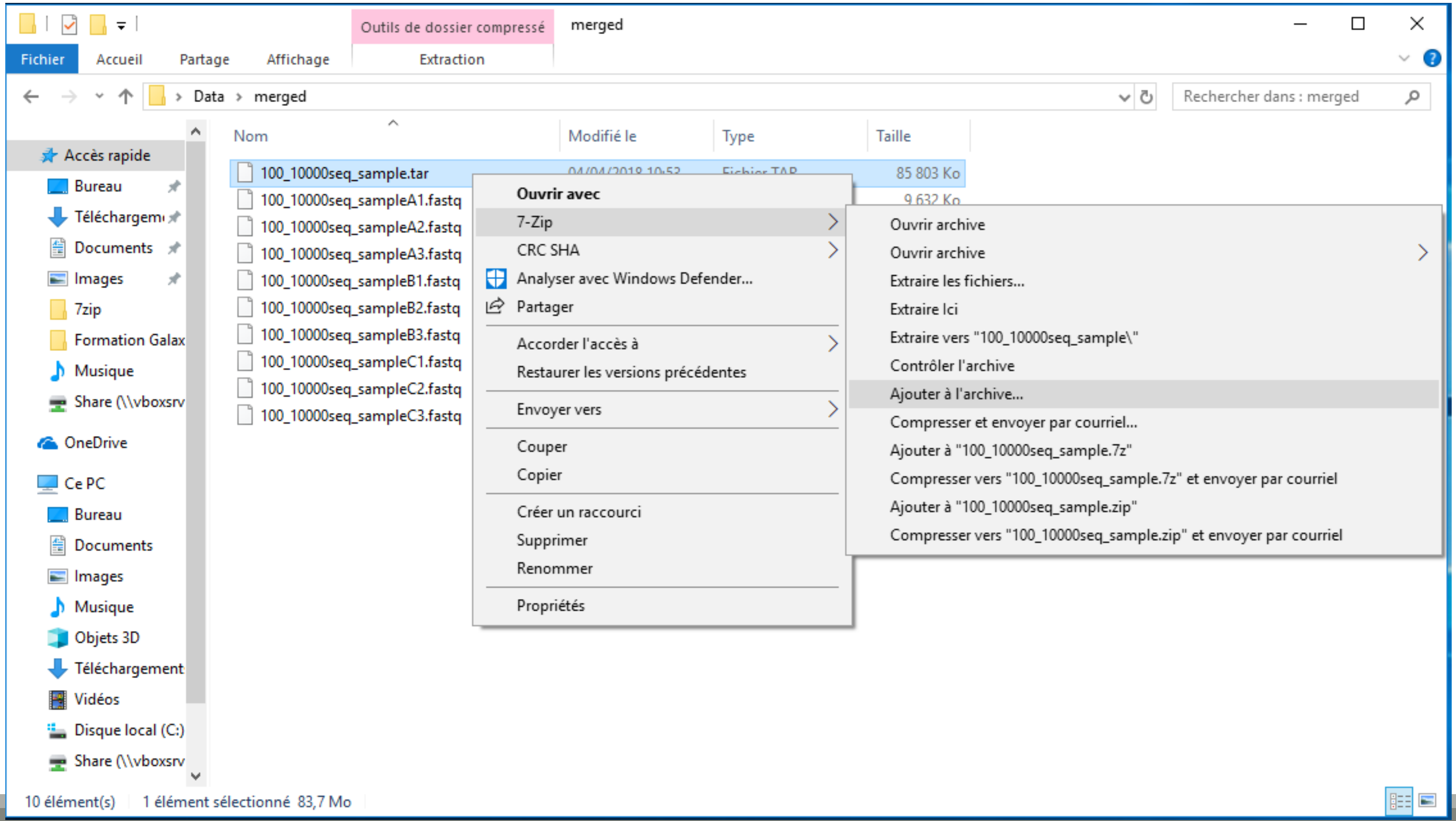
NTFS

☐ Emmagasine liens symboliques

☐ Emmagasine liens solides

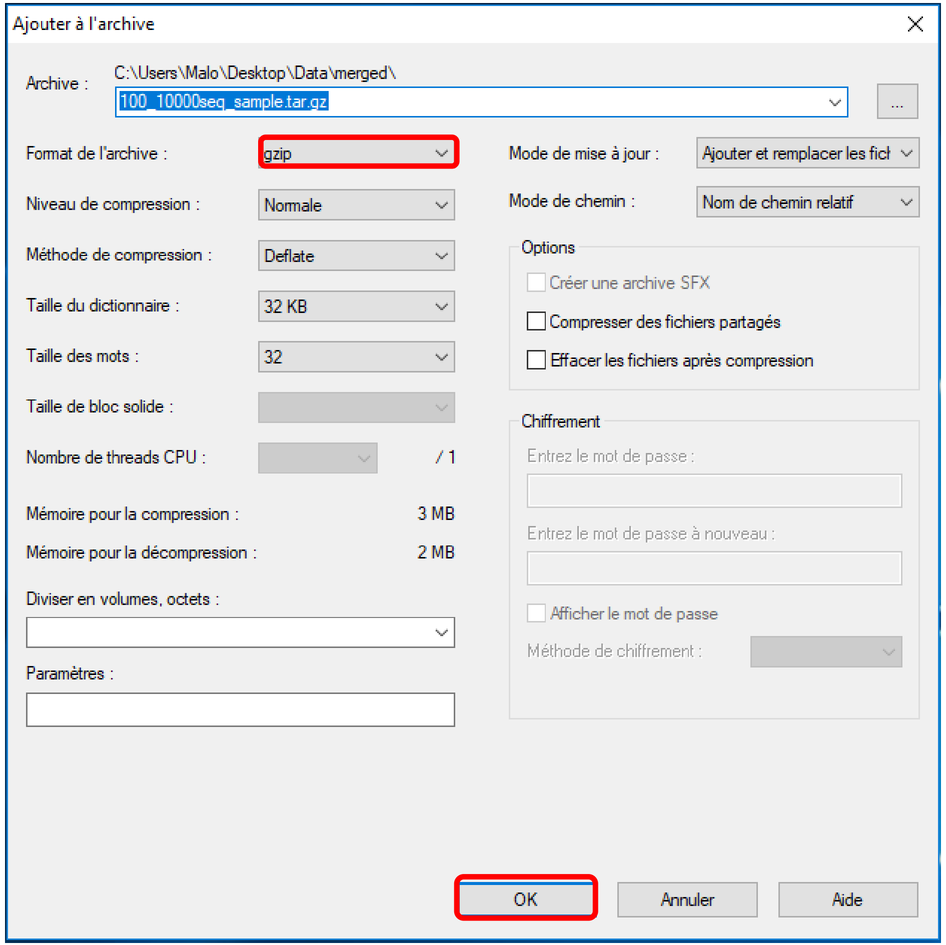
Create a Tar file with 7-Zip

- Select the newly created archive and right click on it > 7-Zip > Add to archive . . .



Create a Tar file with 7-Zip

- Choose « gzip » as archive format .
- Click on OK .



Ajouter à l'archive

Archive : C:\Users\Malo\Desktop\Data\merged\100_10000seq_sample.tar.gz

Format de l'archive : gzip

Niveau de compression : Normale

Méthode de compression : Deflate

Taille du dictionnaire : 32 KB

Taille des mots : 32

Taille de bloc solide :

Nombre de threads CPU : / 1

Mémoire pour la compression : 3 MB

Mémoire pour la décompression : 2 MB

Diviser en volumes, octets :

Paramètres :

Mode de mise à jour : Ajouter et remplacer les fich

Mode de chemin : Nom de chemin relatif

Options

☐ Créer une archive SFX

☐ Compresser des fichiers partagés

☐ Effacer les fichiers après compression

Chiffrement

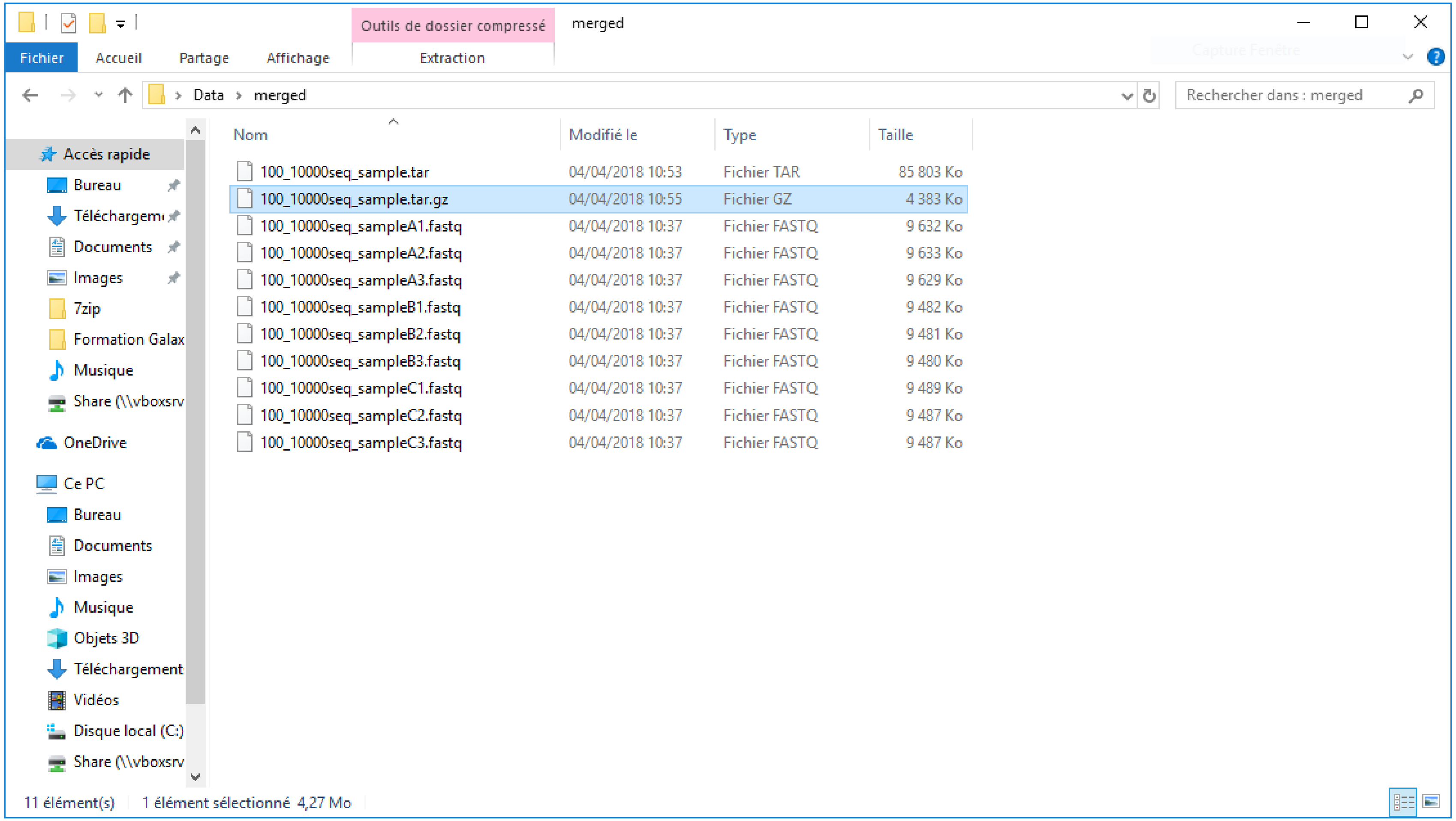
Entrez le mot de passe :

Entrez le mot de passe à nouveau :

☐ Afficher le mot de passe

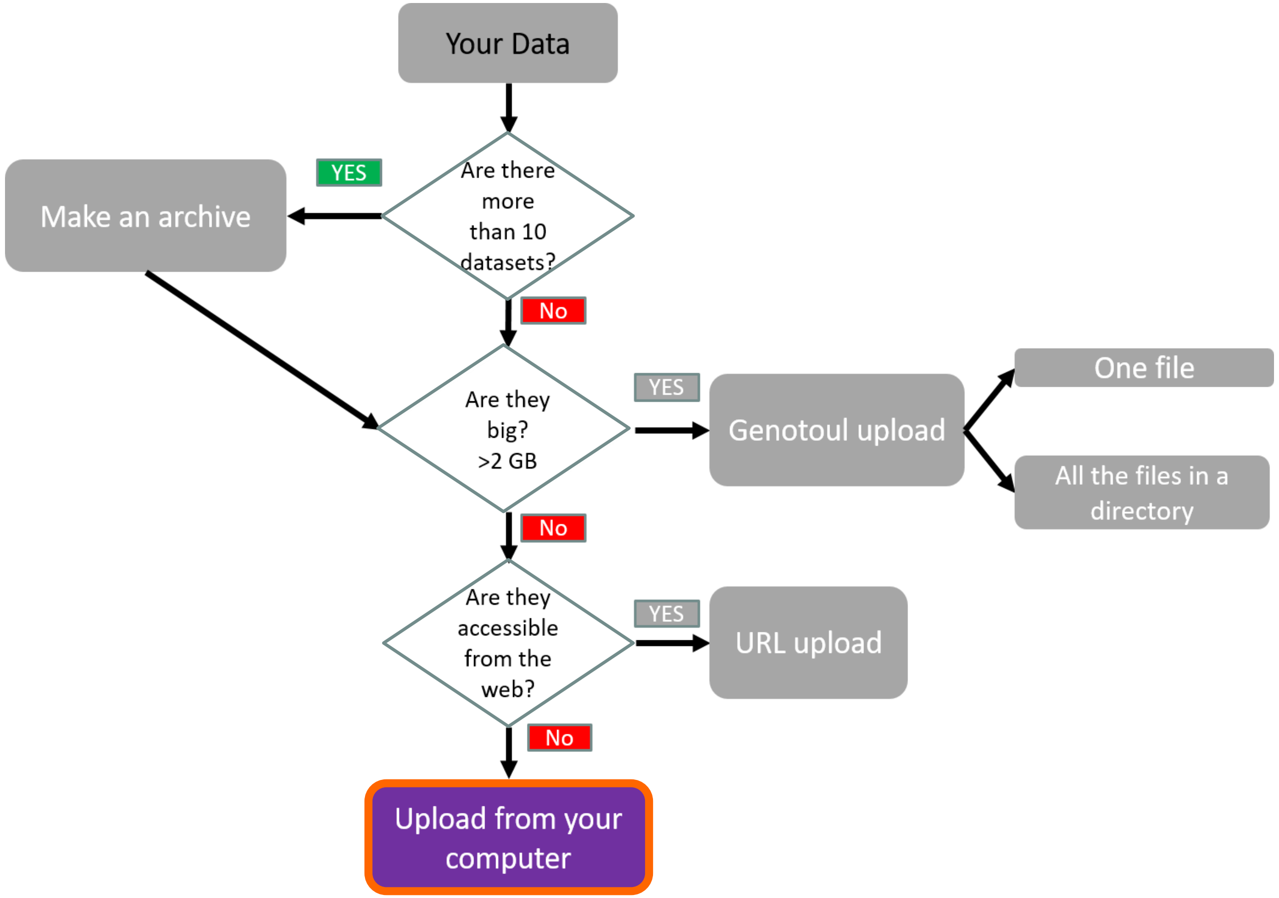
Méthode de chiffrement :

Create a Tar file with 7-Zip



Nom	Modifié le	Type	Taille
100_10000seq_sample.tar	04/04/2018 10:53	Fichier TAR	85 803 Ko
100_10000seq_sample.tar.gz	04/04/2018 10:55	Fichier GZ	4 383 Ko
100_10000seq_sampleA1.fastq	04/04/2018 10:37	Fichier FASTQ	9 632 Ko
100_10000seq_sampleA2.fastq	04/04/2018 10:37	Fichier FASTQ	9 633 Ko
100_10000seq_sampleA3.fastq	04/04/2018 10:37	Fichier FASTQ	9 629 Ko
100_10000seq_sampleB1.fastq	04/04/2018 10:37	Fichier FASTQ	9 482 Ko
100_10000seq_sampleB2.fastq	04/04/2018 10:37	Fichier FASTQ	9 481 Ko
100_10000seq_sampleB3.fastq	04/04/2018 10:37	Fichier FASTQ	9 480 Ko
100_10000seq_sampleC1.fastq	04/04/2018 10:37	Fichier FASTQ	9 489 Ko
100_10000seq_sampleC2.fastq	04/04/2018 10:37	Fichier FASTQ	9 487 Ko
100_10000seq_sampleC3.fastq	04/04/2018 10:37	Fichier FASTQ	9 487 Ko

11 élément(s) 1 élément sélectionné 4,27 Mo


Your Data

Are there more than 10 datasets?

YES

Make an archive

No

Are they big? >2 GB

YES

Genotoul upload

One file

All the files in a directory

No

Are they accessible from the web?

YES

URL upload

No

Upload from your computer

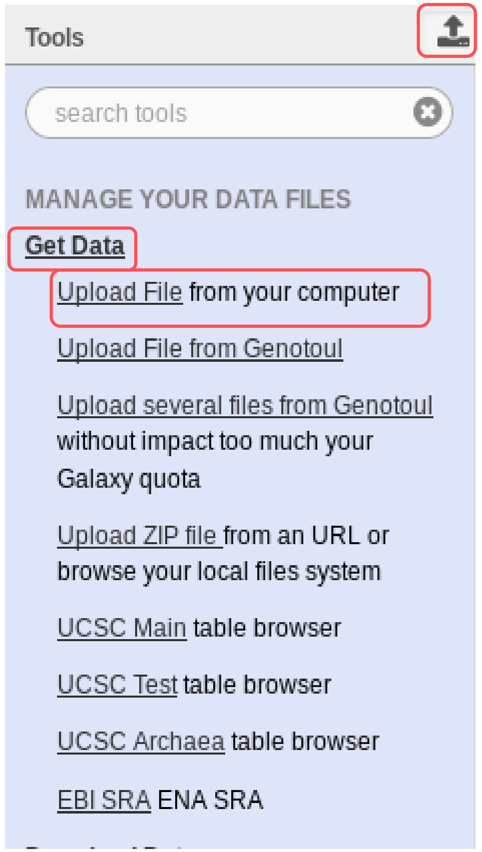
Your Turn!

UPLOAD FILES FROM YOUR COMPUTER

Upload local files

- In Galaxy, your current history must be multiplex.
- Click on the « Get Data » tool.
- Upload file from your computer.

⚠ For files smaller than 2 GB



Tools

search tools

MANAGE YOUR DATA FILES

Get Data

Upload File from your computer

Upload File from Genotoul

Upload several files from Genotoul without impact too much your Galaxy quota

Upload ZIP file from an URL or browse your local files system

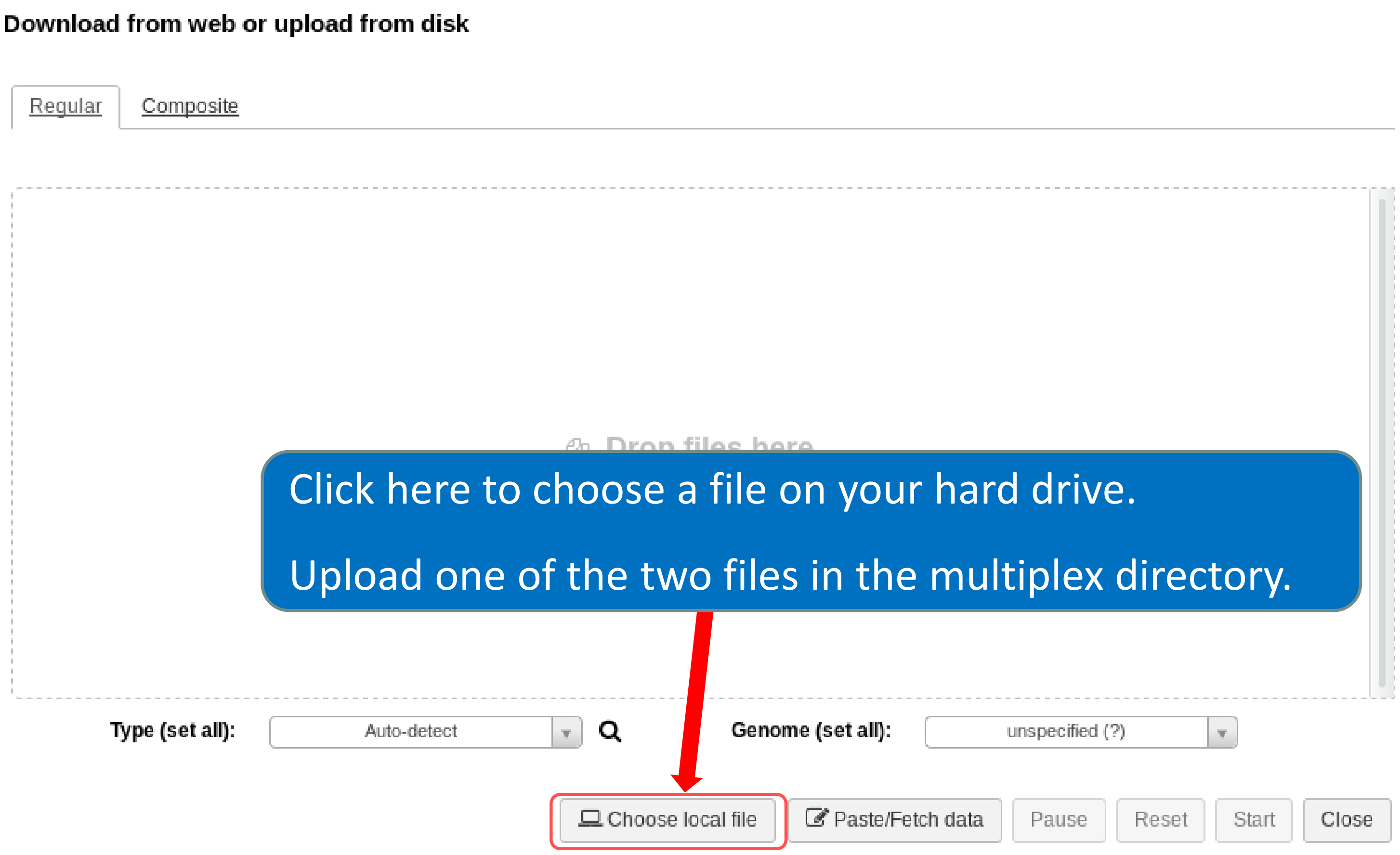
UCSC Main table browser

UCSC Test table browser

UCSC Archaea table browser

EBI SRA ENA SRA

Upload local files



Click here to choose a file on your hard drive.

Upload one of the two files in the multiplex directory.

Upload local files

The datatype of a file indicates the content of it.

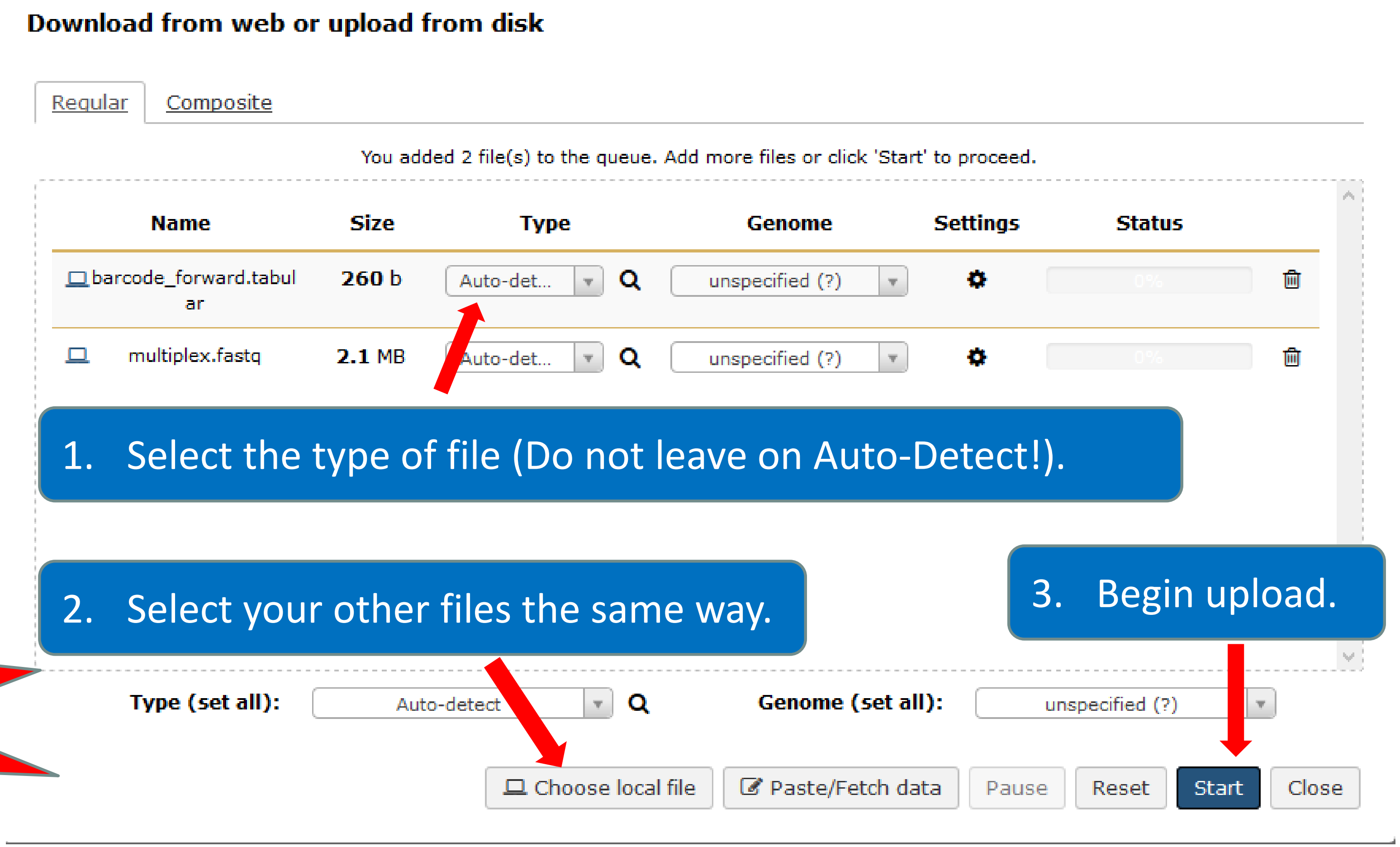
.fasta = fasta

.fastq = fastq

.tar = tar

Select the type corresponding to your file.

Do no not trust blindly the auto-detect!



Download from web or upload from disk

You added 2 file(s) to the queue. Add more files or click 'Start' to proceed.

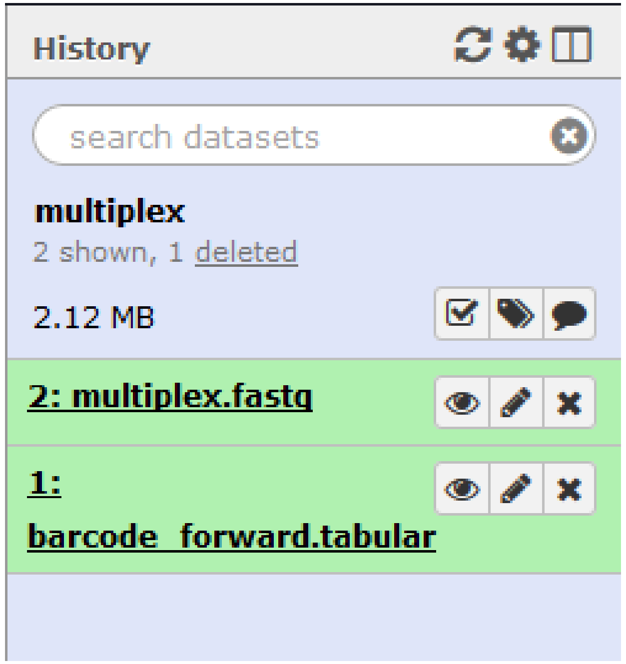
Name	Size	Type	Genome	Settings	Status
barcode_forward.tabular	260 b	Auto-det...	unspecified (?)		0%
multiplex.fastq	2.1 MB	Auto-det...	unspecified (?)		0%

Type (set all): Auto-detect

Genome (set all): unspecified (?)

1. Select the type of file (Do not leave on Auto-Detect!).
2. Select your other files the same way.
3. Begin upload.

Upload local files



History

search datasets

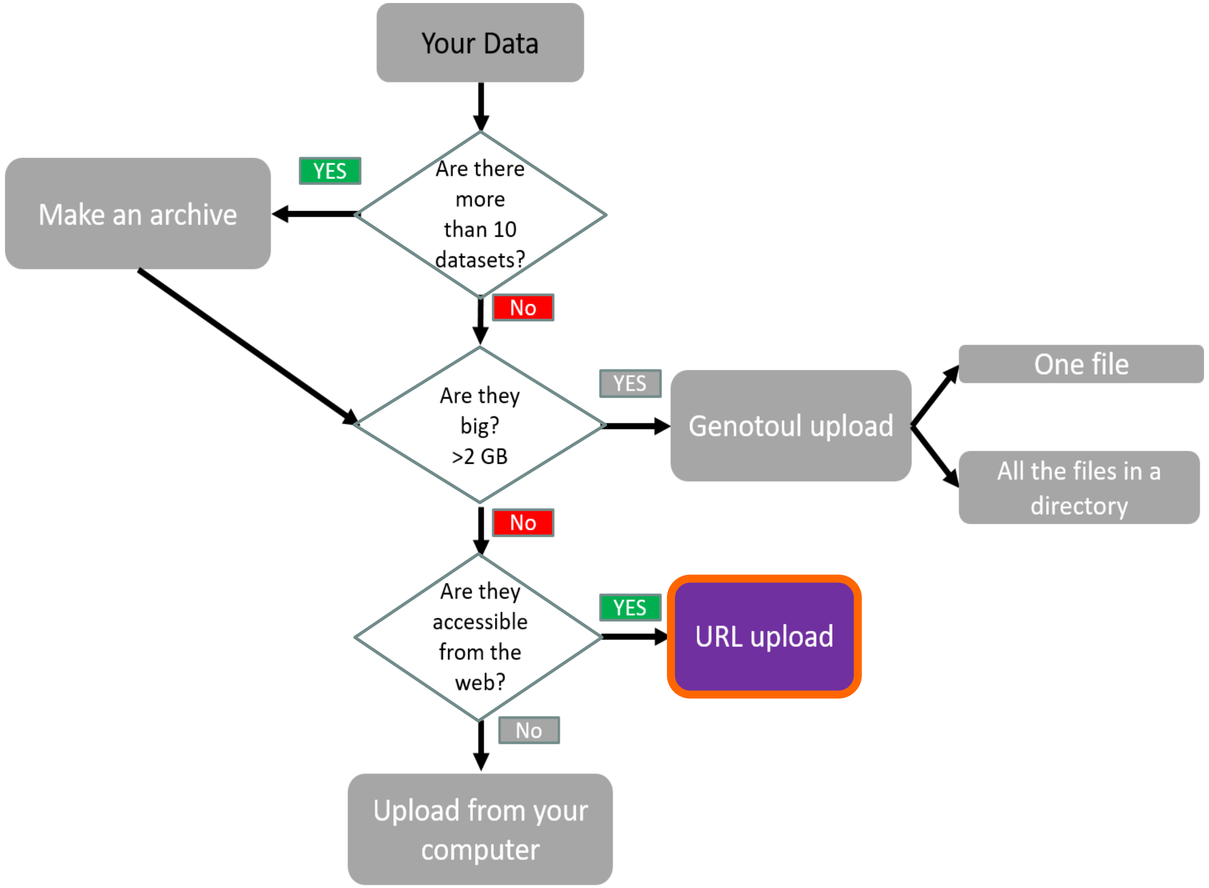
multiplex

2 shown, 1 deleted

2.12 MB

2: multiplex.fastq

1: barcode_forward.tabular



Your Turn!

UPLOAD FILE FROM AN URL

Upload file from URL

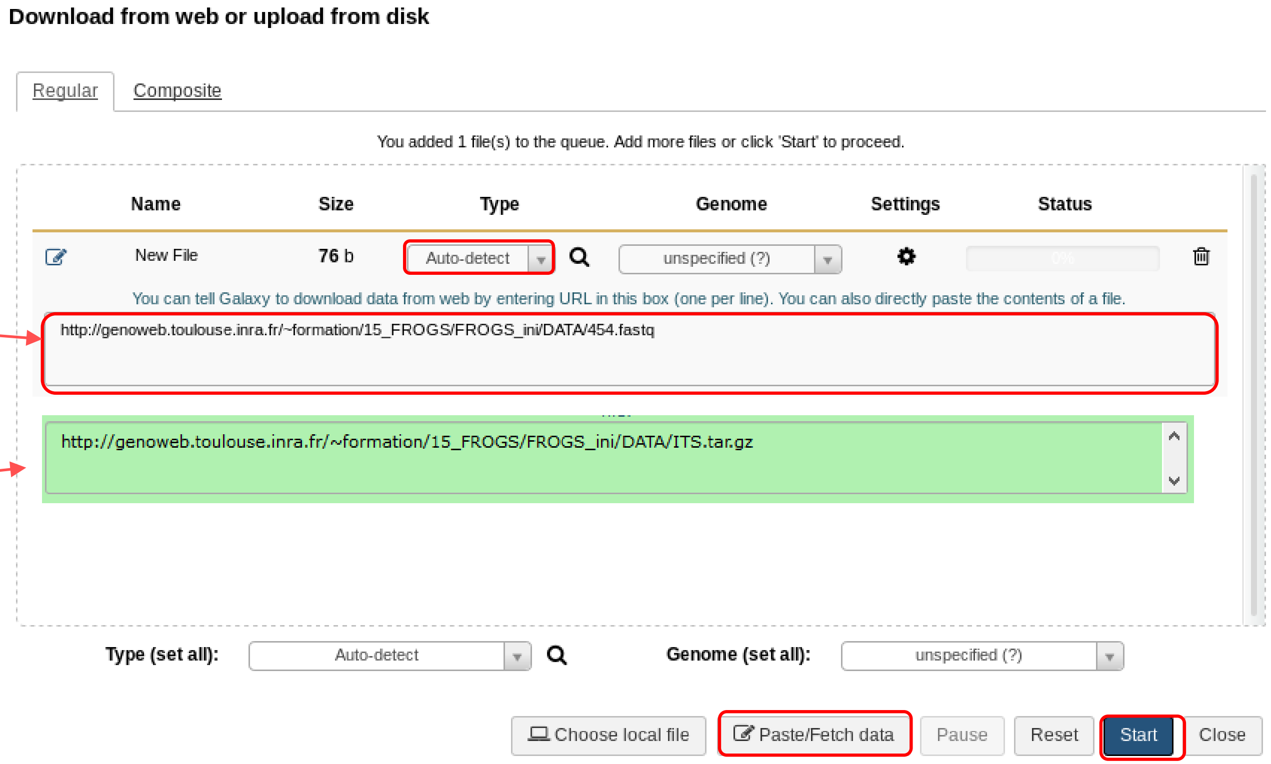
1. Switch to 454 or ITS history as current history.
2. Go to **Get Data** > **Upload File from your computer**
3. Click on **Paste/Fetch Data**
4. Copy the address of the file:

http://genoweb.toulouse.inra.fr/~formation/15_FROGS/FROGS_ini/DATA/454.fastq

Or

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

1. Change the type!
2. Click on Start.
3. You can put one address per line for multiple uploads.



Download from web or upload from disk

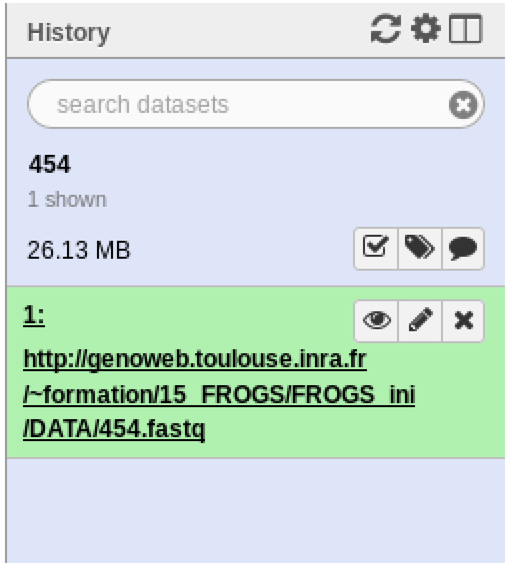
You added 1 file(s) to the queue. Add more files or click 'Start' to proceed.

You can tell Galaxy to download data from web by entering URL in this box (one per line). You can also directly paste the contents of a file.

http://genoweb.toulouse.inra.fr/~formation/15_FROGS/FROGS_ini/DATA/454.fastq

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

Upload file from URL



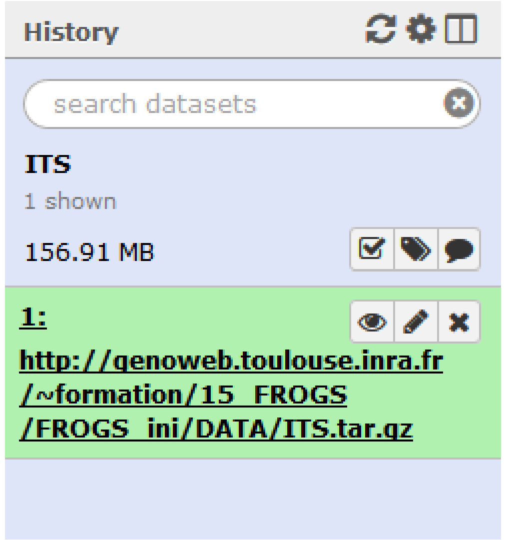
454

1 shown

26.13 MB

1: http://genoweb.toulouse.inra.fr/~formation/15_FROGS/FROGS_ini/DATA/454.fastq

OR

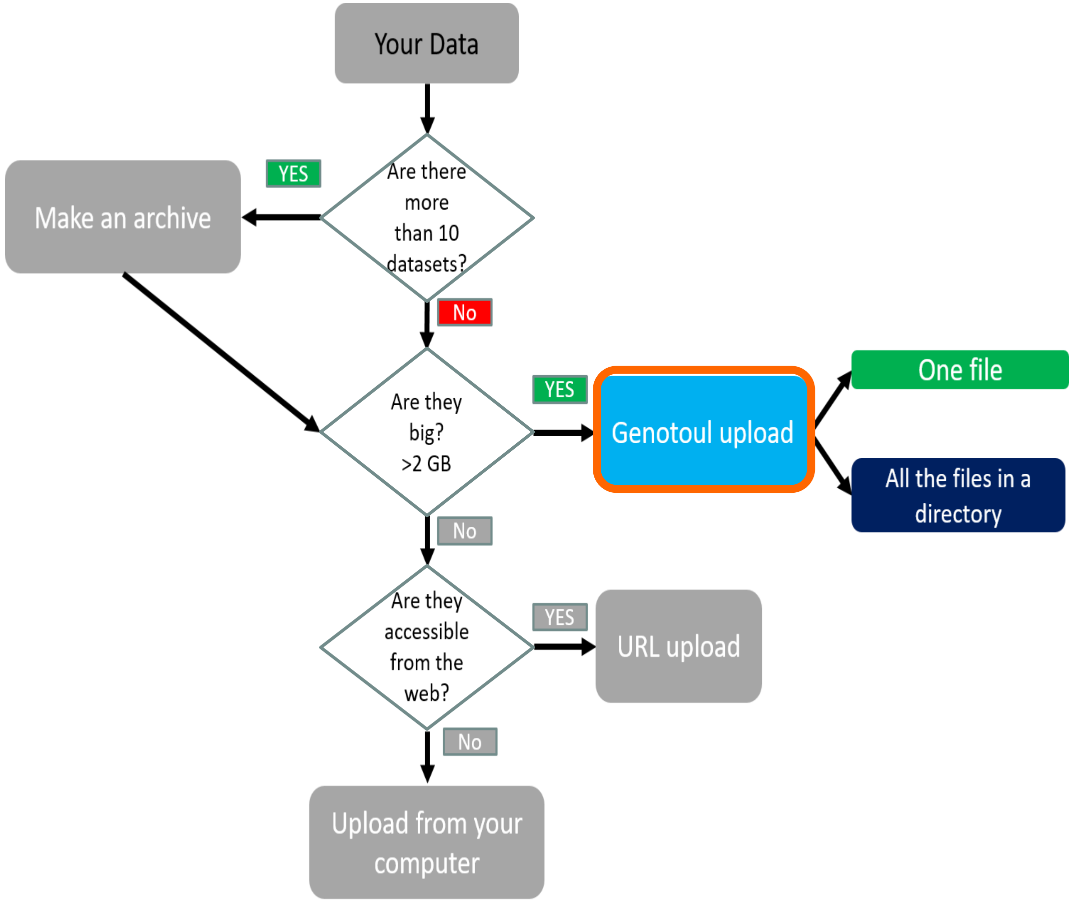


ITS

1 shown

156.91 MB

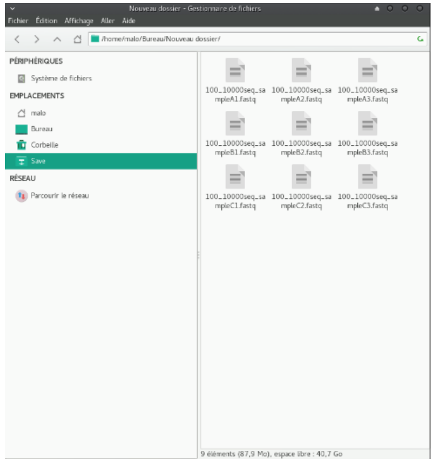
1: http://genoweb.toulouse.inra.fr/~formation/15_FROGS/FROGS_ini/DATA/ITS.tar.gz



Your Turn!

UPLOAD FILES TO GENOTOUL AND LINK IT TO GALAXY

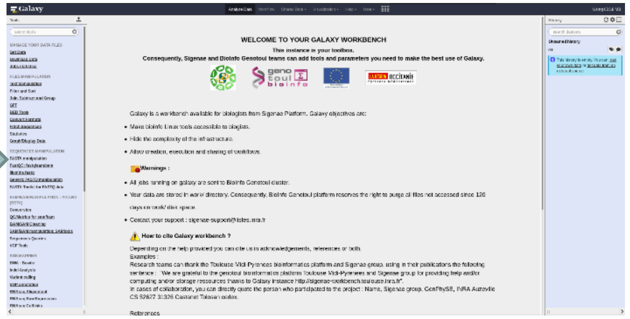
Objectives



/home/malo/Bureau/Nouveau dossier/

100_10000seq_sampleA1.fastq
100_10000seq_sampleA2.fastq
100_10000seq_sampleA3.fastq
100_10000seq_sampleB1.fastq
100_10000seq_sampleB2.fastq
100_10000seq_sampleB3.fastq
100_10000seq_sampleC1.fastq
100_10000seq_sampleC2.fastq
100_10000seq_sampleC3.fastq

9 éléments (87,9 Mo), espace libre : 40,7 Go



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 ?

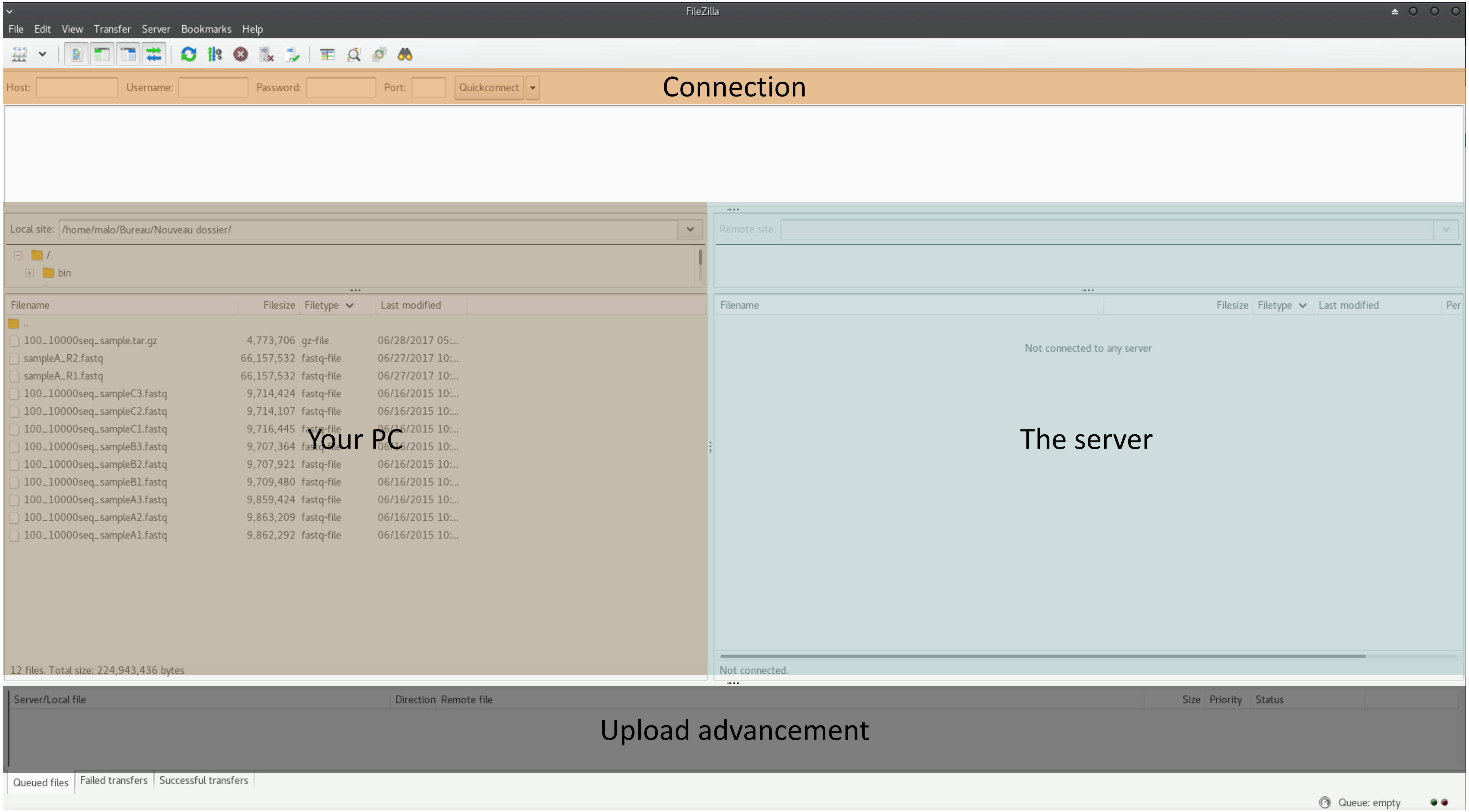
Depending on the help provided you can cite us in acknowledgements, references or both.
Examples :
Research teams can thank the Toulouse Midi-Pyrenees bioinformatics platform and Sigenae group, using in their publications the following sentence : "We are grateful to the genotoul bioinformatics platform 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

Transfer your files to your Genotoul account and link the file to Galaxy.

Preparation

- Open the data directory on your desktop.
- Launch Filezilla.
  - Filezilla is a FTP client *i.e.* can transfer files to a distant server.
  - You can download it on : https://filezilla-project.org/



Local site: /home/malo/Bureau/Nouveau dossier/

/
bin

Filename	Filesize	Filetype	Last modified
..			
100_10000seq_sample.tar.gz	4,773,706	gz-file	06/28/2017 05:...
sampleA_R2.fastq	66,157,532	fastq-file	06/27/2017 10:...
sampleA_R1.fastq	66,157,532	fastq-file	06/27/2017 10:...
100_10000seq_sampleC3.fastq	9,714,424	fastq-file	06/16/2015 10:...
100_10000seq_sampleC2.fastq	9,714,107	fastq-file	06/16/2015 10:...
100_10000seq_sampleC1.fastq	9,716,445	fastq-file	06/16/2015 10:...
100_10000seq_sampleB3.fastq	9,707,364	fastq-file	06/16/2015 10:...
100_10000seq_sampleB2.fastq	9,707,921	fastq-file	06/16/2015 10:...
100_10000seq_sampleB1.fastq	9,709,480	fastq-file	06/16/2015 10:...
100_10000seq_sampleA3.fastq	9,859,424	fastq-file	06/16/2015 10:...
100_10000seq_sampleA2.fastq	9,863,209	fastq-file	06/16/2015 10:...
100_10000seq_sampleA1.fastq	9,862,292	fastq-file	06/16/2015 10:...

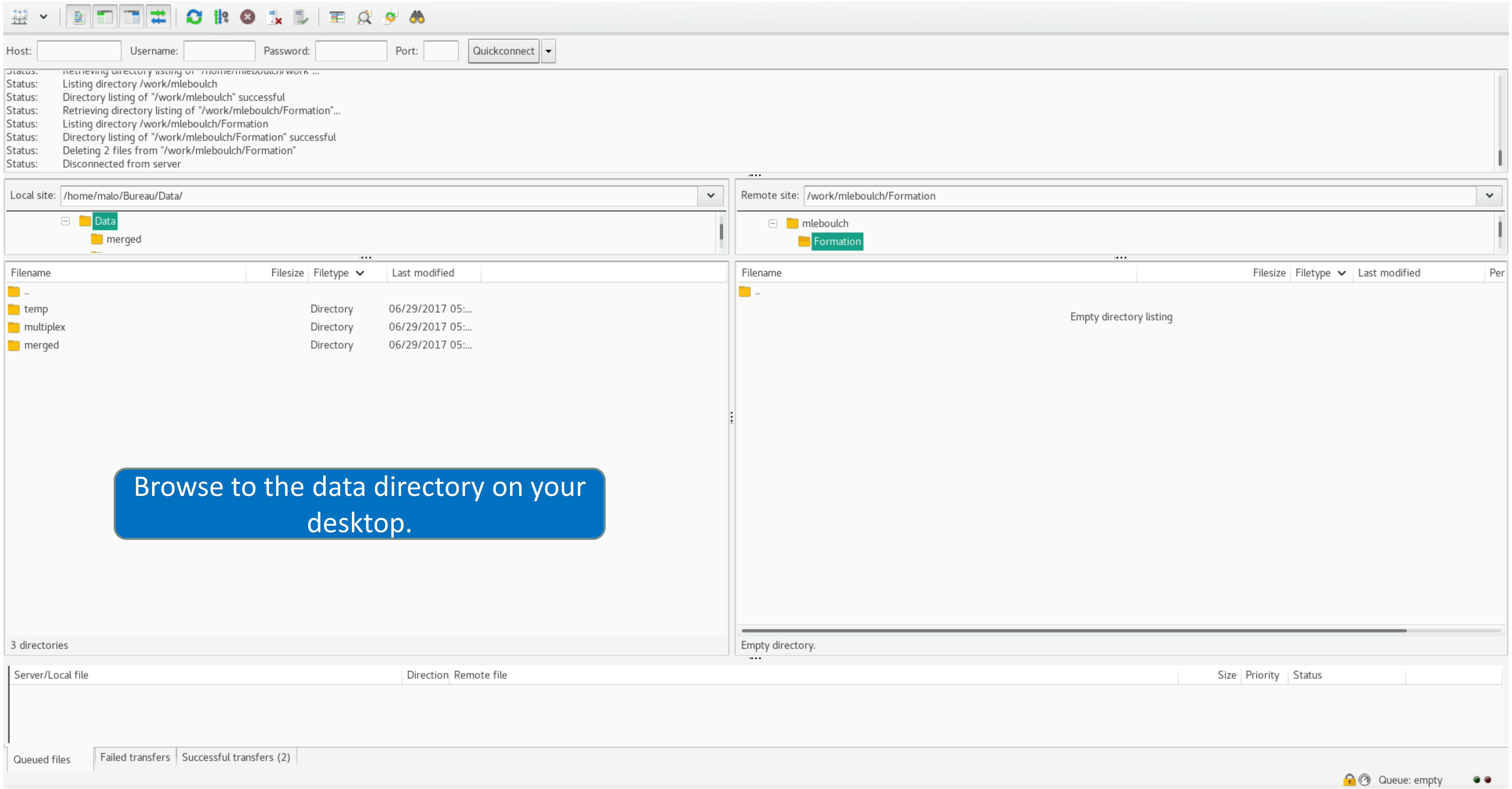
12 files. Total size: 224,943,436 bytes

Filename	Filesize	Filetype	Last modified	Per

Not connected to any server

Not connected.

Queue: empty



```
Status:	Listing directory /work/mleboulch
Status:	Directory listing of "/work/mleboulch" successful
Status:	Retrieving directory listing of "/work/mleboulch/Formation"...
Status:	Listing directory /work/mleboulch/Formation
Status:	Directory listing of "/work/mleboulch/Formation" successful
Status:	Deleting 2 files from "/work/mleboulch/Formation"
Status:	Disconnected from server
```

Local site: /home/malo/Bureau/Data/

Remote site: /work/mleboulch/Formation

Filename	Filesize	Filetype	Last modified
..			
temp		Directory	06/29/2017 05:...
multiplex		Directory	06/29/2017 05:...
merged		Directory	06/29/2017 05:...

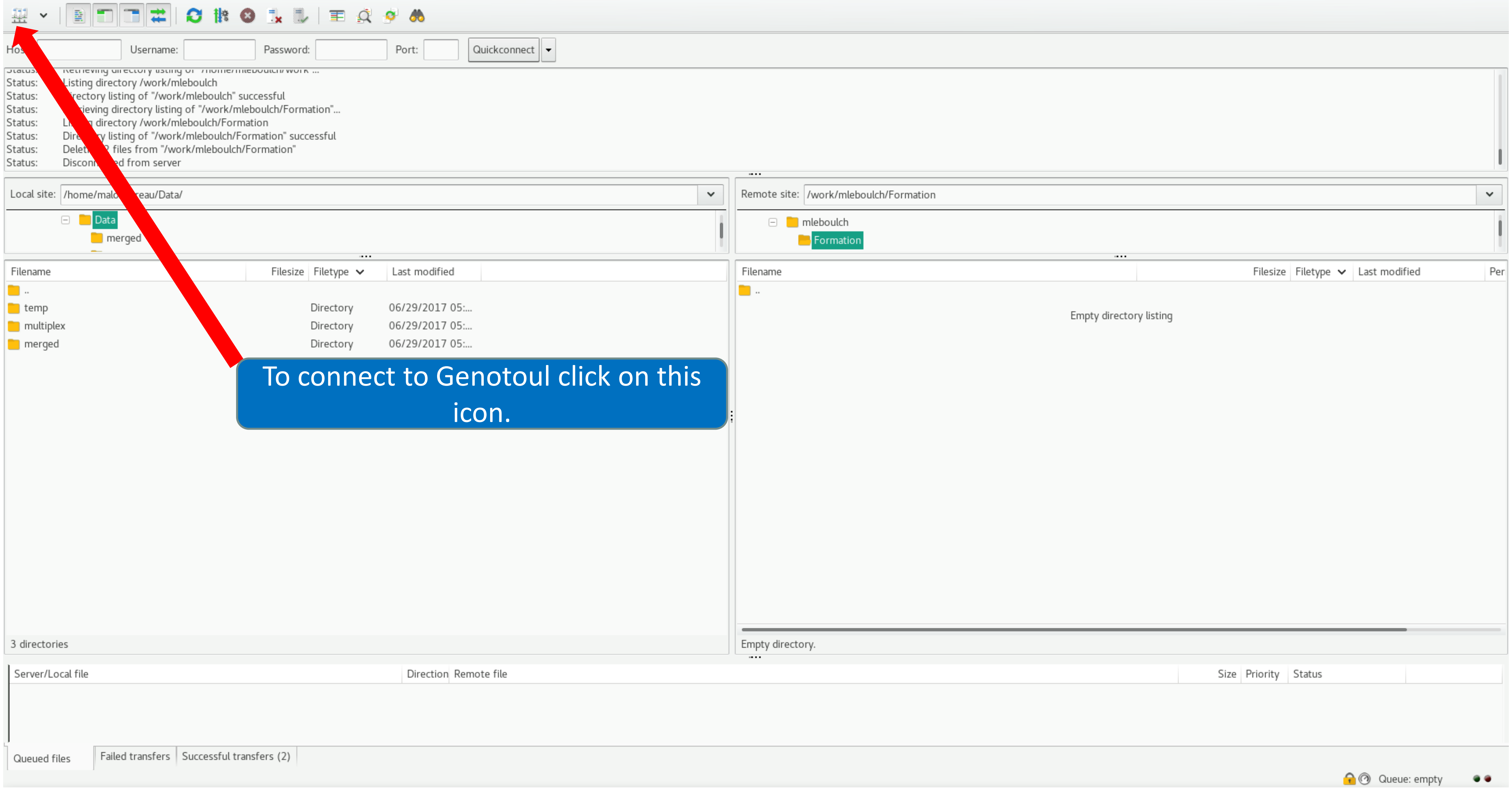
3 directories

Filename	Filesize	Filetype	Last modified	Per
..				

Empty directory listing

Empty directory.

Browse to the data directory on your desktop.



```
Status:     Retrieving directory listing of "/home/mleboulch/work"...
Status:     Listing directory /work/mleboulch
Status:     irectory listing of "/work/mleboulch" successful
Status:        ieving directory listing of "/work/mleboulch/Formation"...
Status:     Li    g directory /work/mleboulch/Formation
Status:     Dire     y listing of "/work/mleboulch/Formation" successful
Status:     Delet     2 files from "/work/mleboulch/Formation"
Status:     Disconn     ed from server
```

Local site: /home/malo   reau/Data/

- Data
  - merged

Filename	Filesize	Filetype	Last modified
..			
temp		Directory	06/29/2017 05:...
multiplex		Directory	06/29/2017 05:...
merged		Directory	06/29/2017 05:...

3 directories

Remote site: /work/mleboulch/Formation

- mleboulch
  - Formation

Filename	Filesize	Filetype	Last modified	Per
..				

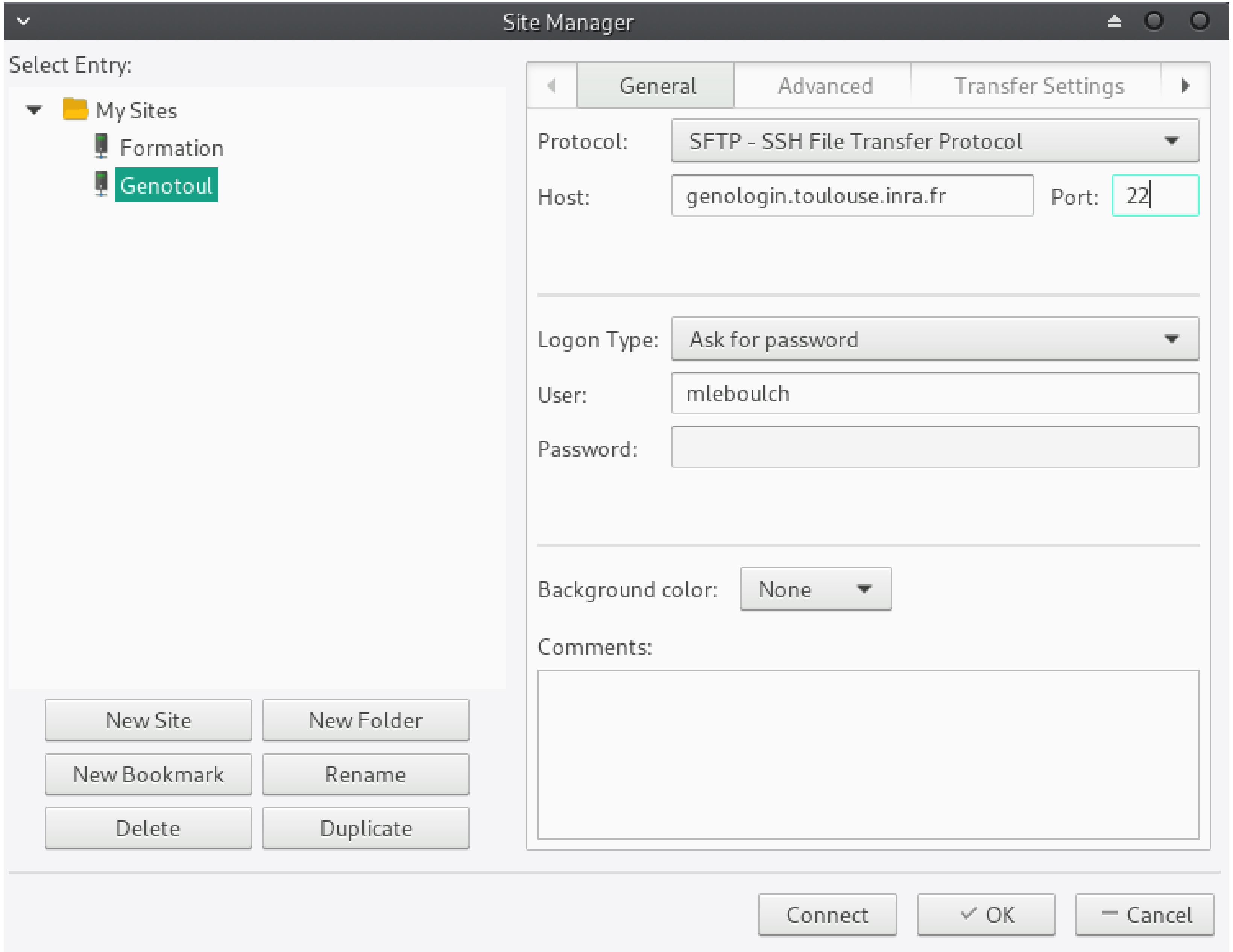
Empty directory listing

Empty directory.

To connect to Genotoul click on this icon.

- Host: genologin.toulouse.inra.fr
- Port: 22
- Protocol: SFTP
- User: your Genotoul login
- Logon Type: Ask for password
- Click on connect.

- Password: your password



Protocol: SFTP - SSH File Transfer Protocol

Host: genologin.toulouse.inra.fr

Port: 22

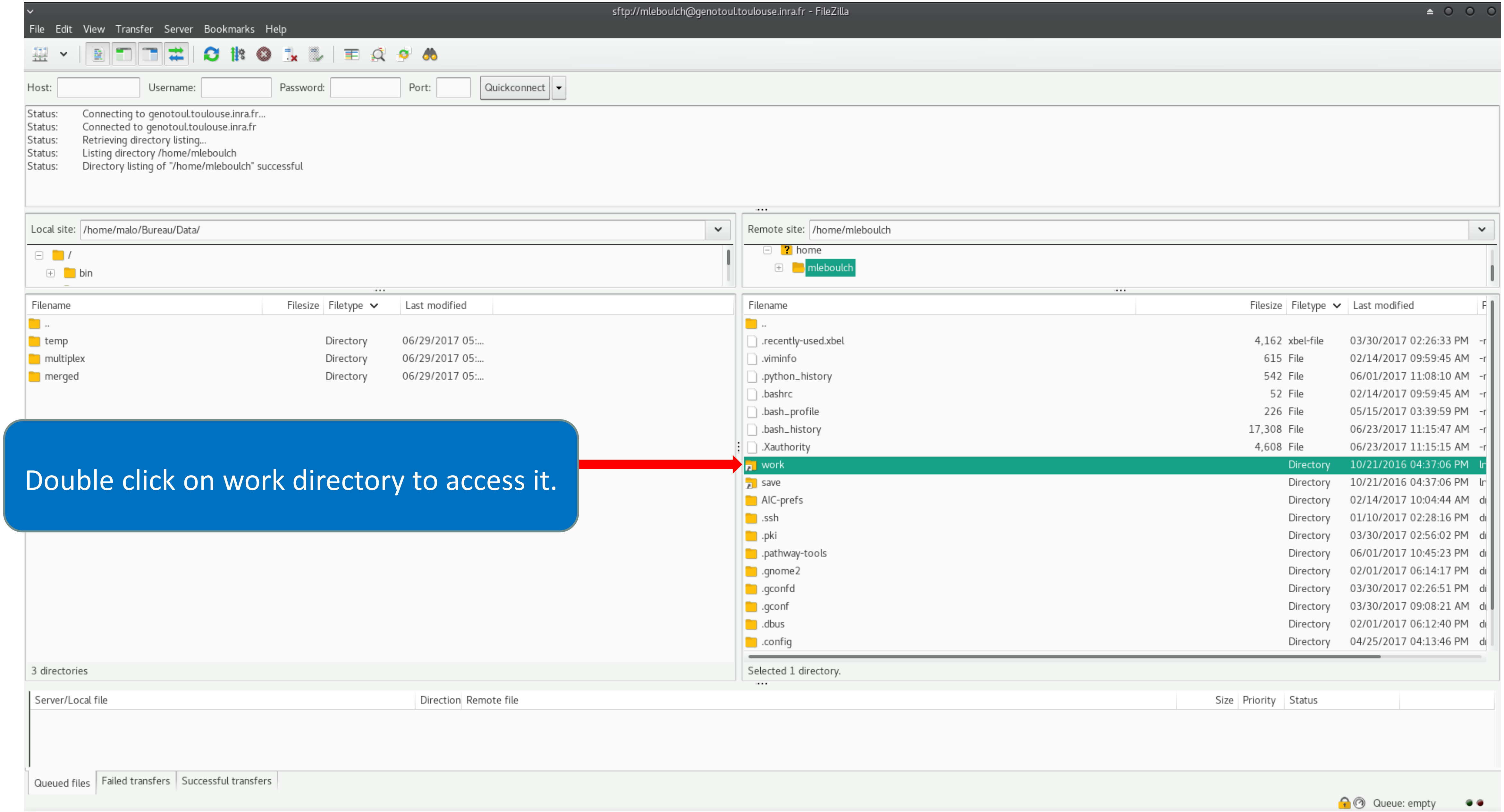
Logon Type: Ask for password

User: mleboulch

Password:

Background color: None

Comments:



```
Status:      Connecting to genotoul.toulouse.inra.fr...
Status:      Connected to genotoul.toulouse.inra.fr
Status:      Retrieving directory listing...
Status:      Listing directory /home/mleboulch
Status:      Directory listing of "/home/mleboulch" successful
```

Local site: /home/malo/Bureau/Data/

- /
  - bin

Filename	Filesize	Filetype	Last modified
..			
temp		Directory	06/29/2017 05:...
multiplex		Directory	06/29/2017 05:...
merged		Directory	06/29/2017 05:...

3 directories

Remote site: /home/mleboulch

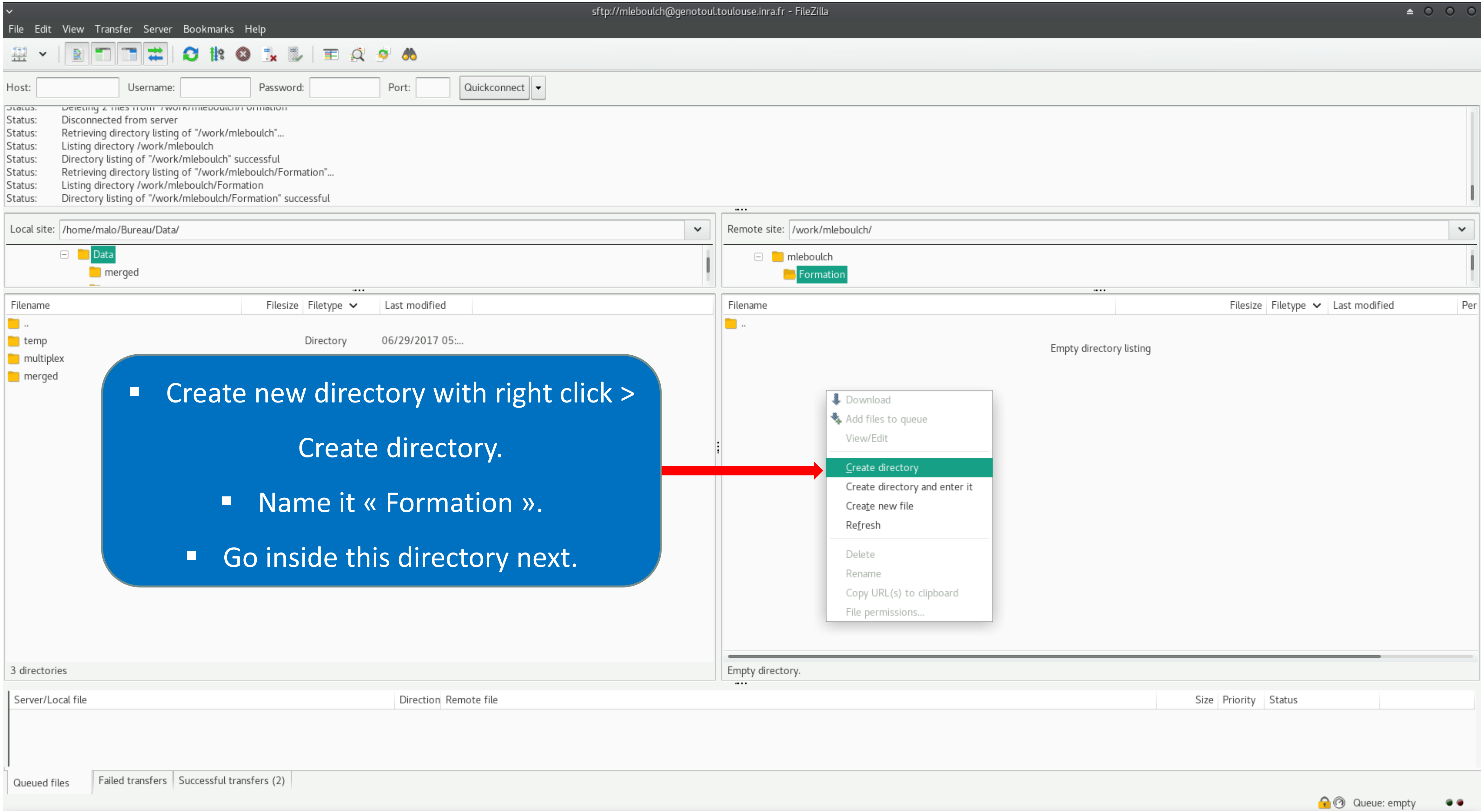
- home
  - mleboulch

Filename	Filesize	Filetype	Last modified	F
..				
.recently-used.xbel	4,162	xbel-file	03/30/2017 02:26:33 PM	-r
.viminfo	615	File	02/14/2017 09:59:45 AM	-r
.python_history	542	File	06/01/2017 11:08:10 AM	-r
.bashrc	52	File	02/14/2017 09:59:45 AM	-r
.bash_profile	226	File	05/15/2017 03:39:59 PM	-r
.bash_history	17,308	File	06/23/2017 11:15:47 AM	-r
.Xauthority	4,608	File	06/23/2017 11:15:15 AM	-r
work		Directory	10/21/2016 04:37:06 PM	lr
save		Directory	10/21/2016 04:37:06 PM	lr
AIC-prefs		Directory	02/14/2017 10:04:44 AM	d
.ssh		Directory	01/10/2017 02:28:16 PM	d
.pki		Directory	03/30/2017 02:56:02 PM	d
.pathway-tools		Directory	06/01/2017 10:45:23 PM	d
.gnome2		Directory	02/01/2017 06:14:17 PM	d
.gconfd		Directory	03/30/2017 02:26:51 PM	d
.gconf		Directory	03/30/2017 09:08:21 AM	d
.dbus		Directory	02/01/2017 06:12:40 PM	d
.config		Directory	04/25/2017 04:13:46 PM	d

Selected 1 directory.

Server/Local file	Direction	Remote file	Size	Priority	Status

Double click on work directory to access it.



```
Status:	Disconnected from server
Status:	Retrieving directory listing of "/work/mleboulch"...
Status:	Listing directory /work/mleboulch
Status:	Directory listing of "/work/mleboulch" successful
Status:	Retrieving directory listing of "/work/mleboulch/Formation"...
Status:	Listing directory /work/mleboulch/Formation
Status:	Directory listing of "/work/mleboulch/Formation" successful
```

Local site: /home/malo/Bureau/Data/

Remote site: /work/mleboulch/

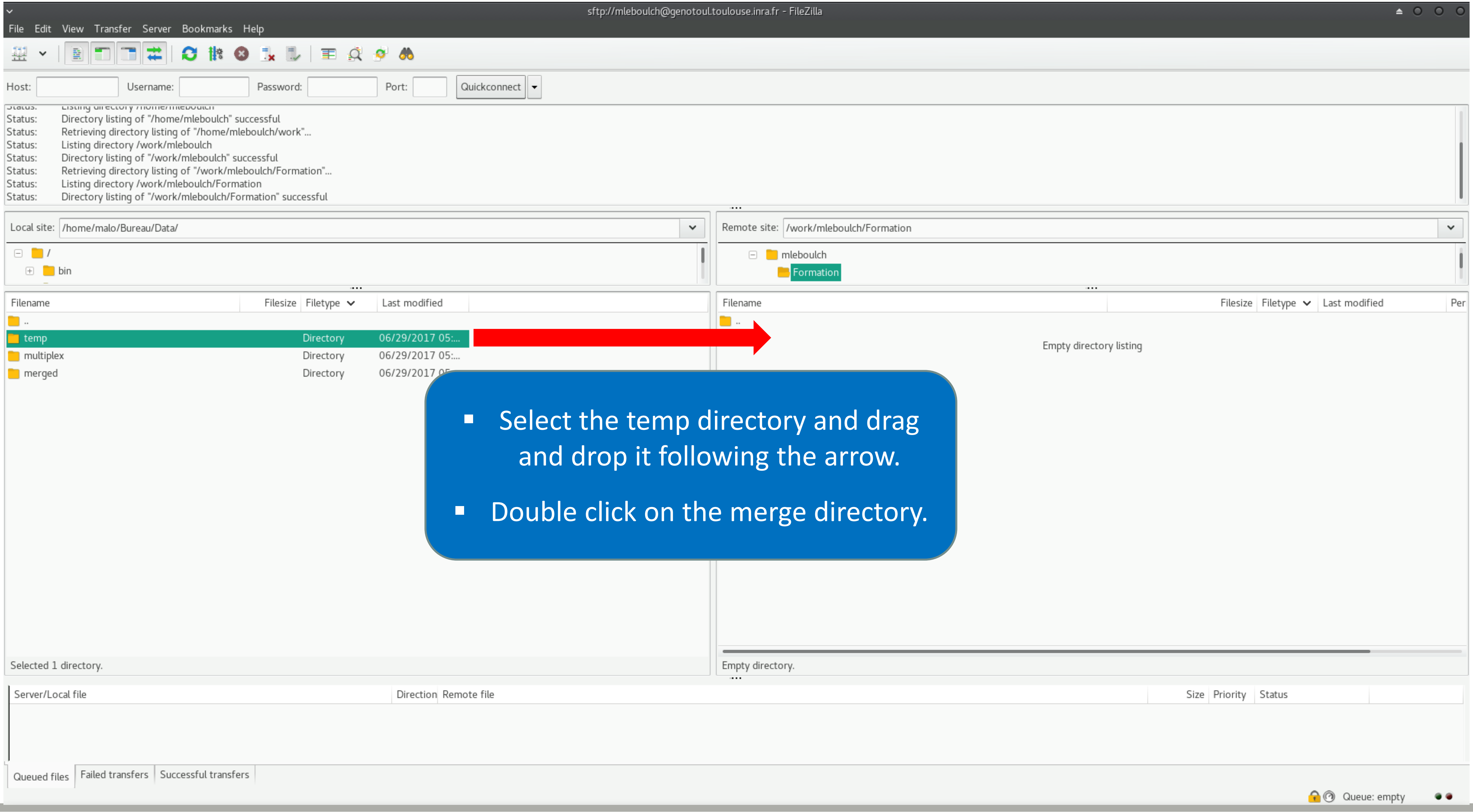
Filename	Filesize	Filetype	Last modified
..			
temp		Directory	06/29/2017 05:...
multiplex			
merged			

3 directories

Empty directory listing

Empty directory.

- Create new directory with right click > Create directory.
  - Name it « Formation ».
  - Go inside this directory next.



```
Status:	Listing directory /home/mleboulch
Status:	Directory listing of "/home/mleboulch" successful
Status:	Retrieving directory listing of "/home/mleboulch/work"...
Status:	Listing directory /work/mleboulch
Status:	Directory listing of "/work/mleboulch" successful
Status:	Retrieving directory listing of "/work/mleboulch/Formation"...
Status:	Listing directory /work/mleboulch/Formation
Status:	Directory listing of "/work/mleboulch/Formation" successful
```

Local site: /home/malo/Bureau/Data/

Remote site: /work/mleboulch/Formation

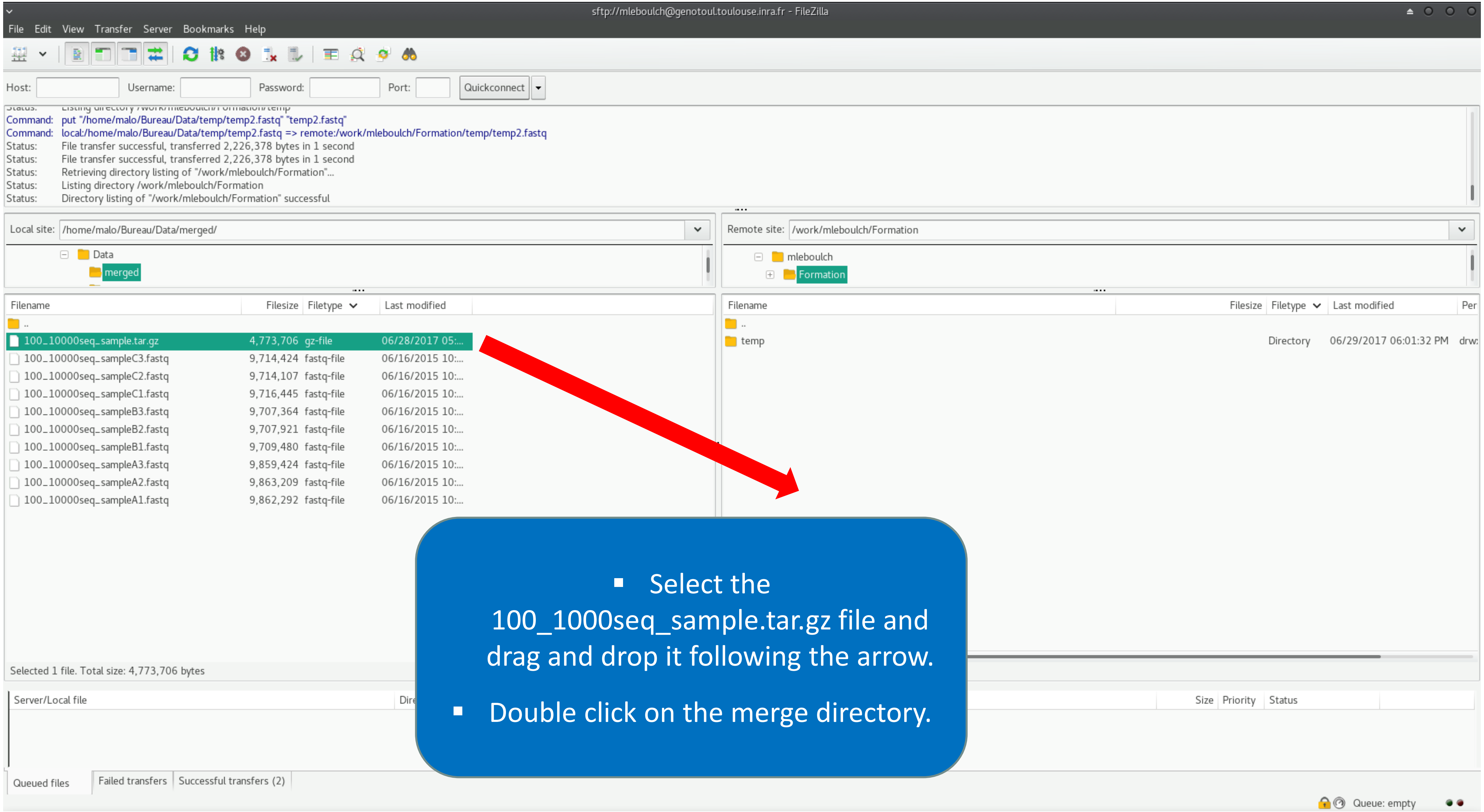
Filename	Filesize	Filetype	Last modified
..			
temp		Directory	06/29/2017 05:...
multiplex		Directory	06/29/2017 05:...
merged		Directory	06/29/2017 05

Selected 1 directory.

Empty directory listing

Empty directory.

- Select the temp directory and drag and drop it following the arrow.
- Double click on the merge directory.



```
Status:     Listing directory /work/mleboulch/Formation/temp
Command:    put "/home/malo/Bureau/Data/temp/temp2.fastq" "temp2.fastq"
Command:    local:/home/malo/Bureau/Data/temp/temp2.fastq => remote:/work/mleboulch/Formation/temp/temp2.fastq
Status:     File transfer successful, transferred 2,226,378 bytes in 1 second
Status:     File transfer successful, transferred 2,226,378 bytes in 1 second
Status:     Retrieving directory listing of "/work/mleboulch/Formation"...
Status:     Listing directory /work/mleboulch/Formation
Status:     Directory listing of "/work/mleboulch/Formation" successful
```

Local site: /home/malo/Bureau/Data/merged/

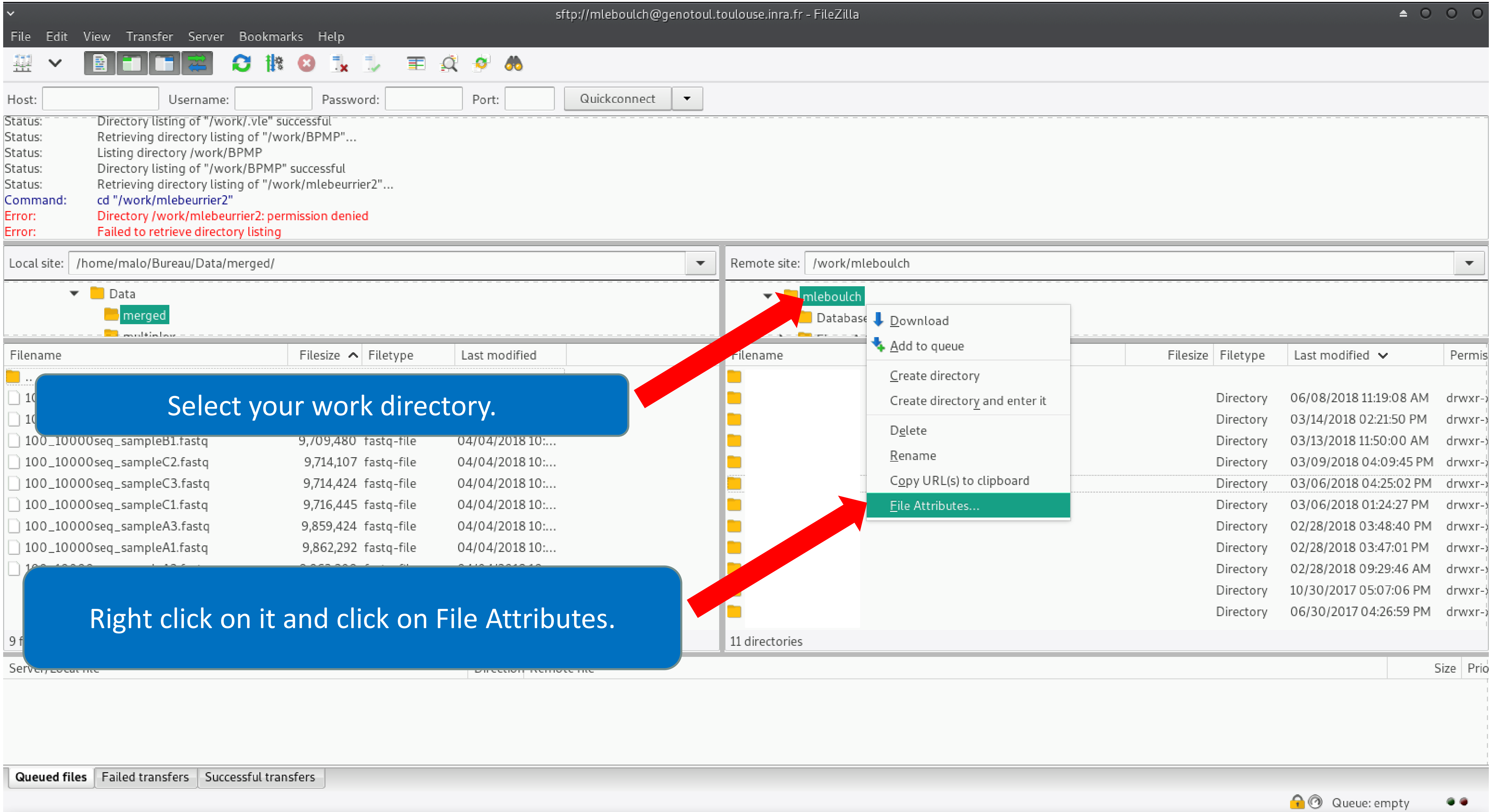
Remote site: /work/mleboulch/Formation

Filename	Filesize	Filetype	Last modified
..			
100_10000seq_sample.tar.gz	4,773,706	gz-file	06/28/2017 05:...
100_10000seq_sampleC3.fastq	9,714,424	fastq-file	06/16/2015 10:...
100_10000seq_sampleC2.fastq	9,714,107	fastq-file	06/16/2015 10:...
100_10000seq_sampleC1.fastq	9,716,445	fastq-file	06/16/2015 10:...
100_10000seq_sampleB3.fastq	9,707,364	fastq-file	06/16/2015 10:...
100_10000seq_sampleB2.fastq	9,707,921	fastq-file	06/16/2015 10:...
100_10000seq_sampleB1.fastq	9,709,480	fastq-file	06/16/2015 10:...
100_10000seq_sampleA3.fastq	9,859,424	fastq-file	06/16/2015 10:...
100_10000seq_sampleA2.fastq	9,863,209	fastq-file	06/16/2015 10:...
100_10000seq_sampleA1.fastq	9,862,292	fastq-file	06/16/2015 10:...

Selected 1 file. Total size: 4,773,706 bytes

Filename	Filesize	Filetype	Last modified	Per
..				
temp		Directory	06/29/2017 06:01:32 PM	drw:

- Select the 100_1000seq_sample.tar.gz file and drag and drop it following the arrow.
- Double click on the merge directory.



```
Status:     Directory listing of "/work/.vle" successful
Status:     Retrieving directory listing of "/work/BPMP"...
Status:     Listing directory /work/BPMP
Status:     Directory listing of "/work/BPMP" successful
Status:     Retrieving directory listing of "/work/mlebeurrier2"...
Command:    cd "/work/mlebeurrier2"
Error:      Directory /work/mlebeurrier2: permission denied
Error:      Failed to retrieve directory listing
```

Local site: /home/malo/Bureau/Data/merged/

Remote site: /work/mleboulch

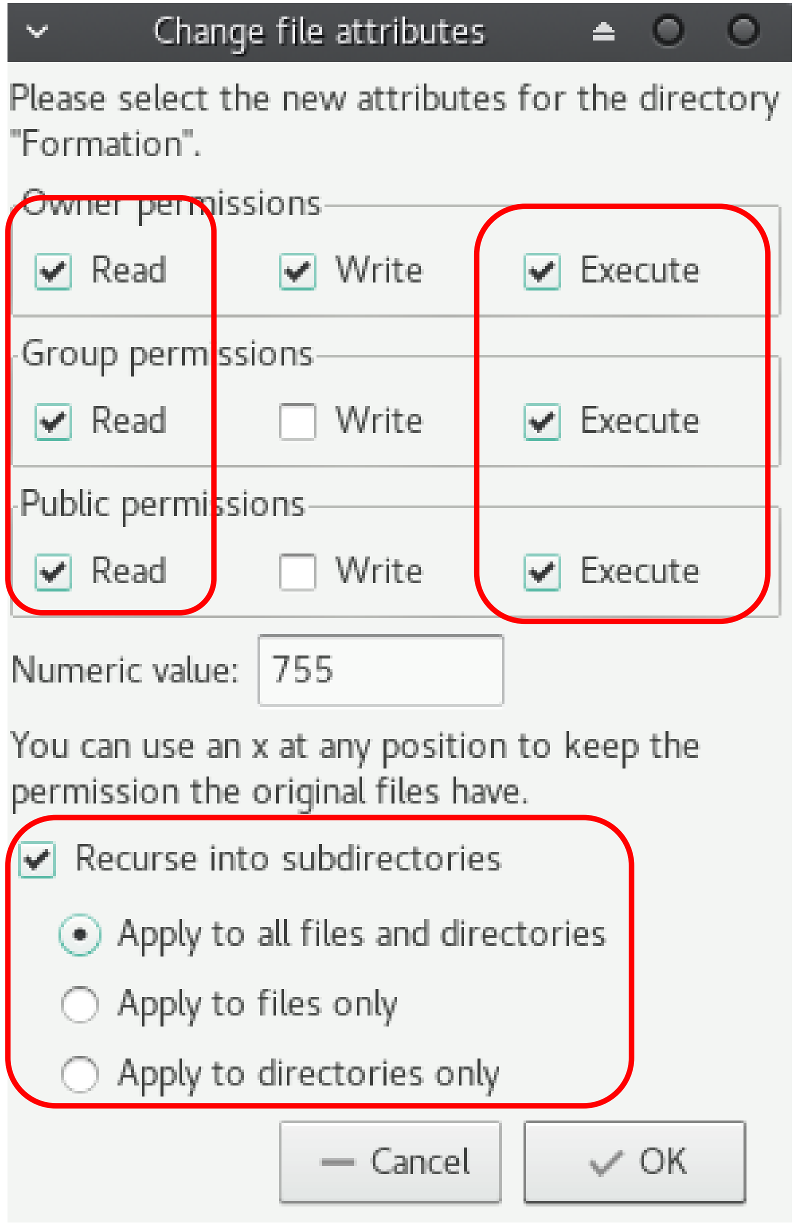
Filename	Filesize	Filetype	Last modified
100_10000seq_sampleB1.fastq	9,709,480	fastq-file	04/04/2018 10:...
100_10000seq_sampleC2.fastq	9,714,107	fastq-file	04/04/2018 10:...
100_10000seq_sampleC3.fastq	9,714,424	fastq-file	04/04/2018 10:...
100_10000seq_sampleC1.fastq	9,716,445	fastq-file	04/04/2018 10:...
100_10000seq_sampleA3.fastq	9,859,424	fastq-file	04/04/2018 10:...
100_10000seq_sampleA1.fastq	9,862,292	fastq-file	04/04/2018 10:...

Filetype	Last modified
Directory	06/08/2018 11:19:08 AM
Directory	03/14/2018 02:21:50 PM
Directory	03/13/2018 11:50:00 AM
Directory	03/09/2018 04:09:45 PM
Directory	03/06/2018 04:25:02 PM
Directory	03/06/2018 01:24:27 PM
Directory	02/28/2018 03:48:40 PM
Directory	02/28/2018 03:47:01 PM
Directory	02/28/2018 09:29:46 AM
Directory	10/30/2017 05:07:06 PM
Directory	06/30/2017 04:26:59 PM

11 directories

Select your work directory.

Right click on it and click on File Attributes.



Change file attributes

Please select the new attributes for the directory "Formation".

Owner permissions
☑ Read ☑ Write ☑ Execute

Group permissions
☑ Read ☐ Write ☑ Execute

Public permissions
☑ Read ☐ Write ☑ Execute

Numeric value: 755

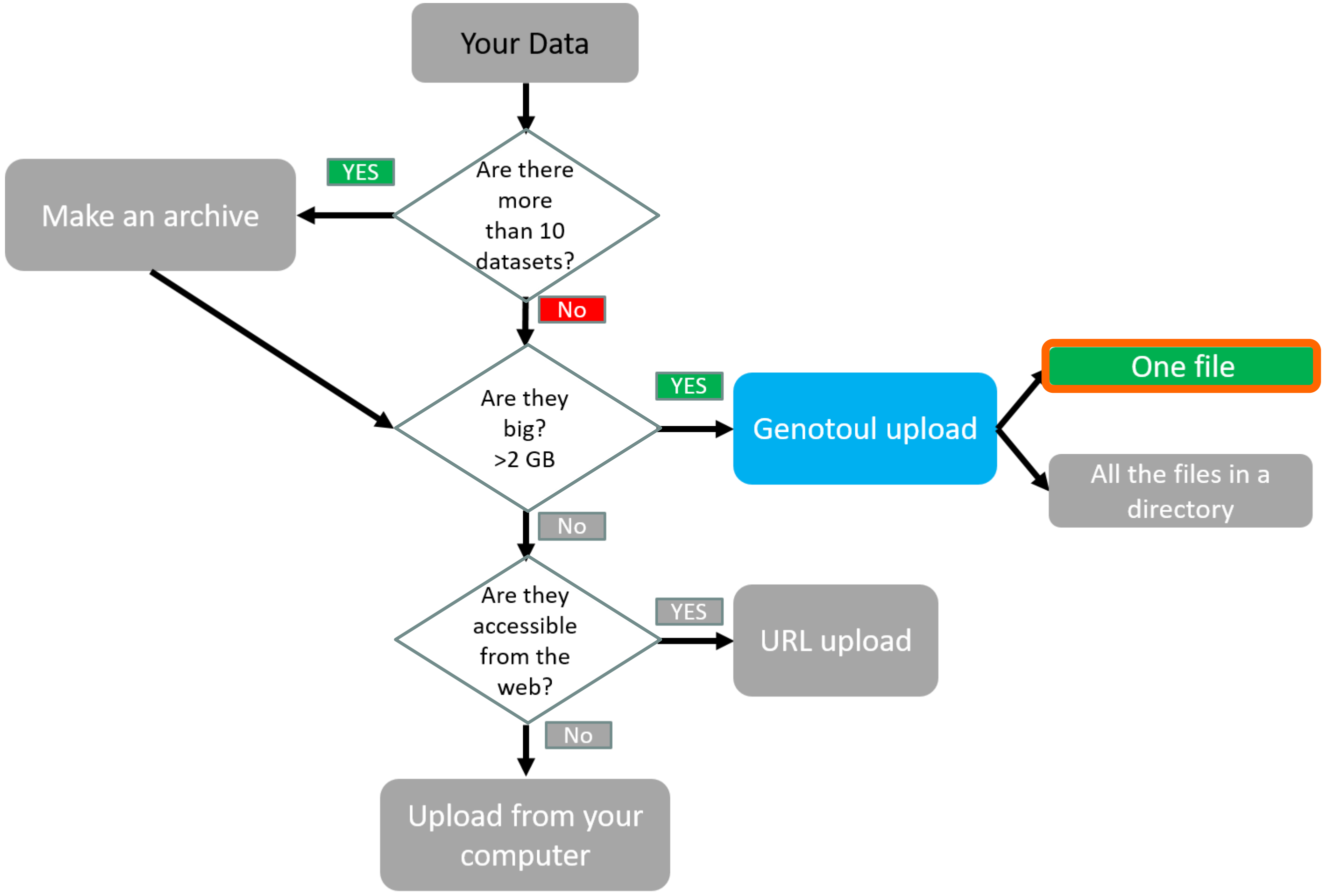
You can use an x at any position to keep the permission the original files have.

☑ Recurse into subdirectories

● Apply to all files and directories
○ Apply to files only
○ Apply to directories only

- Check that all boxes for execute and read are checked.
- Recurse this action to all files and subdirectories.
- That allows Galaxy to access your files on Genotoul.
- Click on Ok.

Upload files from Genotoul



Your Data

Are there more than 10 datasets?

YES

Make an archive

No

Are they big? >2 GB

YES

Genotoul upload

One file

All the files in a directory

No

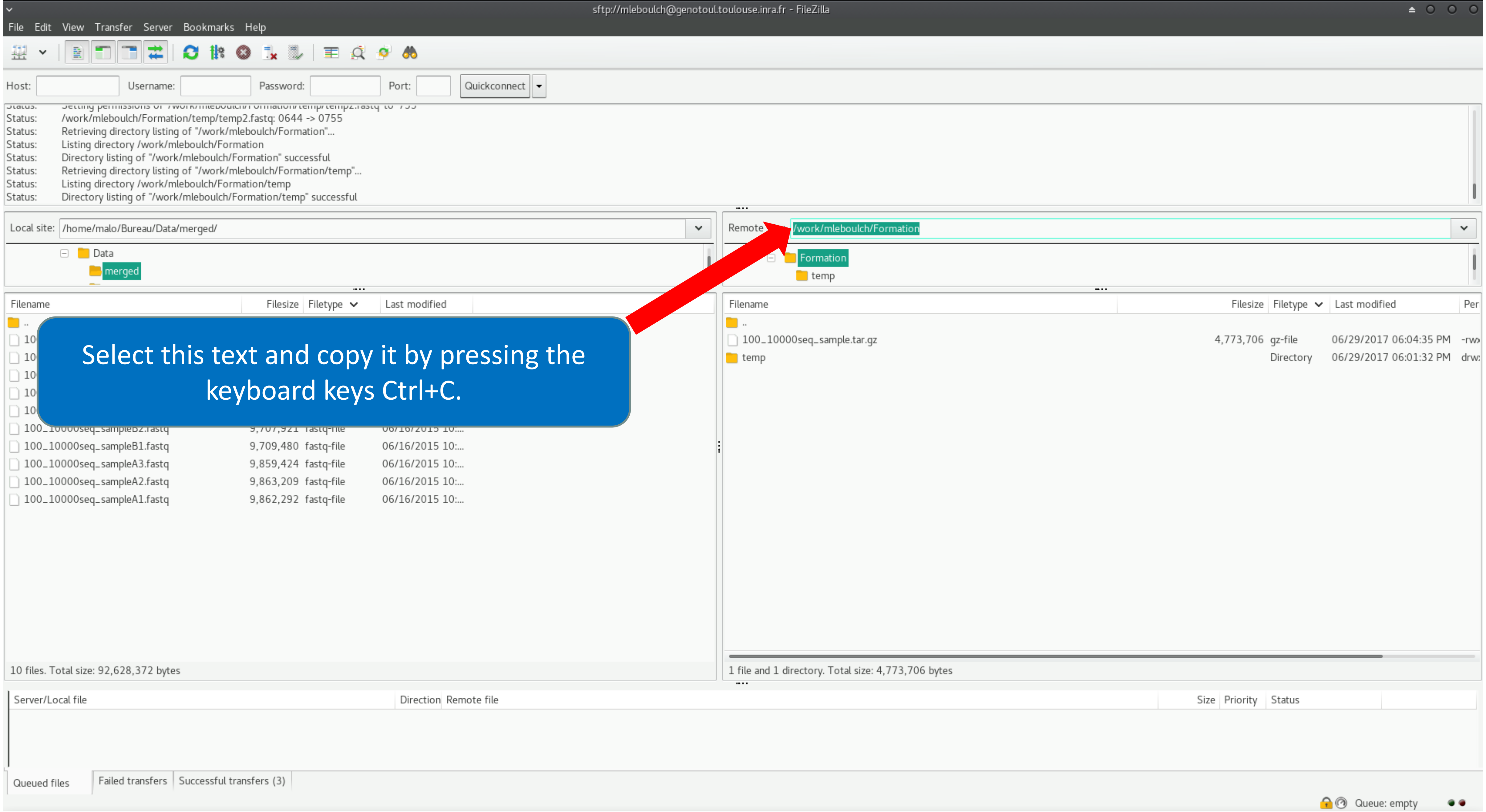
Are they accessible from the web?

YES

URL upload

No

Upload from your computer



```
Status:	/work/mleboulch/Formation/temp/temp2.fastq: 0644 -> 0755
Status:	Retrieving directory listing of "/work/mleboulch/Formation"...
Status:	Listing directory /work/mleboulch/Formation
Status:	Directory listing of "/work/mleboulch/Formation" successful
Status:	Retrieving directory listing of "/work/mleboulch/Formation/temp"...
Status:	Listing directory /work/mleboulch/Formation/temp
Status:	Directory listing of "/work/mleboulch/Formation/temp" successful
```

Local site: /home/malo/Bureau/Data/merged/

Remote site: /work/mleboulch/Formation

Filename	Filesize	Filetype	Last modified
..			
10			
10			
10			
10			
10			
100_10000seq_sampleB2.fastq	9,707,921	fastq-file	06/16/2015 10:...
100_10000seq_sampleB1.fastq	9,709,480	fastq-file	06/16/2015 10:...
100_10000seq_sampleA3.fastq	9,859,424	fastq-file	06/16/2015 10:...
100_10000seq_sampleA2.fastq	9,863,209	fastq-file	06/16/2015 10:...
100_10000seq_sampleA1.fastq	9,862,292	fastq-file	06/16/2015 10:...

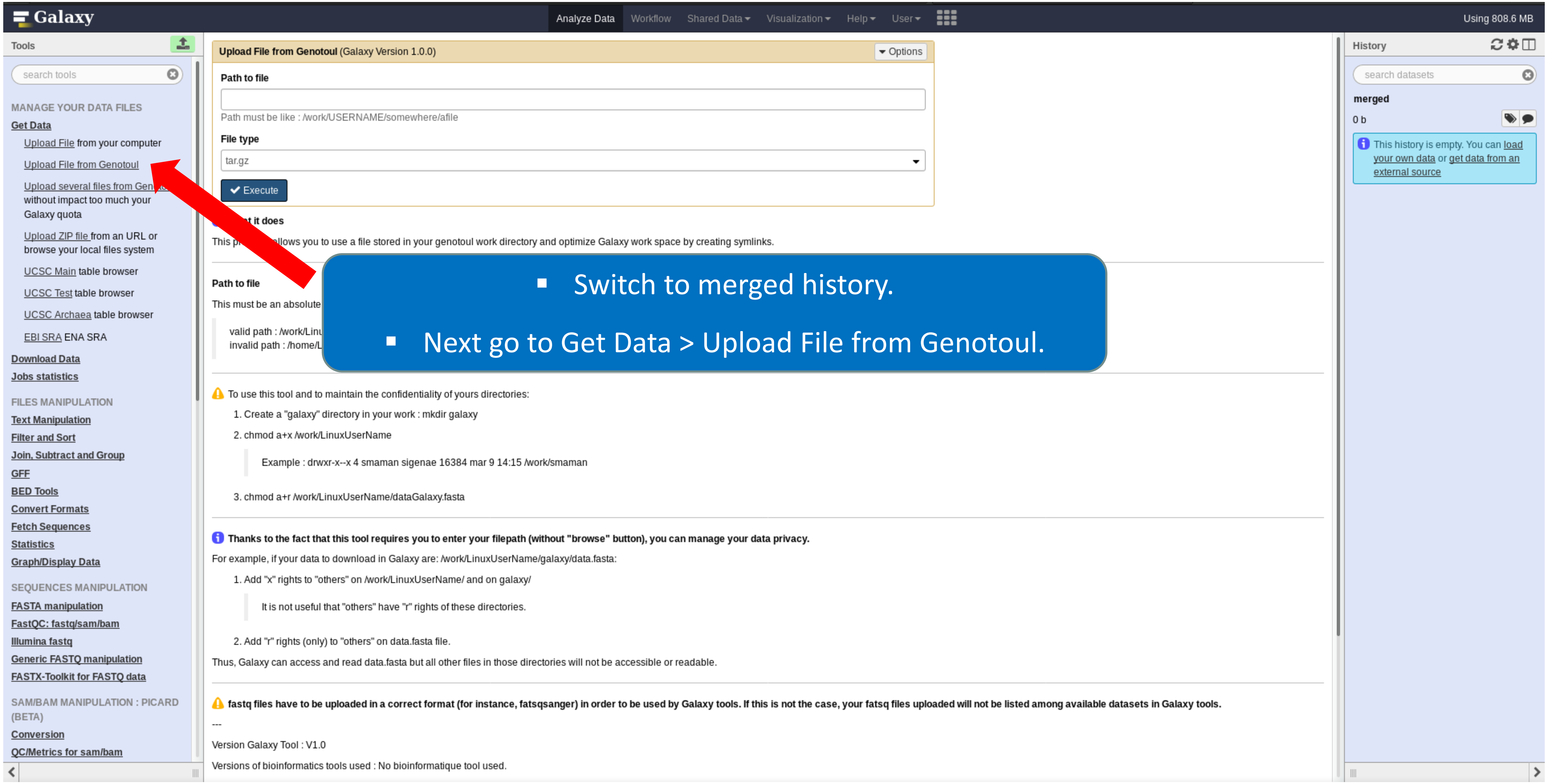
10 files. Total size: 92,628,372 bytes

Filename	Filesize	Filetype	Last modified	Per
..				
100_10000seq_sample.tar.gz	4,773,706	gz-file	06/29/2017 06:04:35 PM	-rw
temp		Directory	06/29/2017 06:01:32 PM	drw

1 file and 1 directory. Total size: 4,773,706 bytes

Server/Local file	Direction	Remote file	Size	Priority	Status

Select this text and copy it by pressing the keyboard keys Ctrl+C.



Upload File from Genotoul (Galaxy Version 1.0.0)

Path to file

Path must be like : /work/USERNAME/somewhere/afile

File type

tar.gz

t it does

This p llows you to use a file stored in your genotoul work directory and optimize Galaxy work space by creating symlinks.

Path to file

This must be an absolute

valid path : /work/Linu
invalid path : /home/L

To use this tool and to maintain the confidentiality of yours directories:

1. Create a "galaxy" directory in your work : mkdir galaxy
2. chmod a+x /work/LinuxUserName

   Example : drwxr-x--x 4 smaman sigenae 16384 mar 9 14:15 /work/smaman

3. chmod a+r /work/LinuxUserName/dataGalaxy.fasta

Thanks to the fact that this tool requires you to enter your filepath (without "browse" button), you can manage your data privacy.

For example, if your data to download in Galaxy are: /work/LinuxUserName/galaxy/data.fasta:

1. Add "x" rights to "others" on /work/LinuxUserName/ and on galaxy/

   It is not useful that "others" have "r" rights of these directories.

2. Add "r" rights (only) to "others" on data.fasta file.

Thus, Galaxy can access and read data.fasta but all other files in those directories will not be accessible or readable.

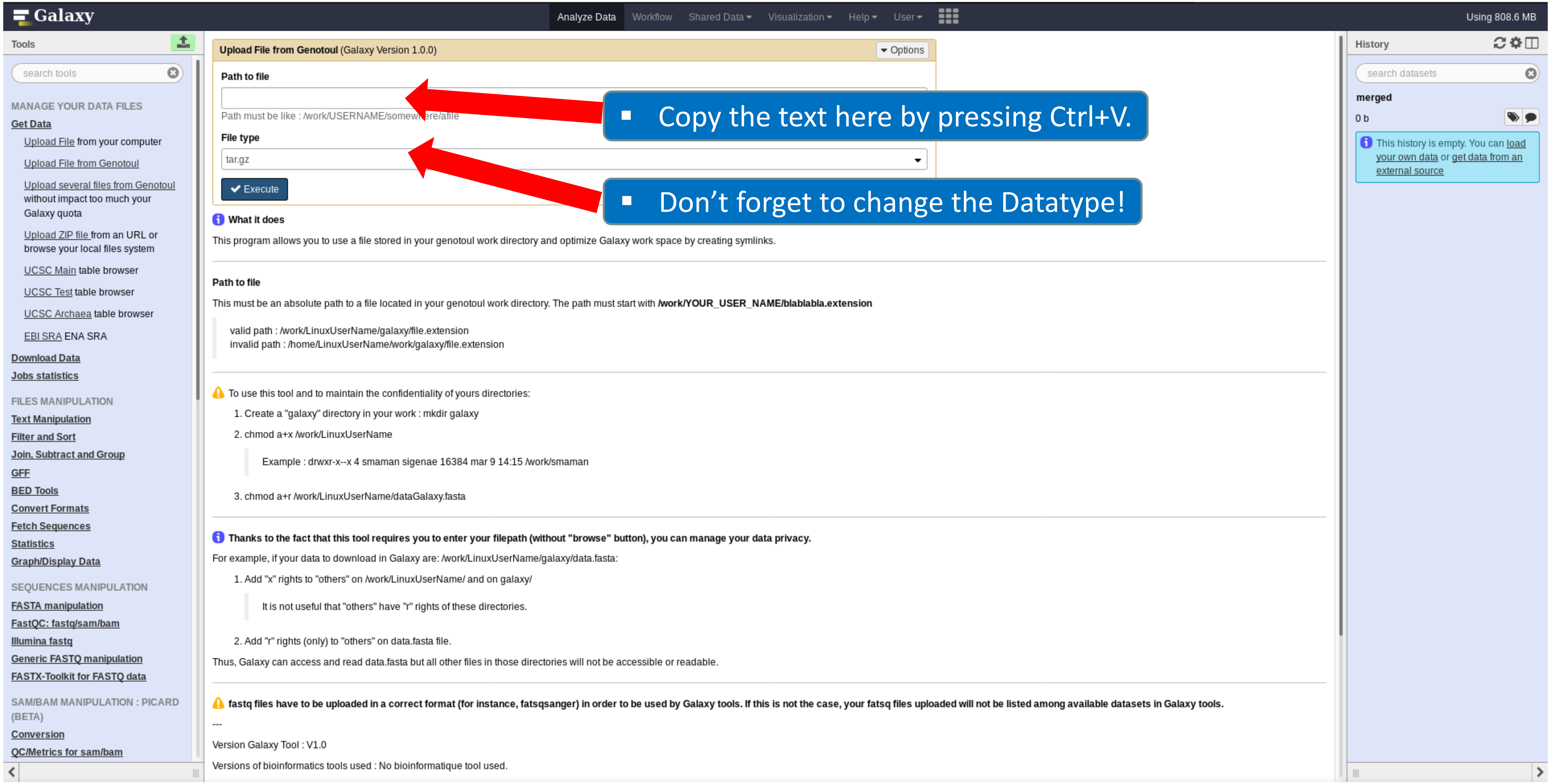
fastq files have to be uploaded in a correct format (for instance, fatsqsanger) in order to be used by Galaxy tools. If this is not the case, your fatsq files uploaded will not be listed among available datasets in Galaxy tools.

Version Galaxy Tool : V1.0

Versions of bioinformatics tools used : No bioinformatique tool used.

This history is empty. You can load your own data or get data from an external source

- Switch to merged history.
- Next go to Get Data > Upload File from Genotoul.



Upload File from Genotoul (Galaxy Version 1.0.0)

Path to file

Path must be like : /work/USERNAME/somewhere/afile

File type

tar.gz

What it does

This program allows you to use a file stored in your genotoul work directory and optimize Galaxy work space by creating symlinks.

Path to file

This must be an absolute path to a file located in your genotoul work directory. The path must start with **/work/YOUR_USER_NAME/blablabla.extension**

> valid path : /work/LinuxUserName/galaxy/file.extension
> invalid path : /home/LinuxUserName/work/galaxy/file.extension

To use this tool and to maintain the confidentiality of yours directories:

1. Create a "galaxy" directory in your work : mkdir galaxy
2. chmod a+x /work/LinuxUserName

   > Example : drwxr-x--x 4 smaman sigenae 16384 mar 9 14:15 /work/smaman

3. chmod a+r /work/LinuxUserName/dataGalaxy.fasta

Thanks to the fact that this tool requires you to enter your filepath (without "browse" button), you can manage your data privacy.

For example, if your data to download in Galaxy are: /work/LinuxUserName/galaxy/data.fasta:

1. Add "x" rights to "others" on /work/LinuxUserName/ and on galaxy/

   > It is not useful that "others" have "r" rights of these directories.

2. Add "r" rights (only) to "others" on data.fasta file.

Thus, Galaxy can access and read data.fasta but all other files in those directories will not be accessible or readable.

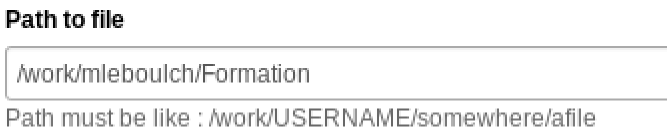
fastq files have to be uploaded in a correct format (for instance, fatsqsanger) in order to be used by Galaxy tools. If this is not the case, your fatsq files uploaded will not be listed among available datasets in Galaxy tools.

Version Galaxy Tool : V1.0

Versions of bioinformatics tools used : No bioinformatique tool used.

This history is empty. You can load your own data or get data from an external source

- Copy the text here by pressing Ctrl+V.
- Don’t forget to change the Datatype!



Path to file

/work/mleboulch/Formation

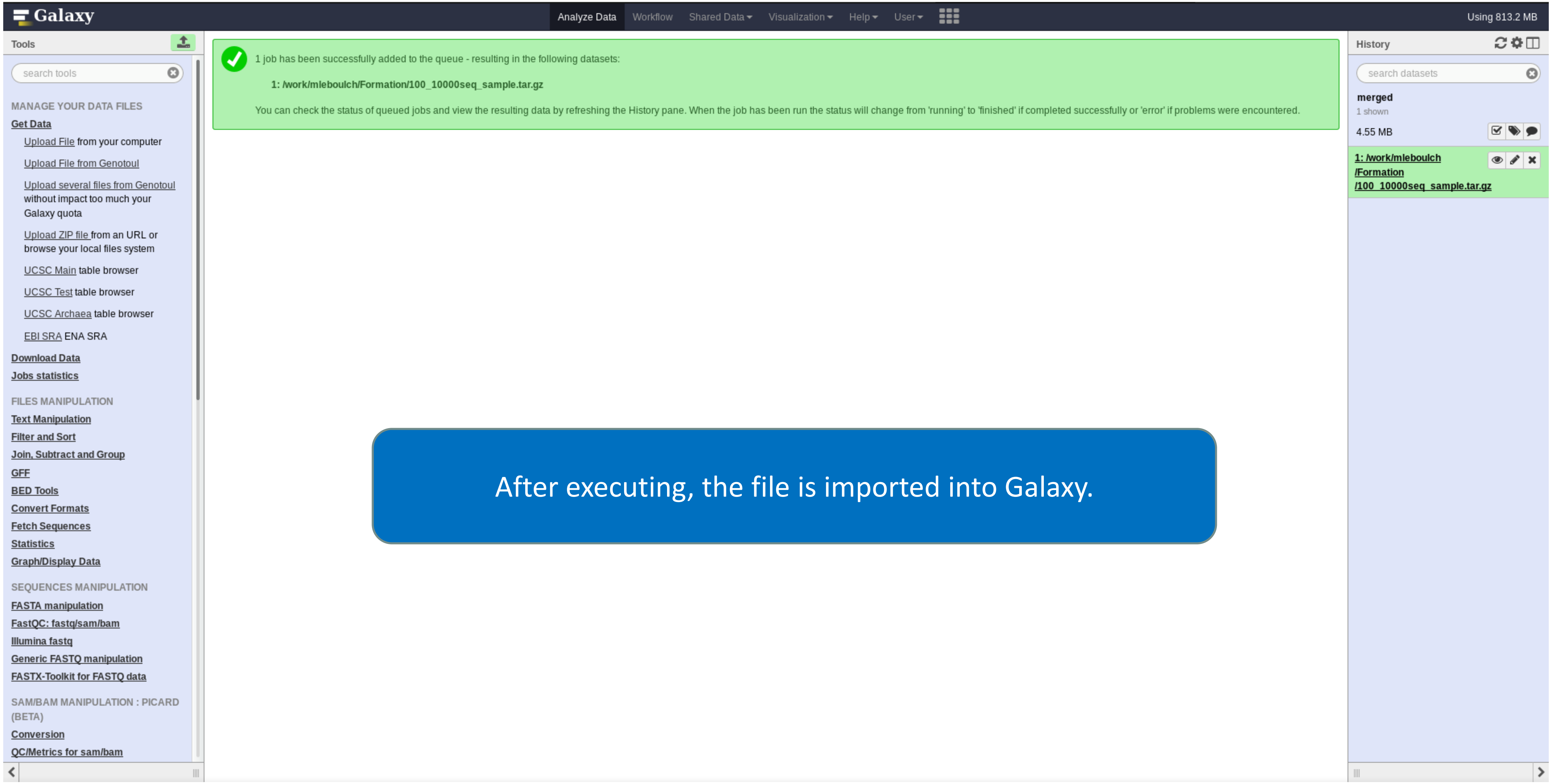
Path must be like : /work/USERNAME/somewhere/afile

Path to file

/work/mleboulch/Formation/100_10000seq_sample.tar.gz

Path must be like : /work/USERNAME/somewhere/afile

You must add « / » and the name of the file at the end of the text.



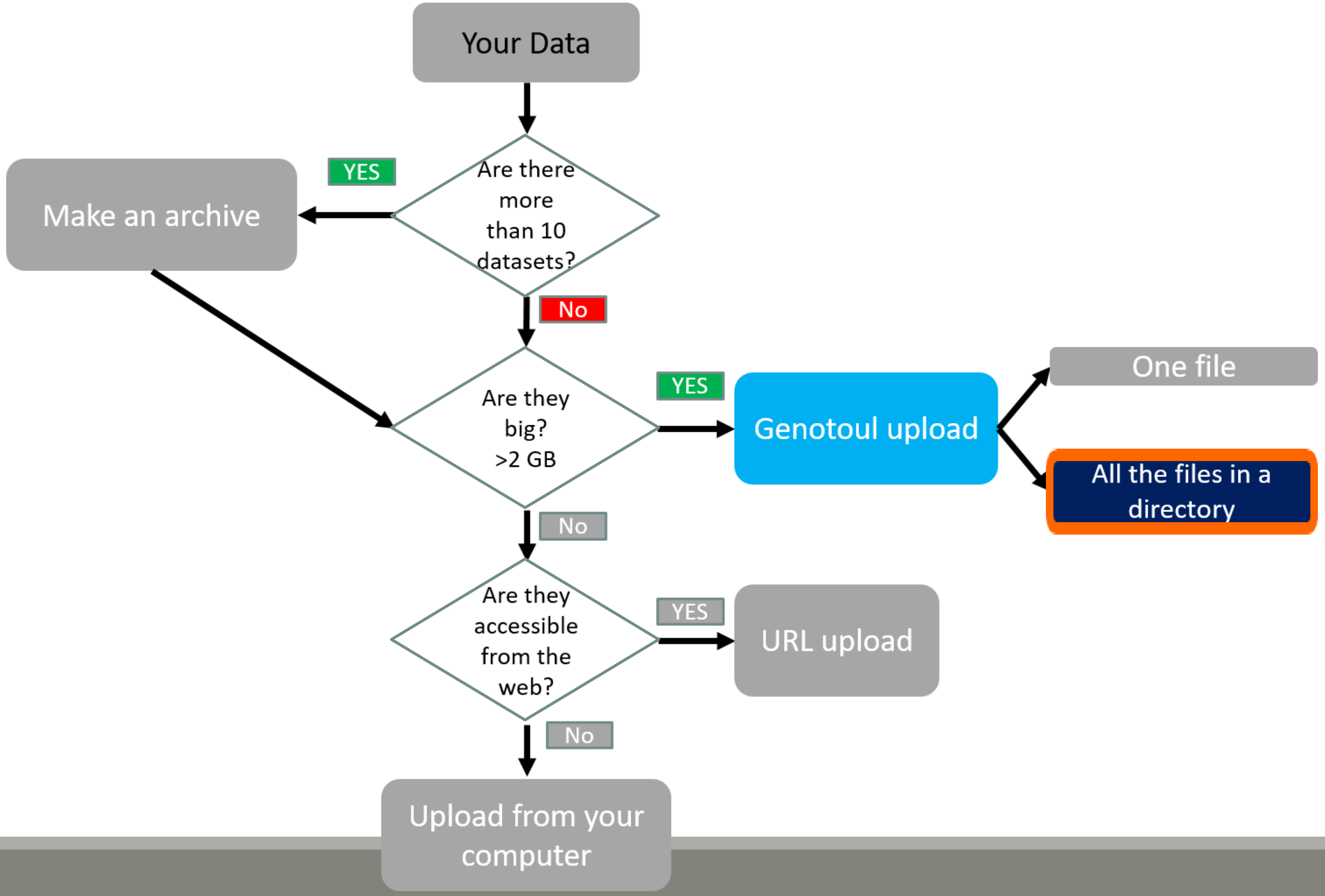
1 job has been successfully added to the queue - resulting in the following datasets:

1: /work/mleboulch/Formation/100_10000seq_sample.tar.gz

You can check the status of queued jobs and view the resulting data by refreshing the History pane. When the job has been run the status will change from 'running' to 'finished' if completed successfully or 'error' if problems were encountered.

After executing, the file is imported into Galaxy.

Upload files from Genotoul



Your Data

Are there more than 10 datasets?

YES

Make an archive

No

Are they big? >2 GB

YES

Genotoul upload

One file

All the files in a directory

No

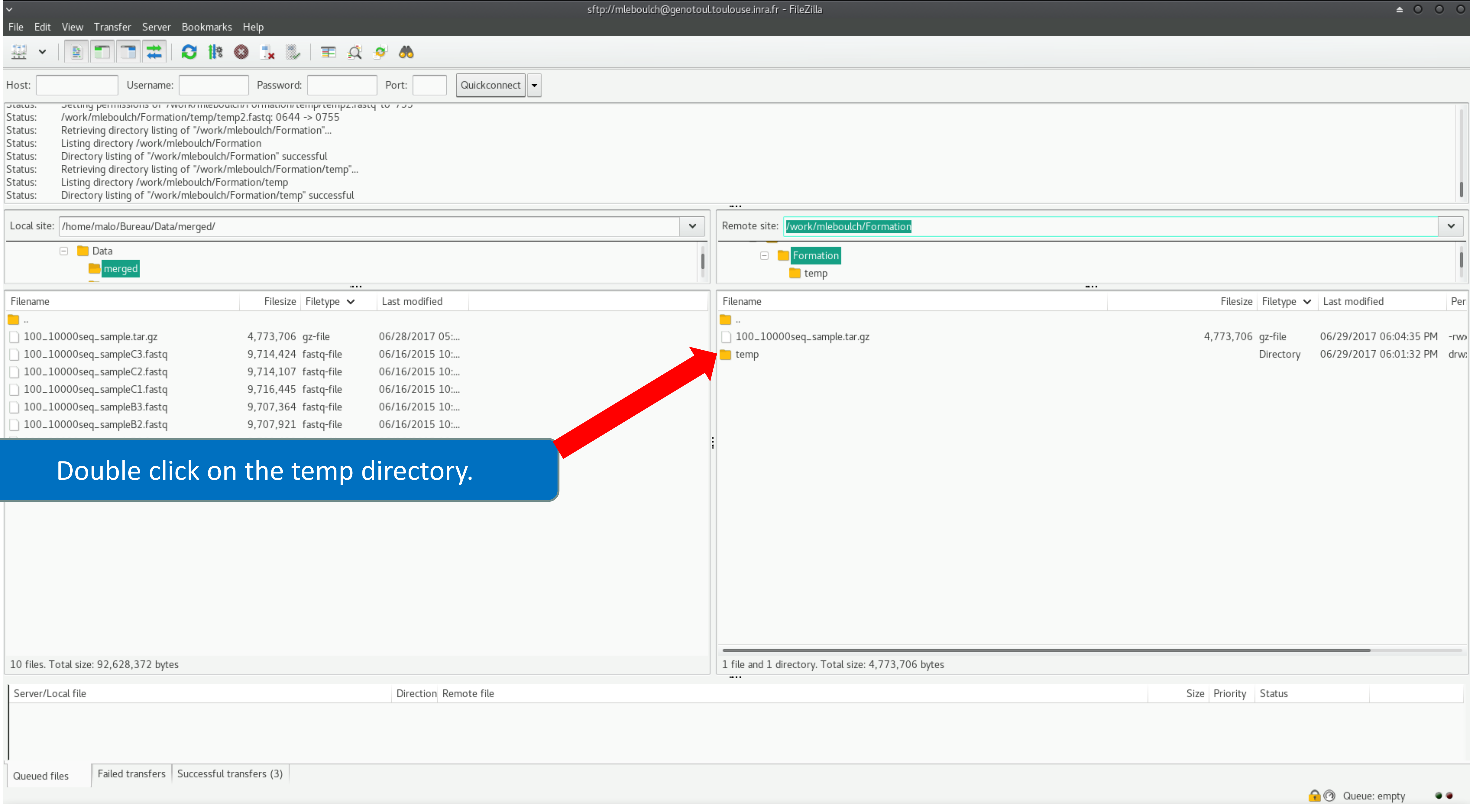
Are they accessible from the web?

YES

URL upload

No

Upload from your computer



```
Status:   Setting permissions of "/work/mleboulch/Formation/temp/temp2.fastq" to 755
Status:   /work/mleboulch/Formation/temp/temp2.fastq: 0644 -> 0755
Status:   Retrieving directory listing of "/work/mleboulch/Formation"...
Status:   Listing directory /work/mleboulch/Formation
Status:   Directory listing of "/work/mleboulch/Formation" successful
Status:   Retrieving directory listing of "/work/mleboulch/Formation/temp"...
Status:   Listing directory /work/mleboulch/Formation/temp
Status:   Directory listing of "/work/mleboulch/Formation/temp" successful
```

Local site: /home/malo/Bureau/Data/merged/

Filename	Filesize	Filetype	Last modified
..			
100_10000seq_sample.tar.gz	4,773,706	gz-file	06/28/2017 05:...
100_10000seq_sampleC3.fastq	9,714,424	fastq-file	06/16/2015 10:...
100_10000seq_sampleC2.fastq	9,714,107	fastq-file	06/16/2015 10:...
100_10000seq_sampleC1.fastq	9,716,445	fastq-file	06/16/2015 10:...
100_10000seq_sampleB3.fastq	9,707,364	fastq-file	06/16/2015 10:...
100_10000seq_sampleB2.fastq	9,707,921	fastq-file	06/16/2015 10:...

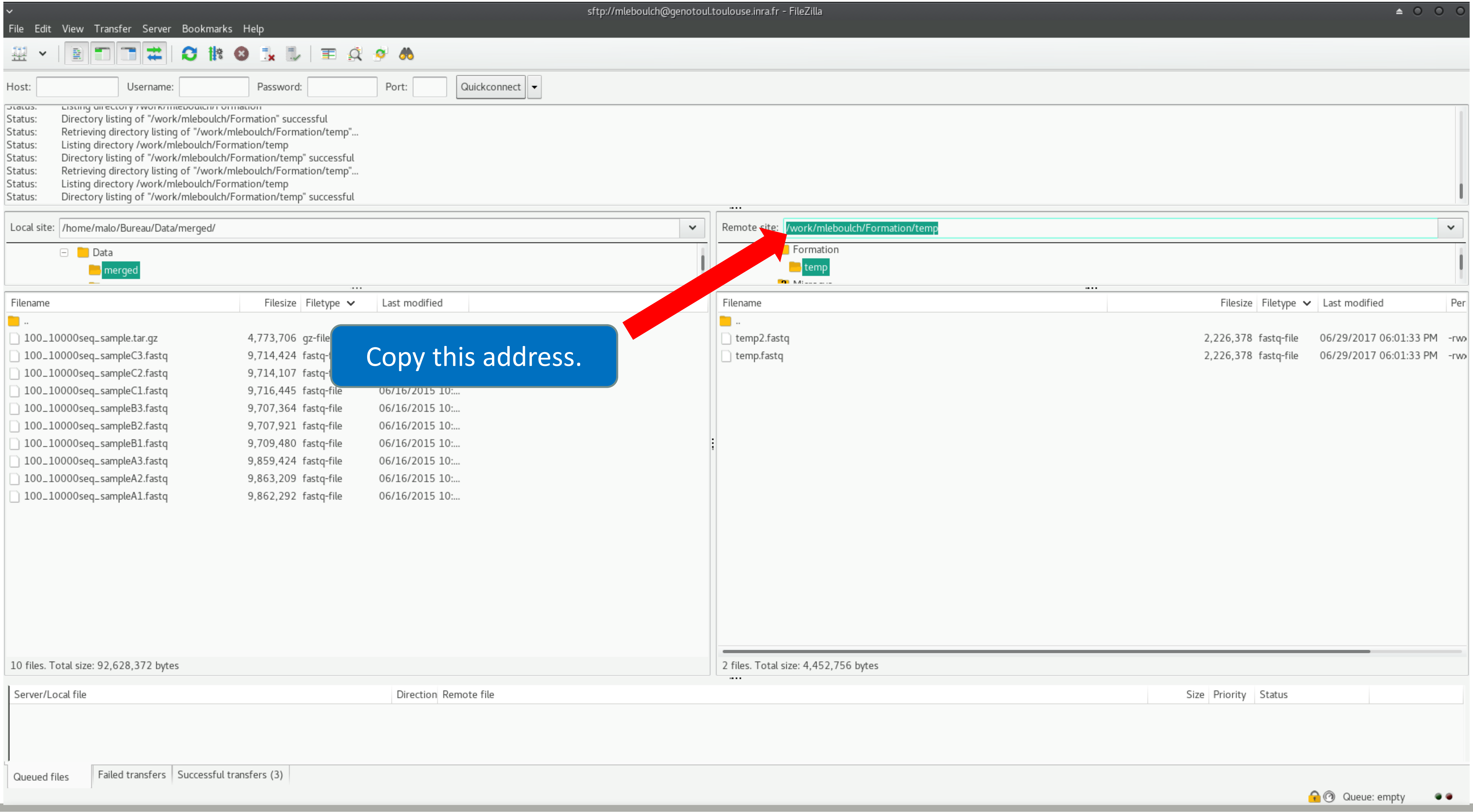
10 files. Total size: 92,628,372 bytes

Remote site: /work/mleboulch/Formation

Filename	Filesize	Filetype	Last modified	Per
..				
100_10000seq_sample.tar.gz	4,773,706	gz-file	06/29/2017 06:04:35 PM	-rw
temp		Directory	06/29/2017 06:01:32 PM	drw

1 file and 1 directory. Total size: 4,773,706 bytes

Double click on the temp directory.



```
Status:    Directory listing of "/work/mleboulch/Formation" successful
Status:    Retrieving directory listing of "/work/mleboulch/Formation/temp"...
Status:    Listing directory /work/mleboulch/Formation/temp
Status:    Directory listing of "/work/mleboulch/Formation/temp" successful
Status:    Retrieving directory listing of "/work/mleboulch/Formation/temp"...
Status:    Listing directory /work/mleboulch/Formation/temp
Status:    Directory listing of "/work/mleboulch/Formation/temp" successful
```

Local site: /home/malo/Bureau/Data/merged/

- Data
  - merged

Filename	Filesize	Filetype	Last modified
..			
100_10000seq_sample.tar.gz	4,773,706	gz-file	
100_10000seq_sampleC3.fastq	9,714,424	fastq-f	
100_10000seq_sampleC2.fastq	9,714,107	fastq-f	
100_10000seq_sampleC1.fastq	9,716,445	fastq-file	06/16/2015 10:...
100_10000seq_sampleB3.fastq	9,707,364	fastq-file	06/16/2015 10:...
100_10000seq_sampleB2.fastq	9,707,921	fastq-file	06/16/2015 10:...
100_10000seq_sampleB1.fastq	9,709,480	fastq-file	06/16/2015 10:...
100_10000seq_sampleA3.fastq	9,859,424	fastq-file	06/16/2015 10:...
100_10000seq_sampleA2.fastq	9,863,209	fastq-file	06/16/2015 10:...
100_10000seq_sampleA1.fastq	9,862,292	fastq-file	06/16/2015 10:...

10 files. Total size: 92,628,372 bytes

Remote site: /work/mleboulch/Formation/temp

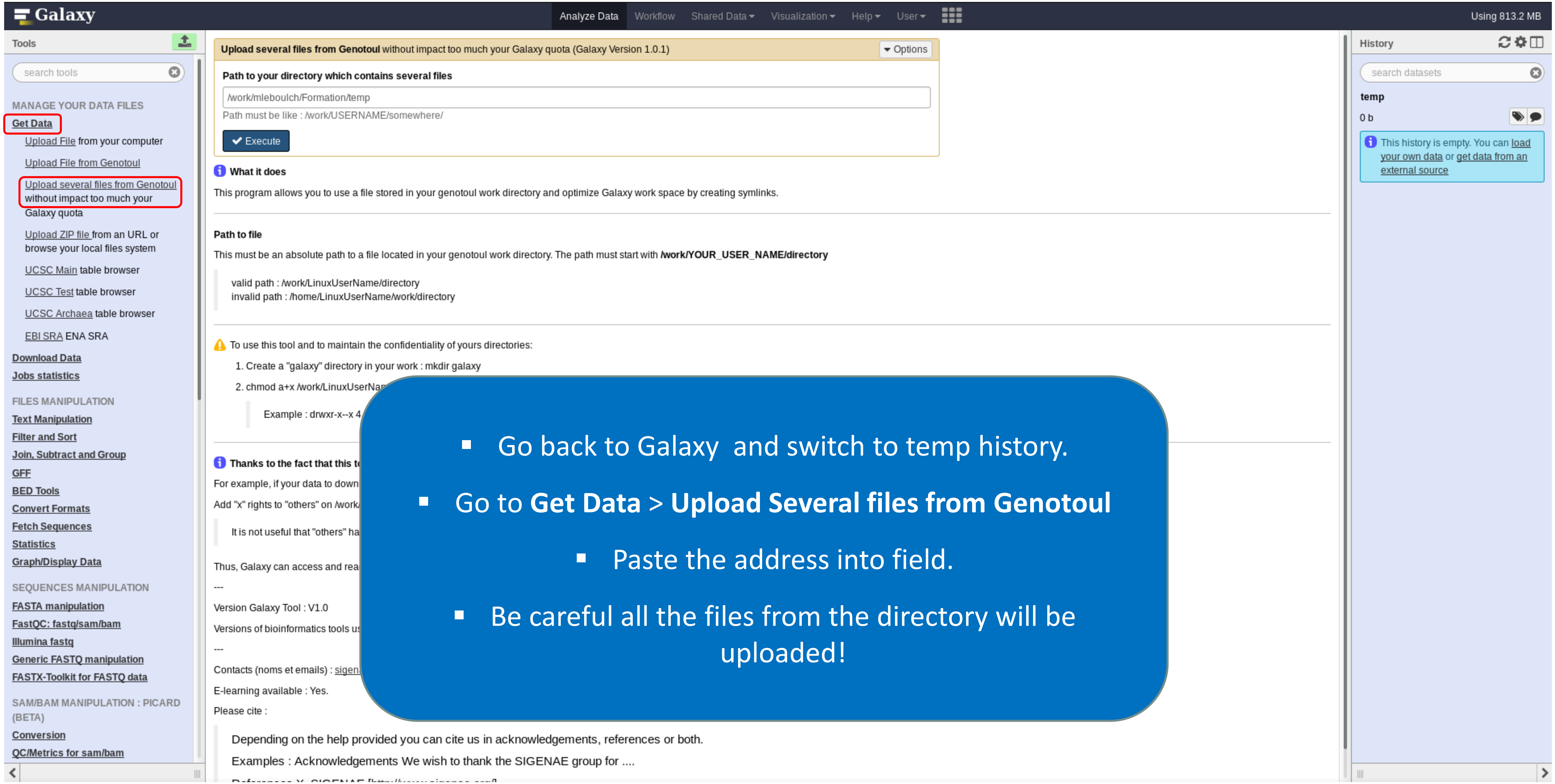
- Formation
  - temp

Filename	Filesize	Filetype	Last modified	Per
..				
temp2.fastq	2,226,378	fastq-file	06/29/2017 06:01:33 PM	-rw
temp.fastq	2,226,378	fastq-file	06/29/2017 06:01:33 PM	-rw

2 files. Total size: 4,452,756 bytes

Server/Local file	Direction	Remote file	Size	Priority	Status

Copy this address.



Upload several files from Genotoul without impact too much your Galaxy quota (Galaxy Version 1.0.1)

Path to your directory which contains several files

/work/mleboulch/Formation/temp

Path must be like : /work/USERNAME/somewhere/

This history is empty. You can load your own data or get data from an external source

What it does

This program allows you to use a file stored in your genotoul work directory and optimize Galaxy work space by creating symlinks.

Path to file

This must be an absolute path to a file located in your genotoul work directory. The path must start with **/work/YOUR_USER_NAME/directory**

valid path : /work/LinuxUserName/directory
invalid path : /home/LinuxUserName/work/directory

To use this tool and to maintain the confidentiality of yours directories:

1. Create a "galaxy" directory in your work : mkdir galaxy
2. chmod a+x /work/LinuxUserNa

Example : drwxr-x--x 4

Thanks to the fact that this t

For example, if your data to down

Add "x" rights to "others" on /work/

It is not useful that "others" ha

Thus, Galaxy can access and rea

Version Galaxy Tool : V1.0

Versions of bioinformatics tools us

Contacts (noms et emails) : sigen

E-learning available : Yes.

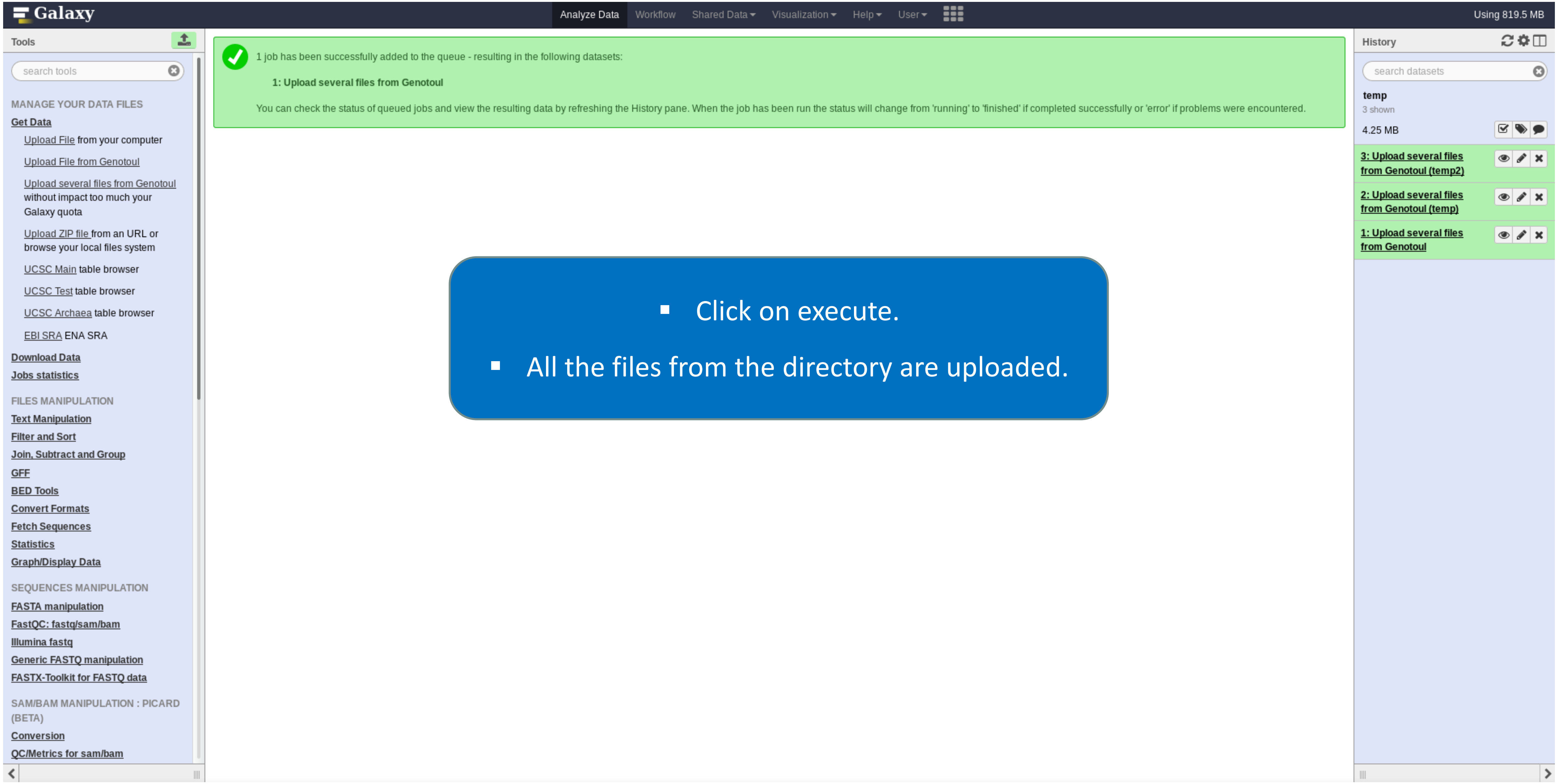
Please cite :

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

Examples : Acknowledgements We wish to thank the SIGENAE group for ....

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

- Go back to Galaxy and switch to temp history.
- Go to **Get Data** > **Upload Several files from Genotoul**
- Paste the address into field.
- Be careful all the files from the directory will be uploaded!



1 job has been successfully added to the queue - resulting in the following datasets:

1: Upload several files from Genotoul

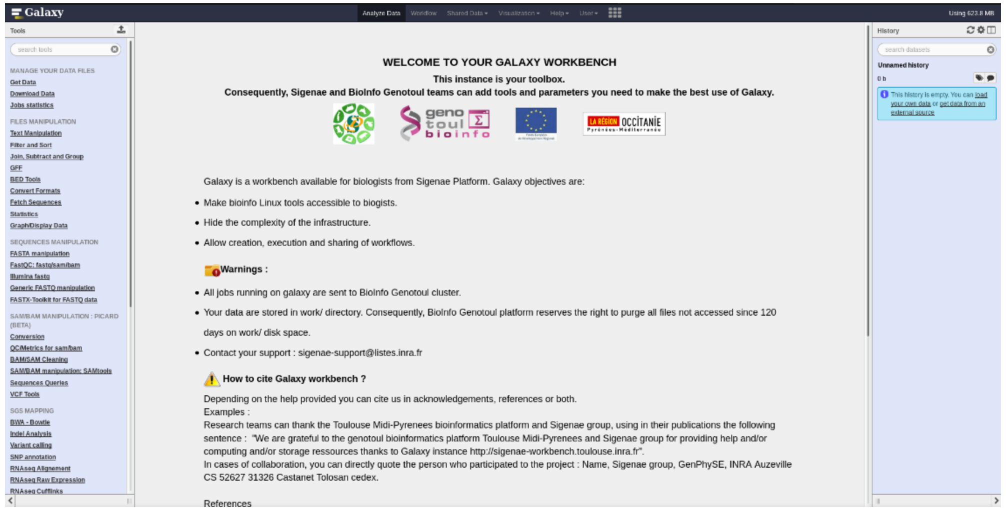
You can check the status of queued jobs and view the resulting data by refreshing the History pane. When the job has been run the status will change from 'running' to 'finished' if completed successfully or 'error' if problems were encountered.

temp
3 shown
4.25 MB

3: Upload several files from Genotoul (temp2)
2: Upload several files from Genotoul (temp)
1: Upload several files from Genotoul

- Click on execute.
- All the files from the directory are uploaded.

Upload by Genotoul



This history is empty. You can load your own data or get data from an external source

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

1TB of space

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

- This method allows you to have more disk space and to upload bigger files.
- Files not used in the last 180 days can be purged at any time.

Share a History

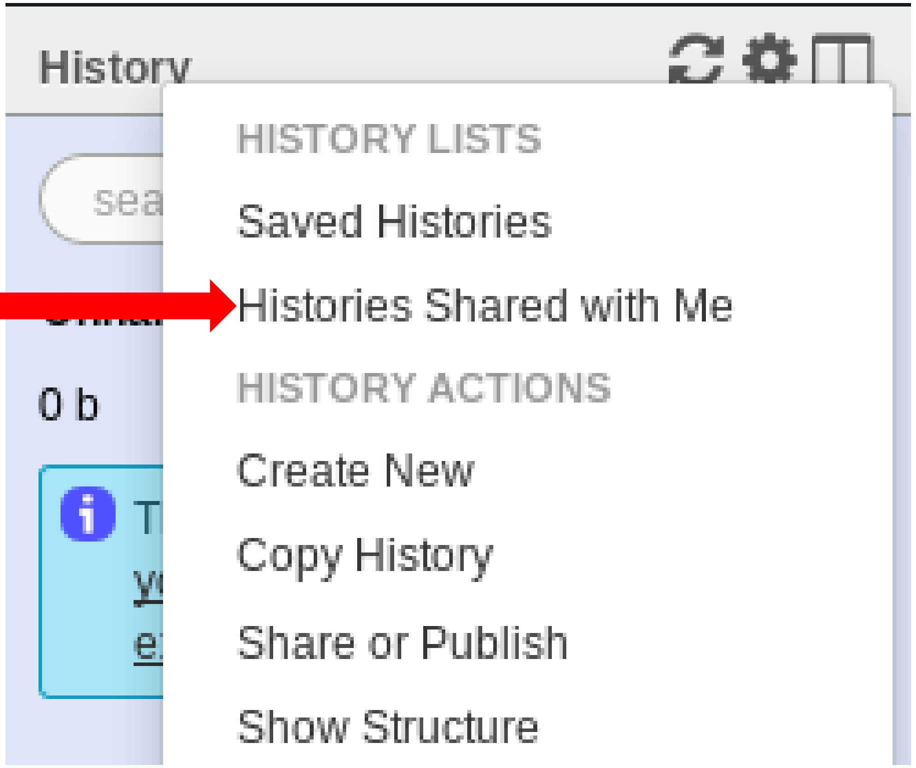
Share a history

- 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.

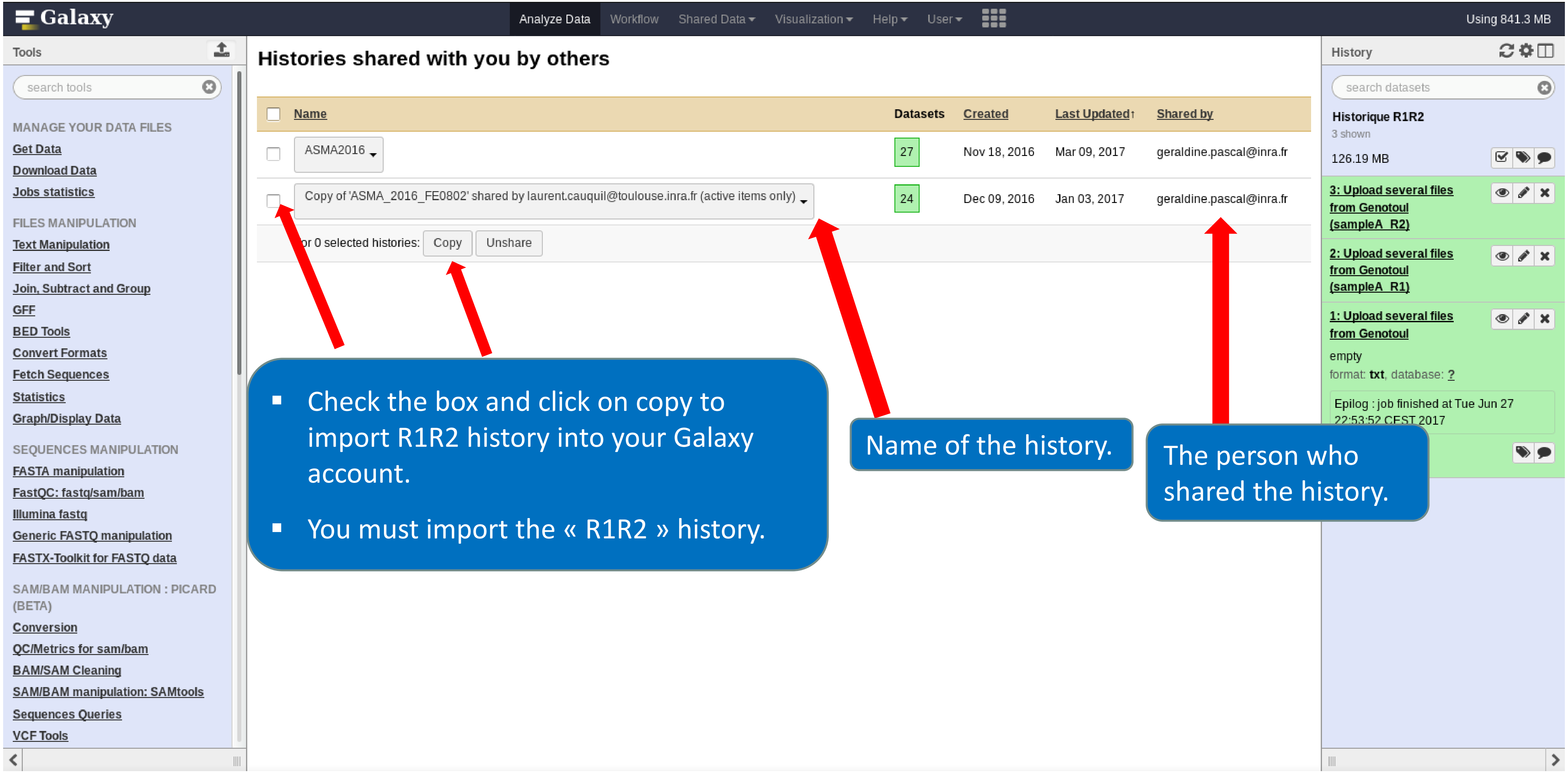
Your Turn!

IMPORT A SHARED HISTORY TO YOUR ACCOUNT

Import a shared history



Click on the wheel and click here.



Histories shared with you by others

☐	Name	Datasets	Created	Last Updated↑	Shared by
☐	ASMA2016	27	Nov 18, 2016	Mar 09, 2017	geraldine.pascal@inra.fr
☐	Copy of 'ASMA_2016_FE0802' shared by laurent.cauquil@toulouse.inra.fr (active items only)	24	Dec 09, 2016	Jan 03, 2017	geraldine.pascal@inra.fr

or 0 selected histories: Copy Unshare

Historique R1R2

3 shown

126.19 MB

3: Upload several files from Genotoul (sampleA_R2)

2: Upload several files from Genotoul (sampleA_R1)

1: Upload several files from Genotoul

empty

format: txt, database: ?

Epilog : job finished at Tue Jun 27 22:53:52 CEST 2017

- Check the box and click on copy to import R1R2 history into your Galaxy account.
- You must import the « R1R2 » history.

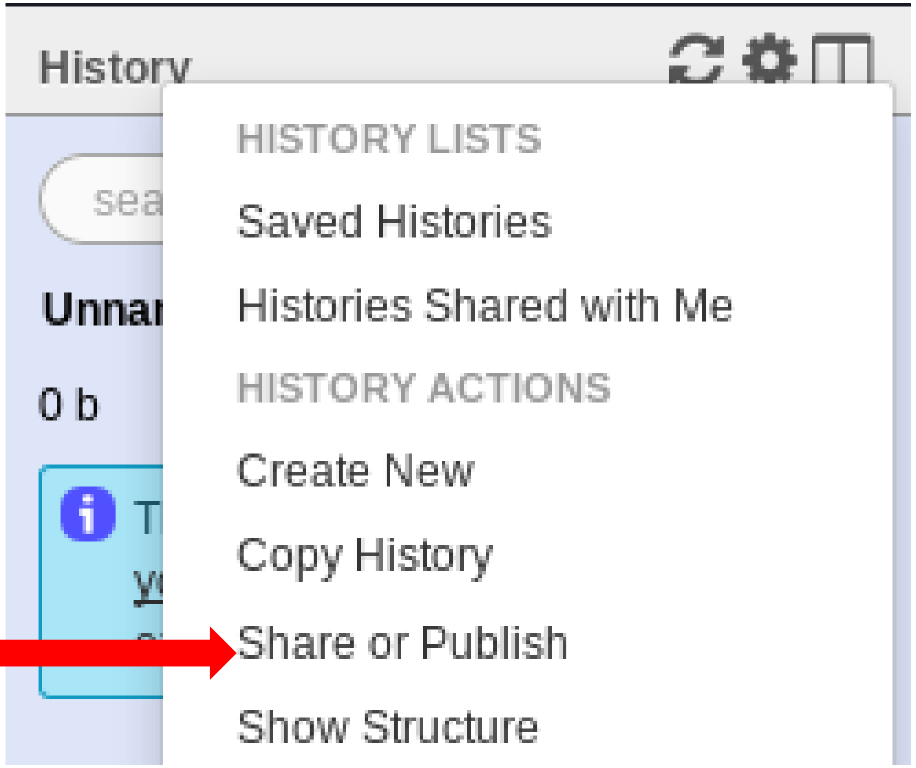
Name of the history.

The person who shared the history.

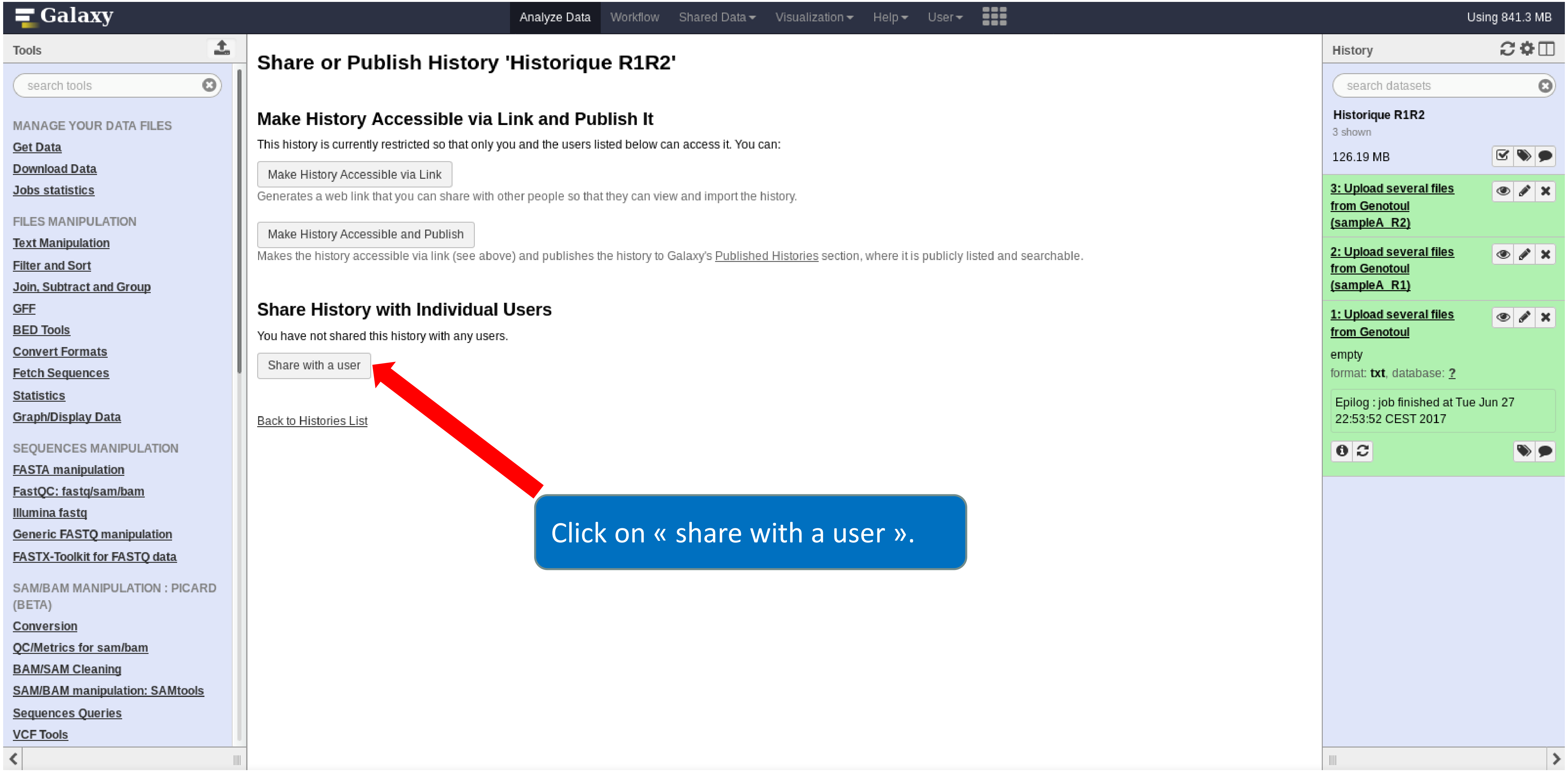
Your Turn!

SHARE A HISTORY WITH YOUR NEIGHBOUR

Share a history



- Switch to the history you want to share.
- Click on the wheel and click here.



Share or Publish History 'Historique R1R2'

Make History Accessible via Link and Publish It

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

Make History Accessible via Link

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

Make History Accessible and Publish

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

Share History with Individual Users

You have not shared this history with any users.

Share with a user

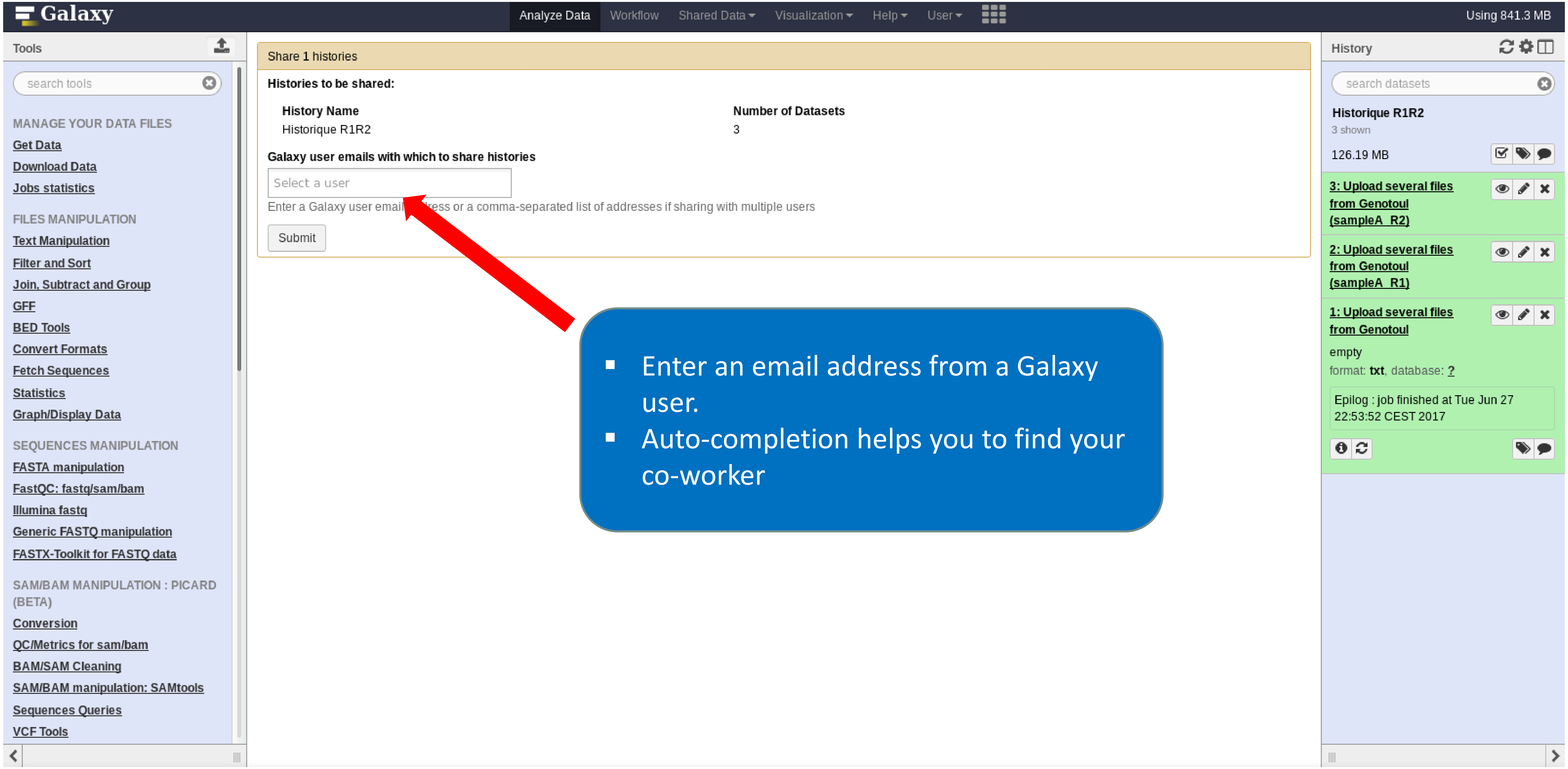
Back to Histories List

empty

format: txt, database: ?

Epilog : job finished at Tue Jun 27 22:53:52 CEST 2017

Click on « share with a user ».



Share 1 histories

Histories to be shared:

History Name	**Number of Datasets**
Historique R1R2	3

Galaxy user emails with which to share histories

Select a user

Enter a Galaxy user email address or a comma-separated list of addresses if sharing with multiple users

Historique R1R2
3 shown
126.19 MB

3: Upload several files from Genotoul (sampleA_R2)

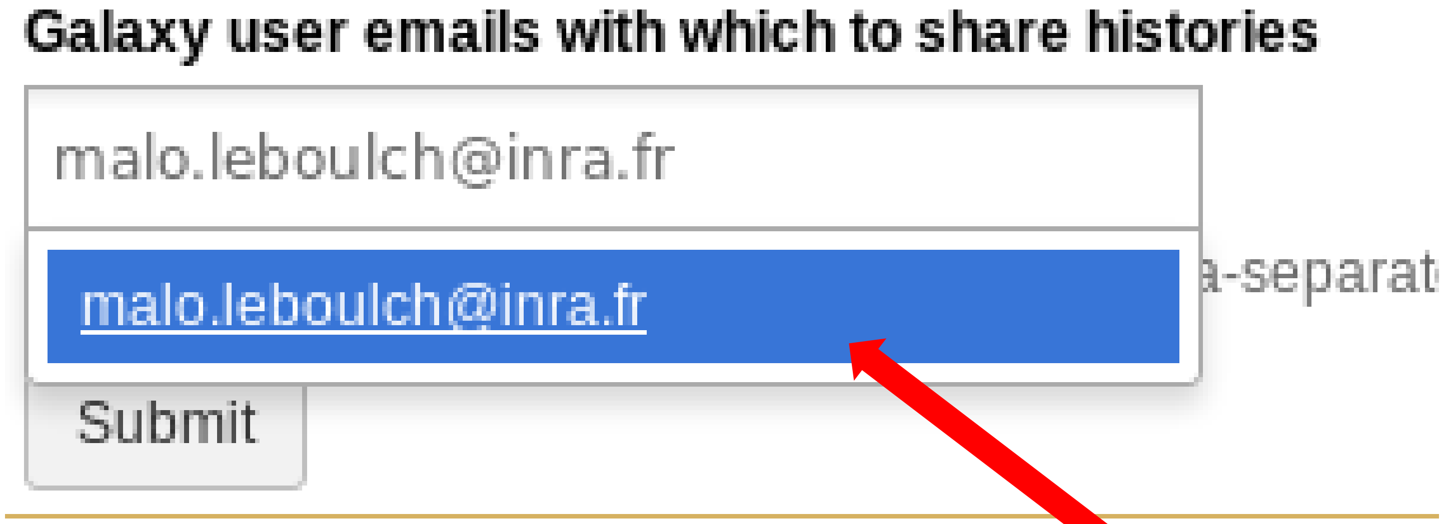
2: Upload several files from Genotoul (sampleA_R1)

1: Upload several files from Genotoul

empty
format: txt, database: ?

Epilog : job finished at Tue Jun 27 22:53:52 CEST 2017

- Enter an email address from a Galaxy user.
- Auto-completion helps you to find your co-worker



Galaxy user emails with which to share histories

malo.leboulch@inra.fr

malo.leboulch@inra.fr

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

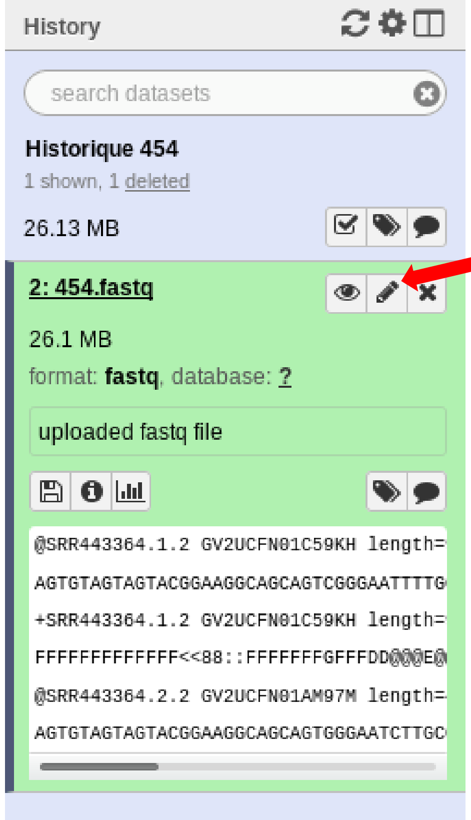
Manipulate datasets

Your Turn!

RENAME A DATASET

Rename a dataset

- Switch to 454 history or ITS history



Historique 454

1 shown, 1 deleted

26.13 MB

2: 454.fastq

26.1 MB

format: **fastq**, database: ?

uploaded fastq file

```
@SRR443364.1.2 GV2UCFN01C59KH length=
AGTGTAGTAGTACGGAAGGCAGCAGTCGGGAATTTTG
+SRR443364.1.2 GV2UCFN01C59KH length=
FFFFFFFFFFFF<<88::FFFFFFFGFFFDD@@@E@
@SRR443364.2.2 GV2UCFN01AM97M length=
AGTGTAGTAGTACGGAAGGCAGCAGTGGGAATCTTGC
```

Historique ITS

1 show

156.91 MB

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

156.9 MB

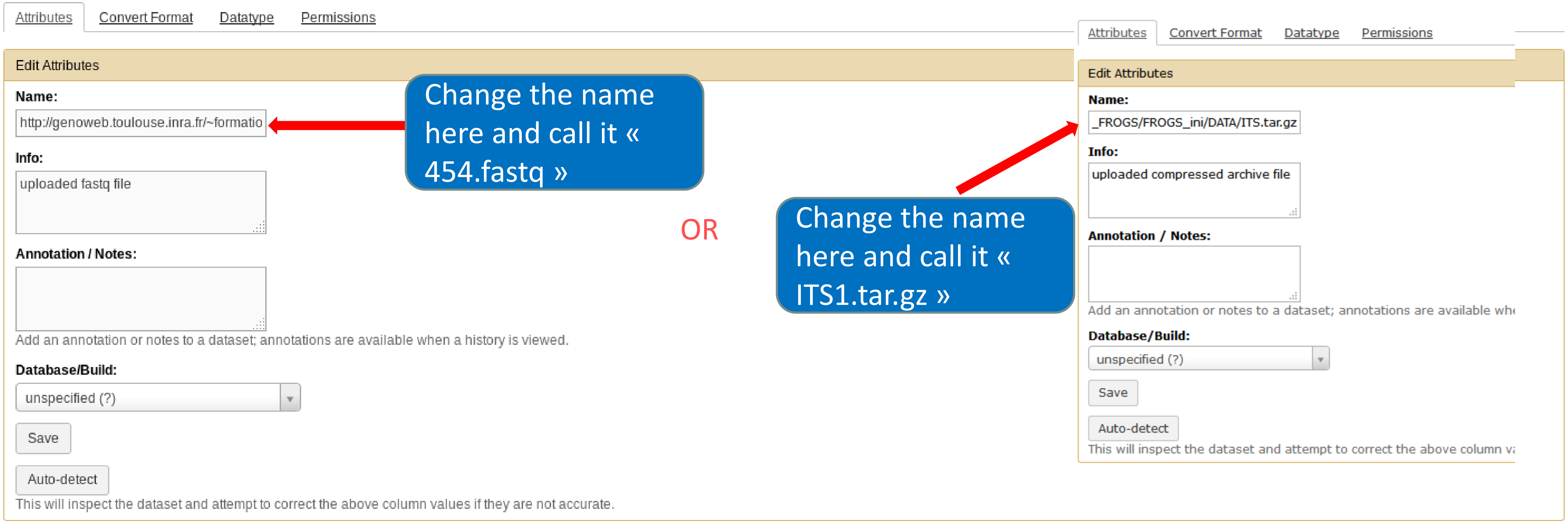
format: **tar**, database: ?

uploaded compressed archive file

```
Compressed binary file
```

Click here to display attributes and change the name.

Rename a dataset



Edit Attributes

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

Info: uploaded fastq file

Annotation / Notes:

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

Database/Build: unspecified (?)

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

Change the name here and call it « 454.fastq »

OR

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

Edit Attributes

Name: _FROGS/FROGS_ini/DATA/ITS.tar.gz

Info: uploaded compressed archive file

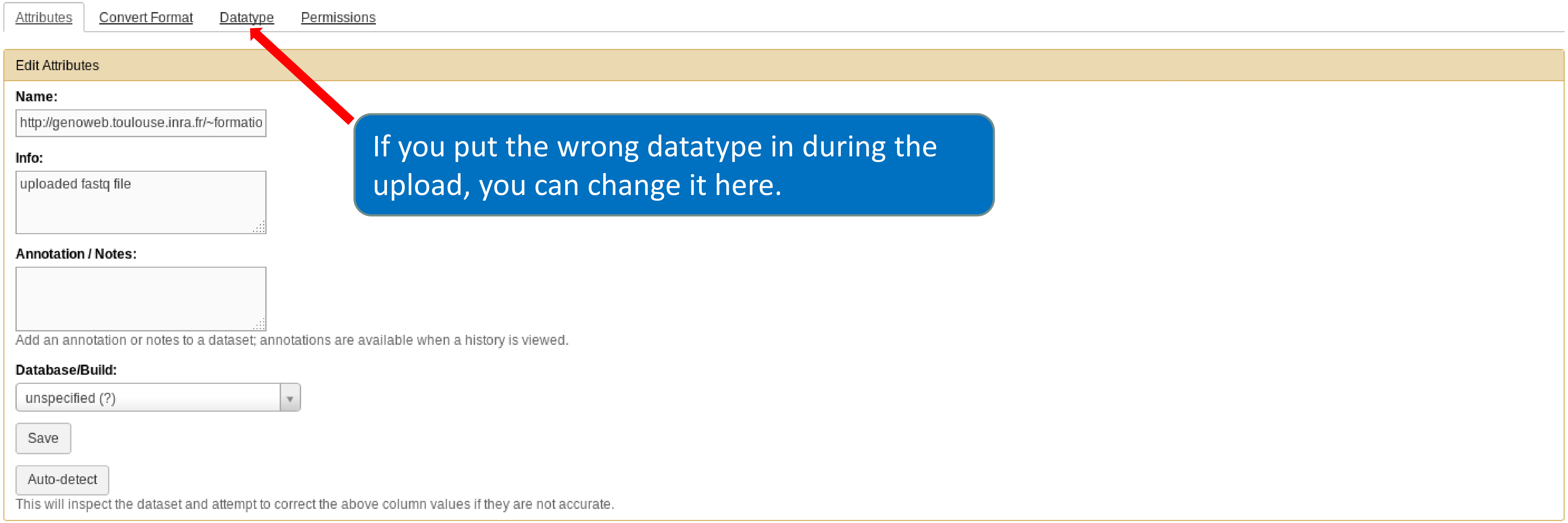
Annotation / Notes:

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

Database/Build: unspecified (?)

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

Rename a dataset



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

Info: uploaded fastq file

Annotation / Notes:

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

Database/Build: unspecified (?)

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

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

Rename a dataset

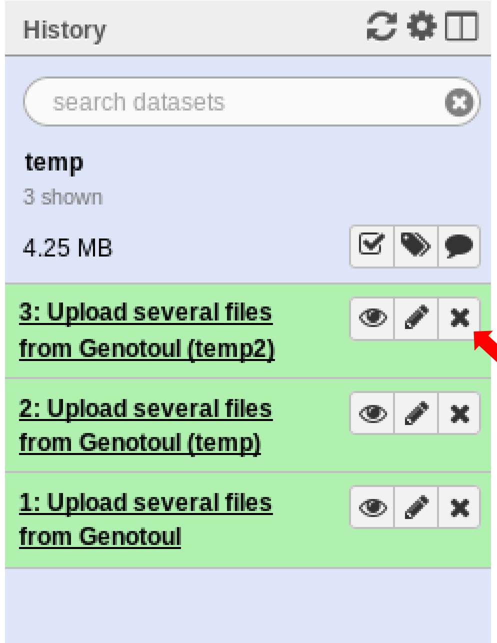
- Do the same with the merged history:
  - Switch to the merged history.
  - Change the name of the file to «100_10000seq_sample.tar.gz ».

Your Turn!

DELETE A DATASET

Delete a dataset

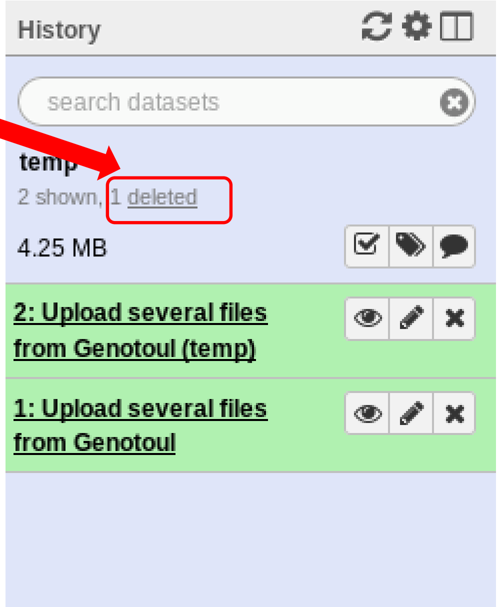
- Switch to temp history.



Click here to delete a dataset.

Delete a dataset

Click here.



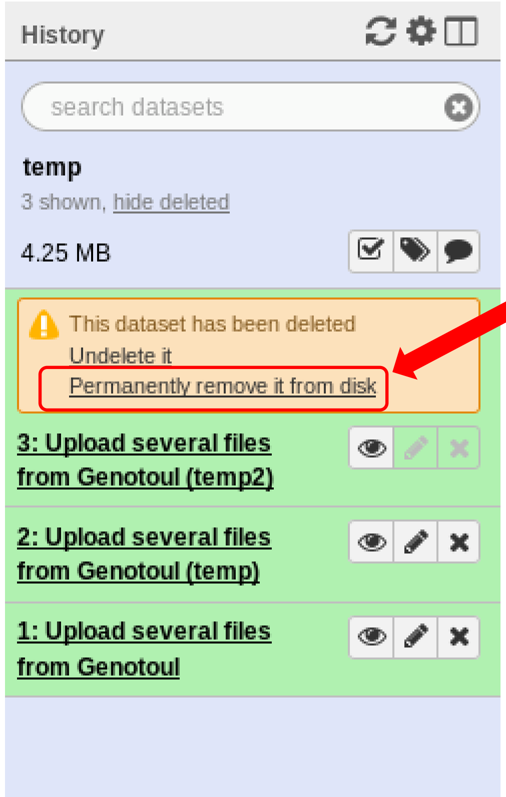
temp

2 shown, 1 deleted

4.25 MB

2: Upload several files from Genotoul (temp)

1: Upload several files from Genotoul



temp

3 shown, hide deleted

4.25 MB

This dataset has been deleted

Undelete it

Permanently remove it from disk

3: Upload several files from Genotoul (temp2)

2: Upload several files from Genotoul (temp)

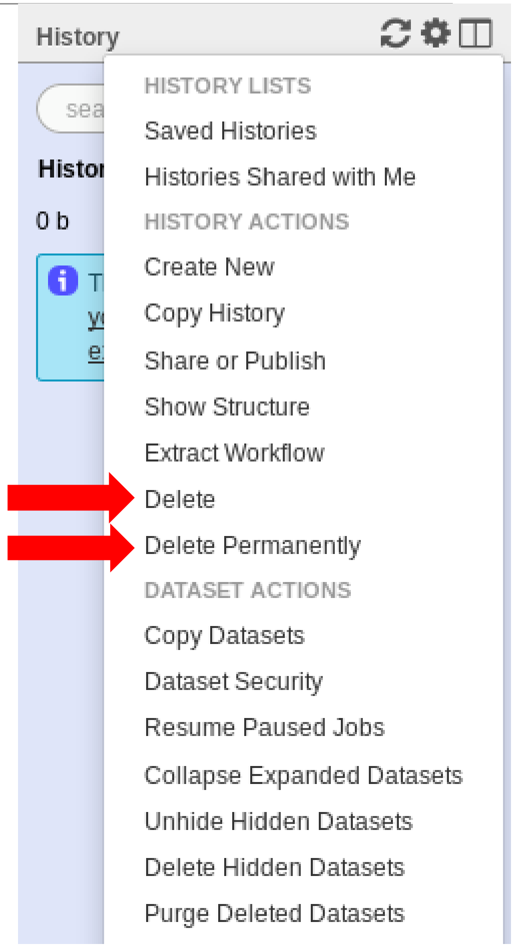
1: Upload several files from Genotoul

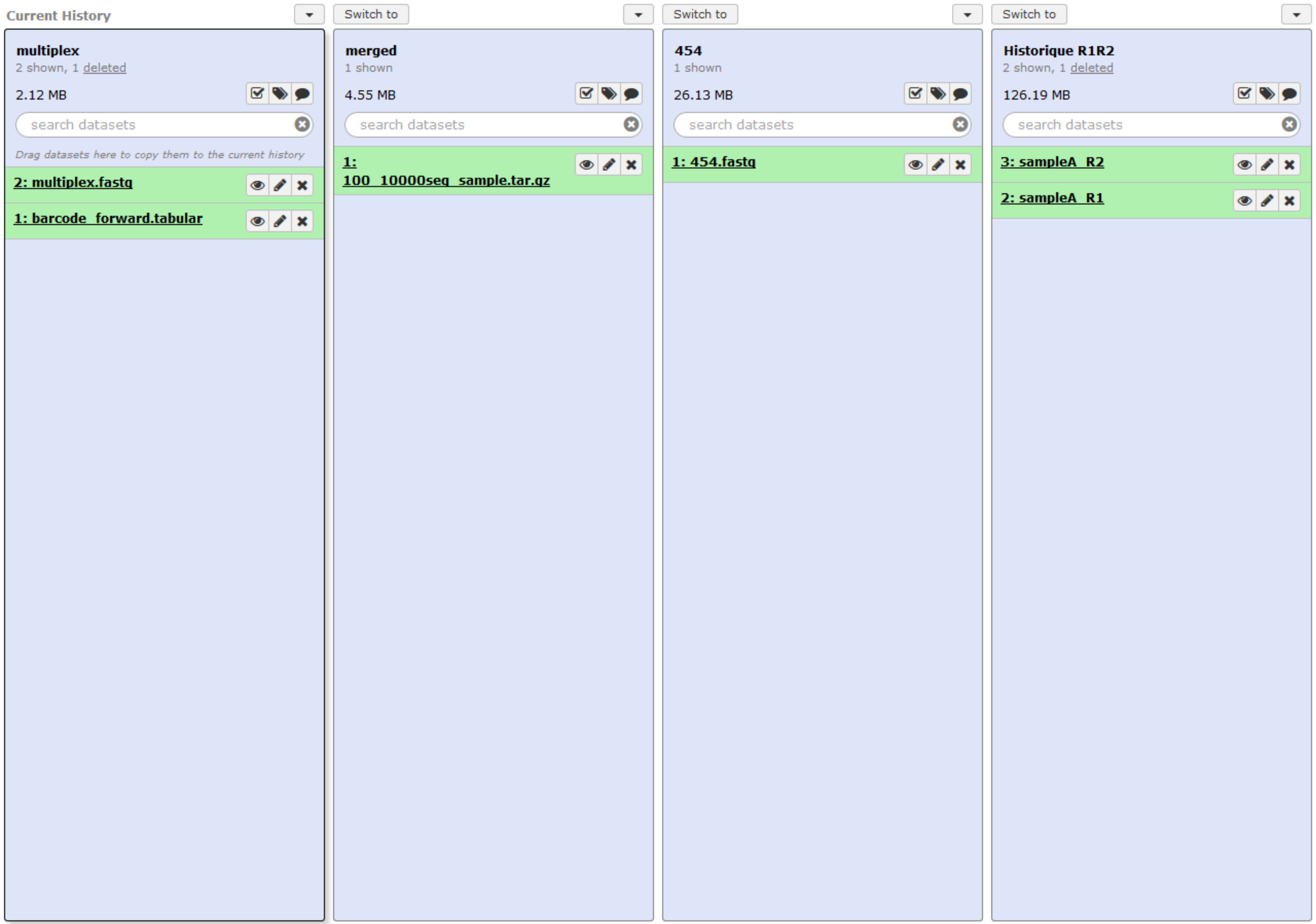
Delete this dataset permanently.

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

How to delete a history?

- Stay in the temp history.
- Click on the wheel.
- Click on delete.
- A deleted item on Galaxy is recoverable.
- To definitively delete it:
  click on « Delete Permanently ».





multiplex
2 shown, 1 deleted
2.12 MB
Drag datasets here to copy them to the current history
2: multiplex.fastq
1: barcode_forward.tabular

merged
1 shown
4.55 MB
1: 100_10000seq_sample.tar.gz

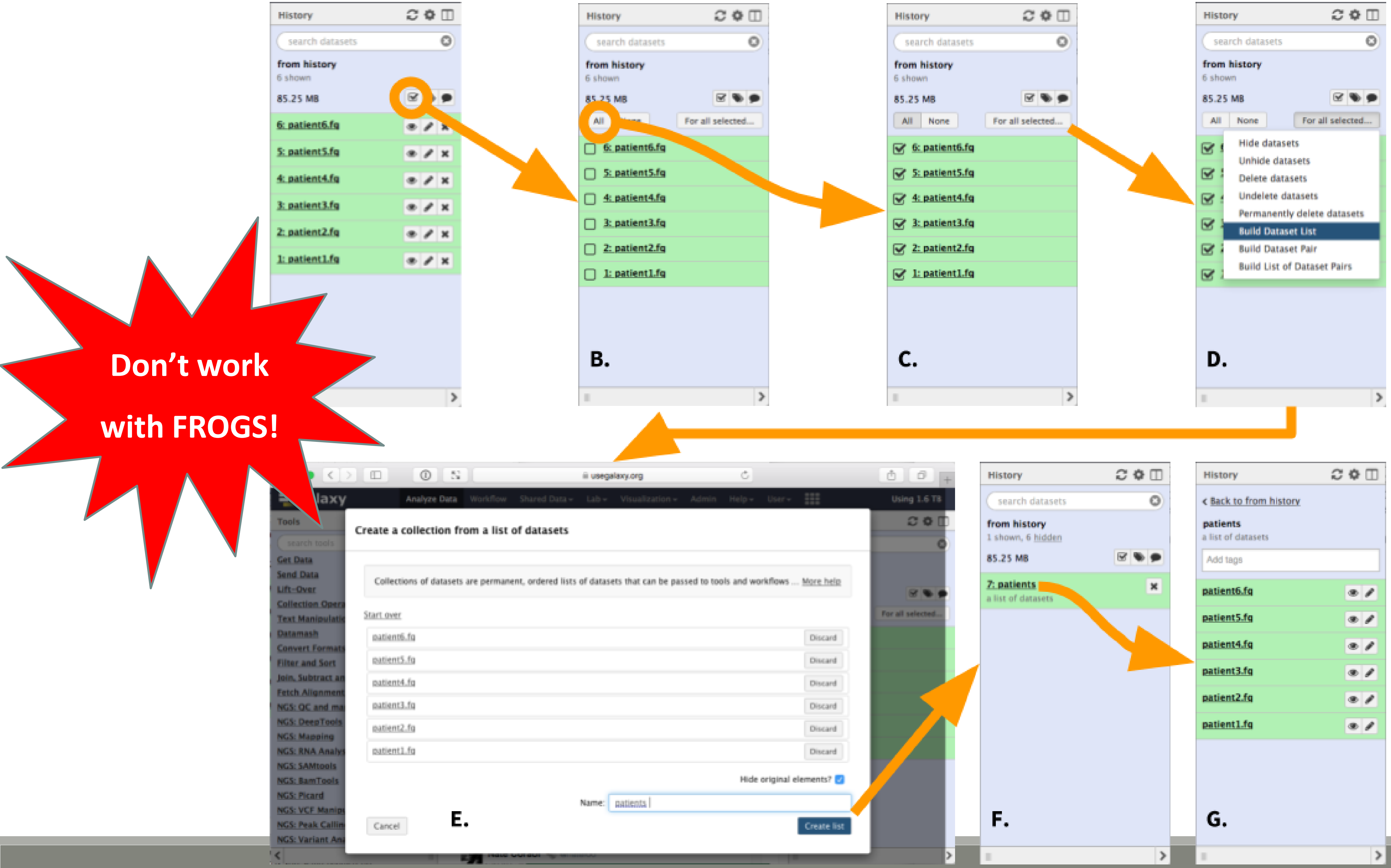
454
1 shown
26.13 MB
1: 454.fastq

Historique R1R2
2 shown, 1 deleted
126.19 MB
3: sampleA_R2
2: sampleA_R1

One word about data collections

- Used to perform the same analysis on multiple files.
- Gather multiple datasets in one collection.

Don’t work with FROGS!



Don't work

with FROGS!

Create a collection from a list of datasets

Collections of datasets are permanent, ordered lists of datasets that can be passed to tools and workflows ... More help

Hide original elements? ☑

Name: patients

Others tools available on galaxy

- Text Manipulation
- Filter and Sort
- FASTA manipulation

Galaxy support

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

How to cite Galaxy?

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

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

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

How to cite Genotoul Galaxy workbench?

- Research teams can thank the Toulouse 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 resources 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.