Formation FROGS juillet 2019

Géraldine Pascal, Laurent Cauquil

▶ **To cite this version:**

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

Submitted on 5 Jun 2020

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

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

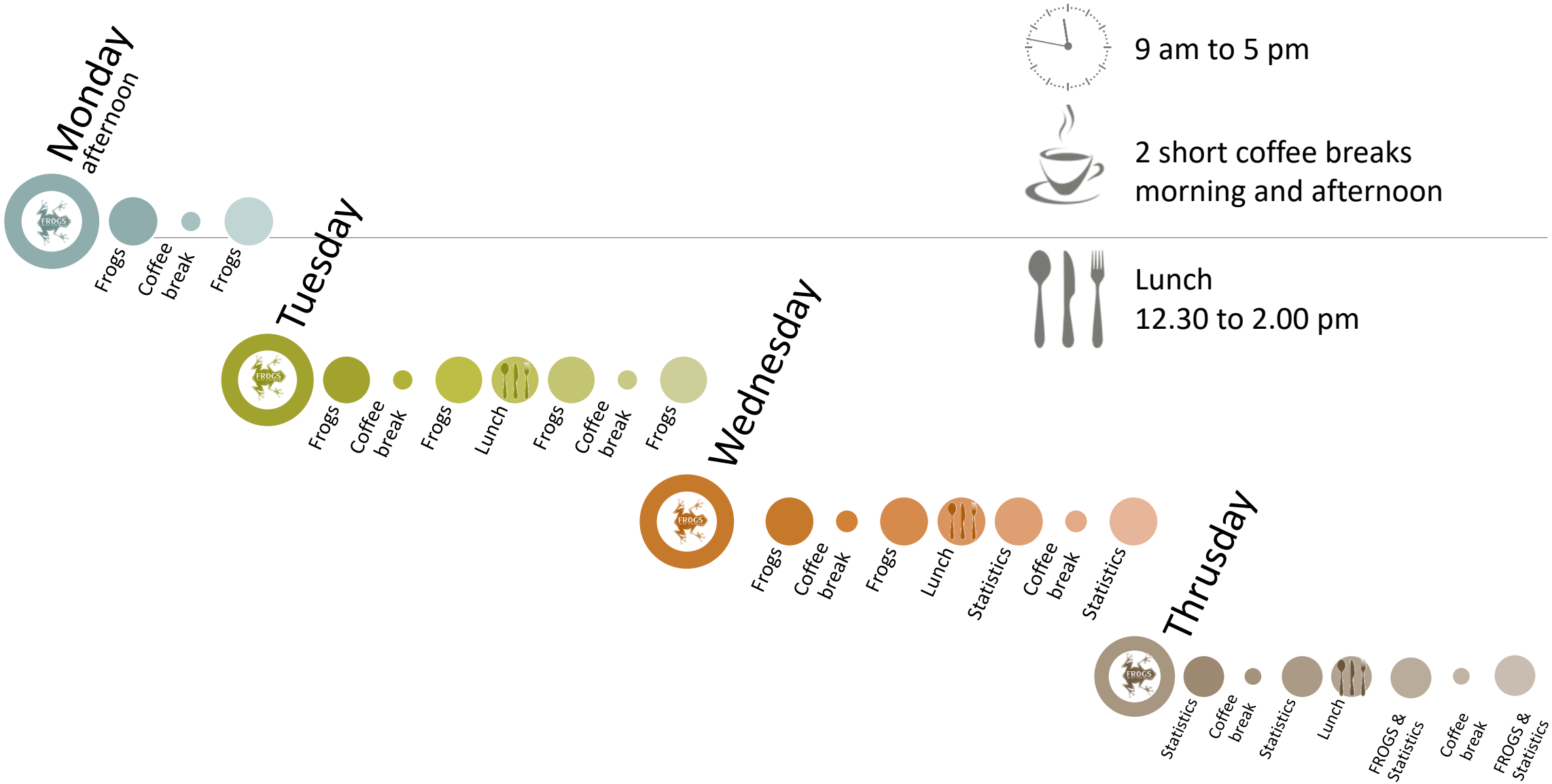
MALO LE BOULCH, SARAH MAMAN, GÉRALDINE PASCAL

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

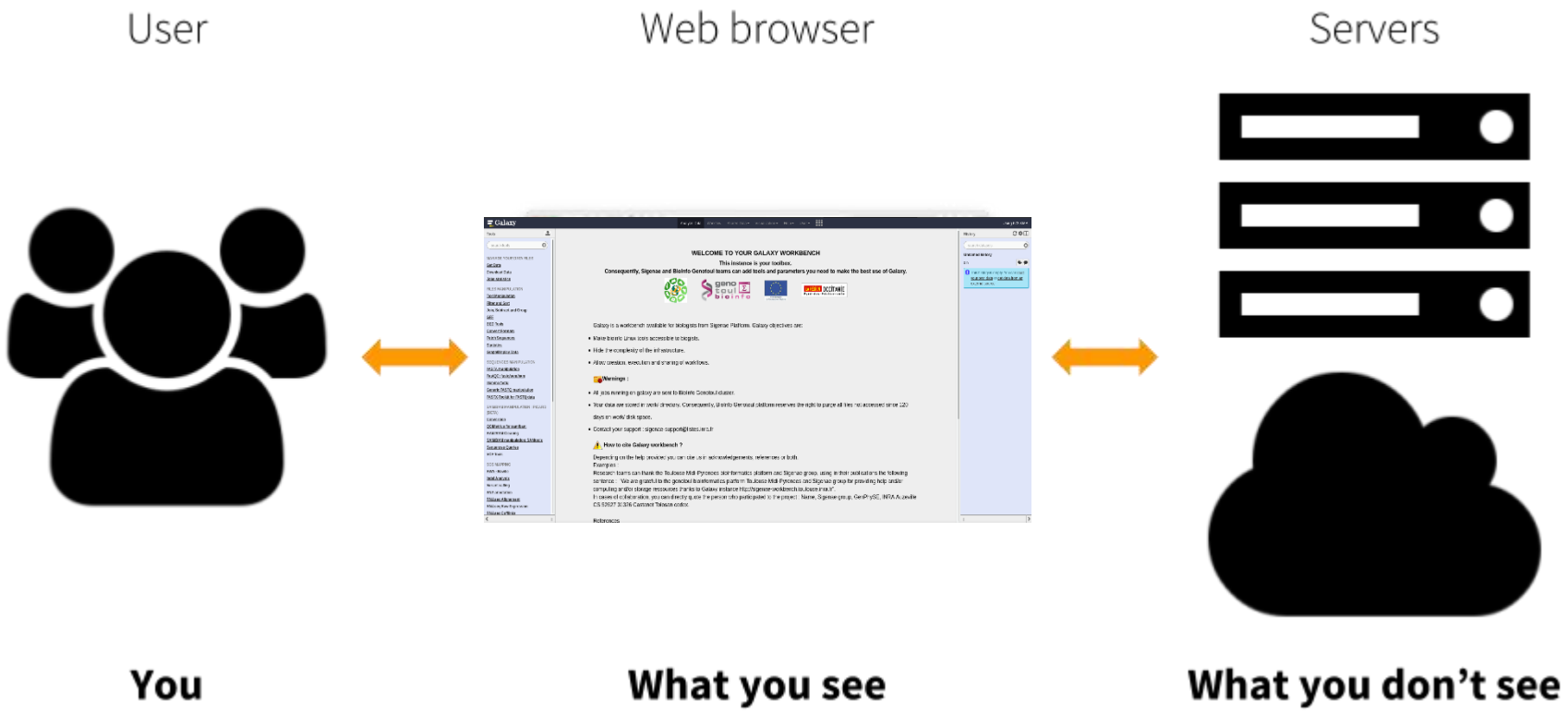
- Learn the basics of Galaxy
- Being independent when using it
- Prepare the datasets for FROGS formation

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



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?

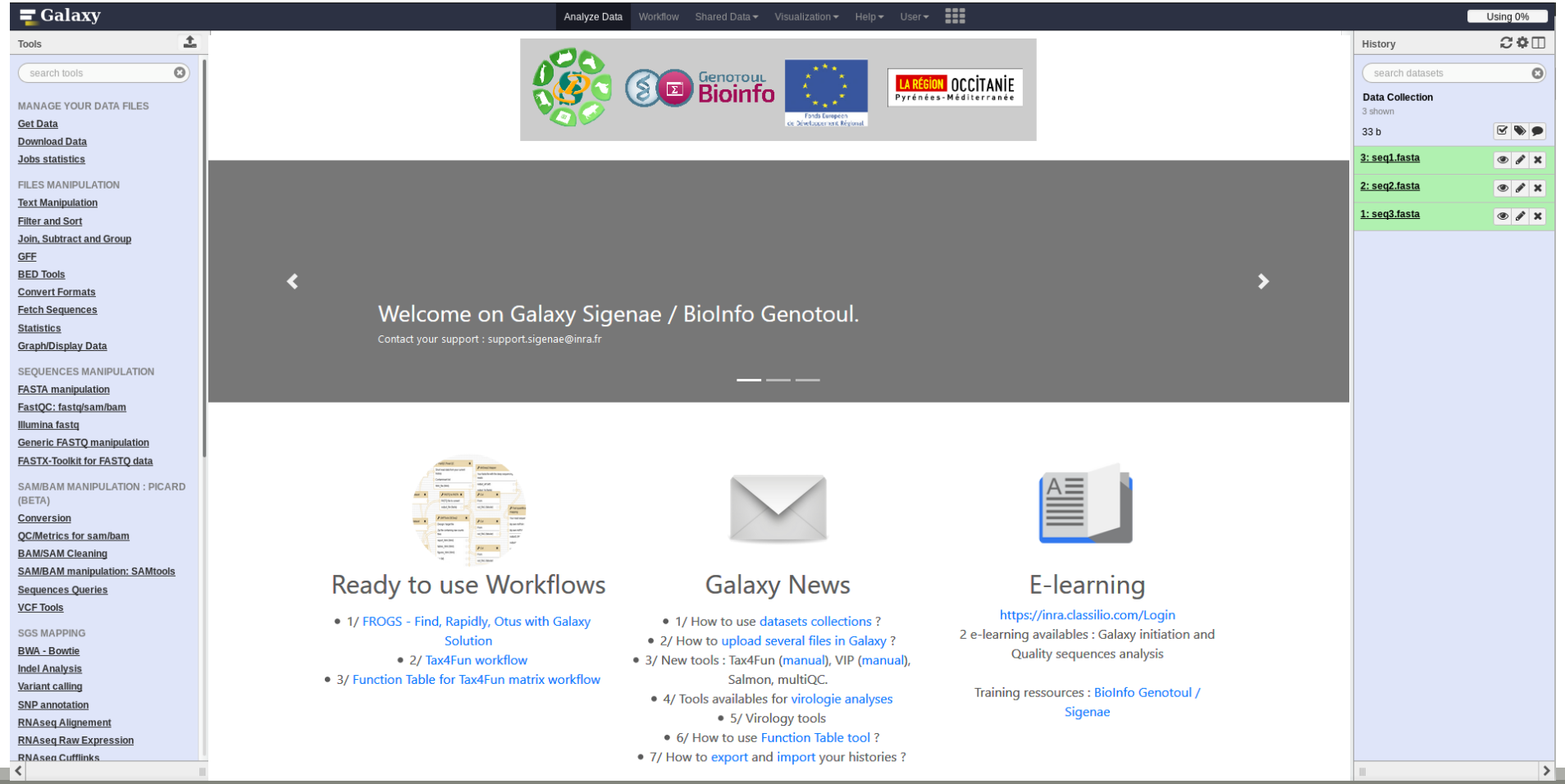
- Our Galaxy platform is http://sigenae-workbench.toulouse.inra.fr/galaxy/

⚠ 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

Exemple of 2 INRA Galaxy platforms



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.

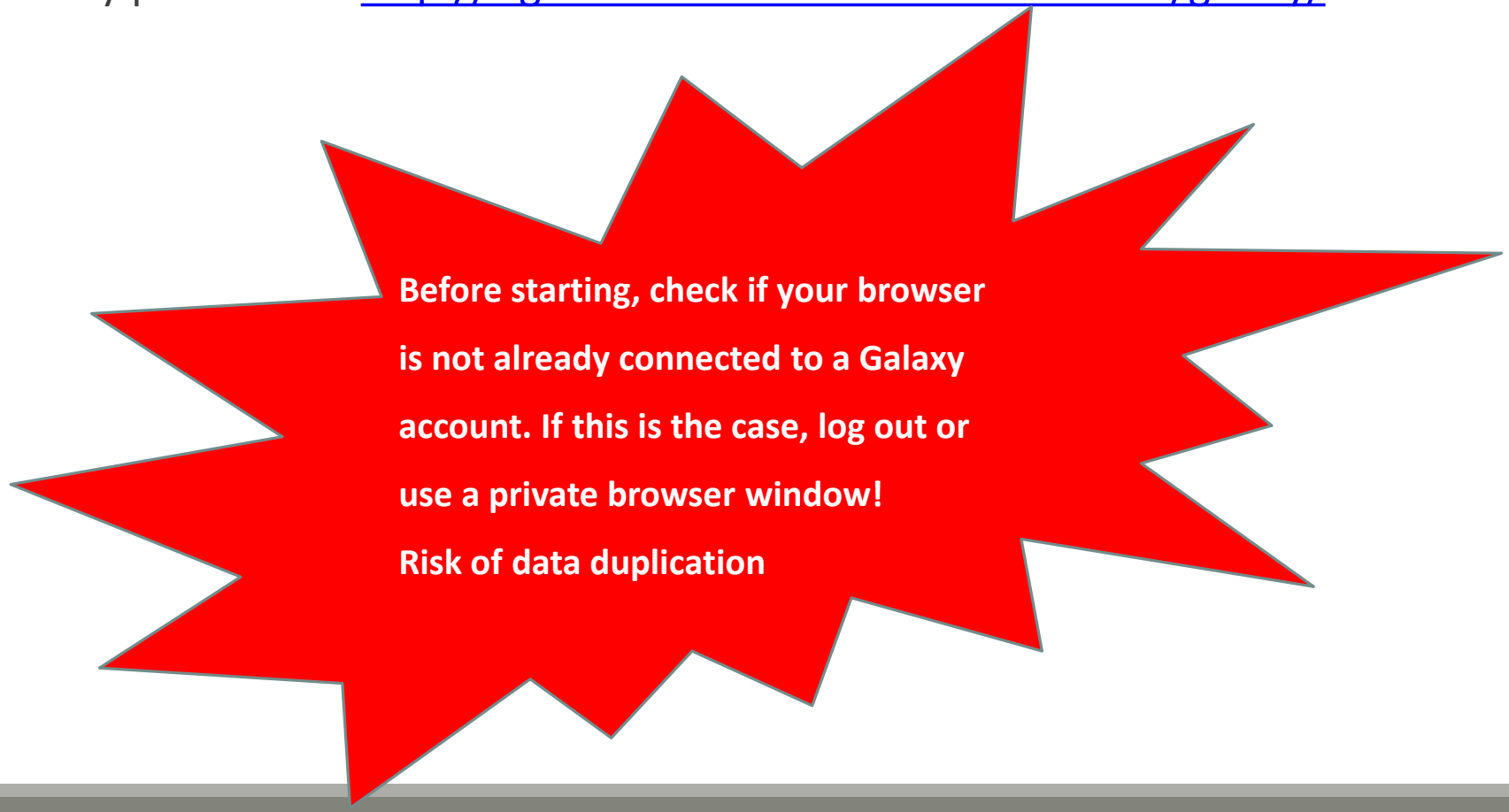
This account will save all the work done during this week and your future work.

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/

Exercise

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

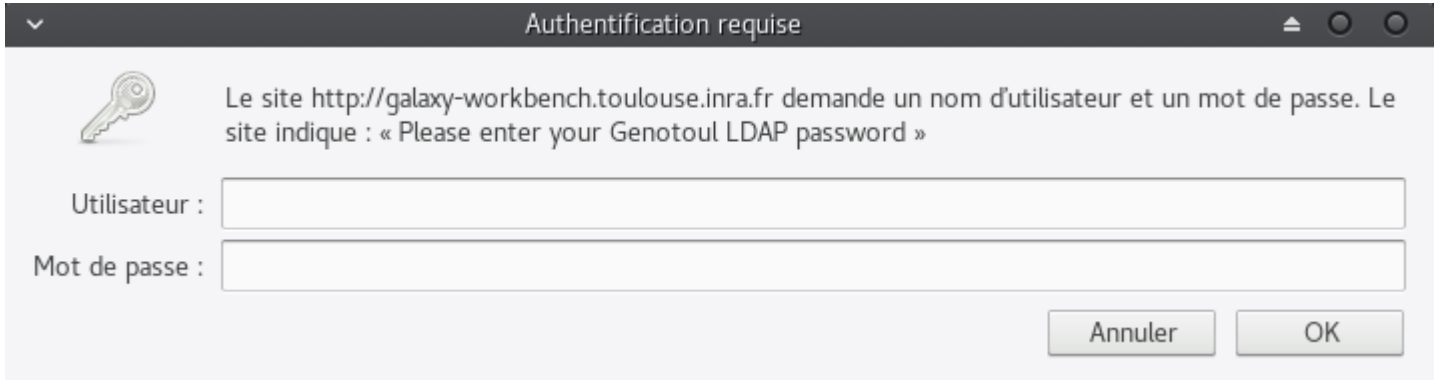


Exercise

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

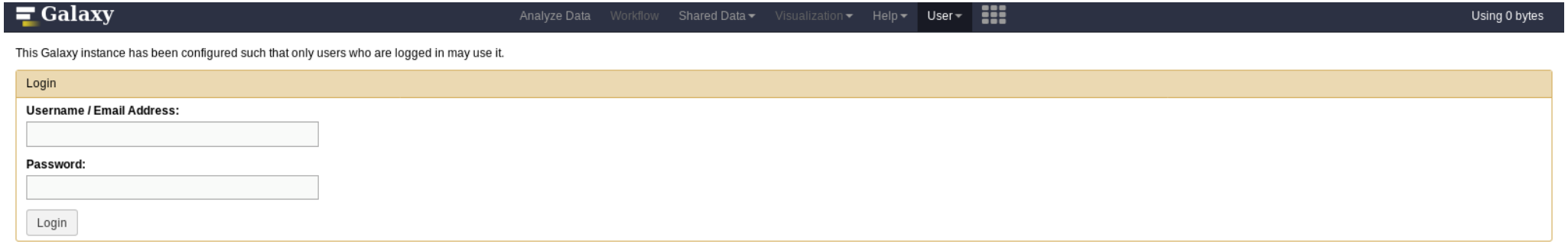
Be careful, to fully login you must enter your credentials twice:

- The first time in this pop-up window:



Exercise

- And a second time, in the browser:



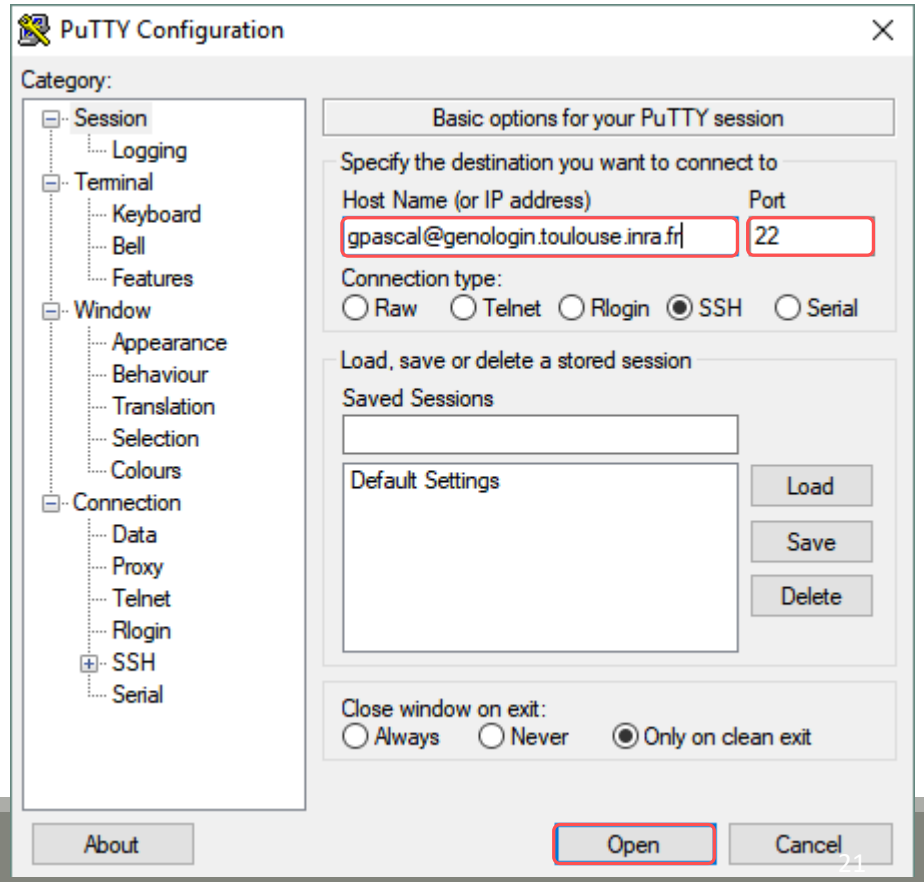
Change your password

- 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

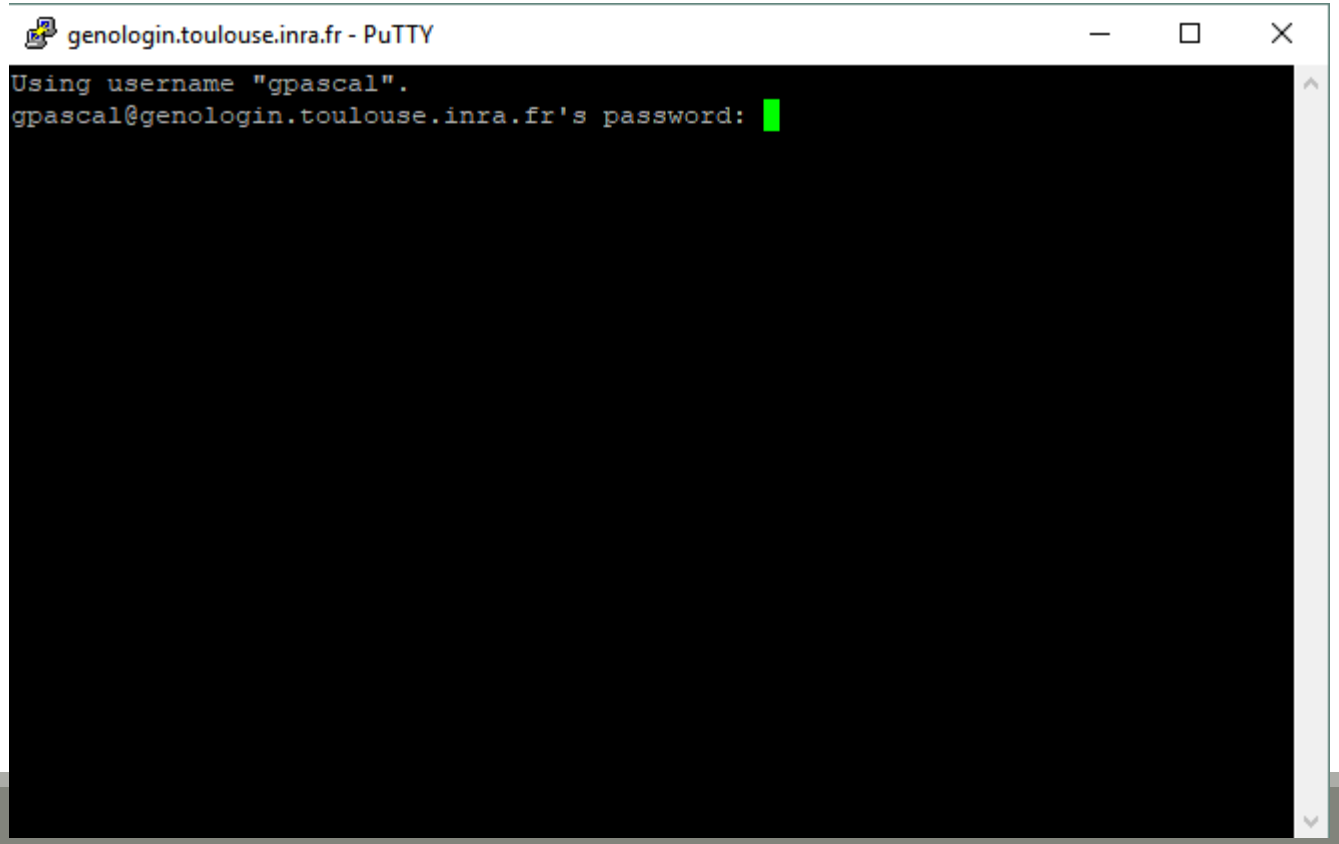
- Launch PuTTY.
- In the following window, you must enter your host name which is:

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



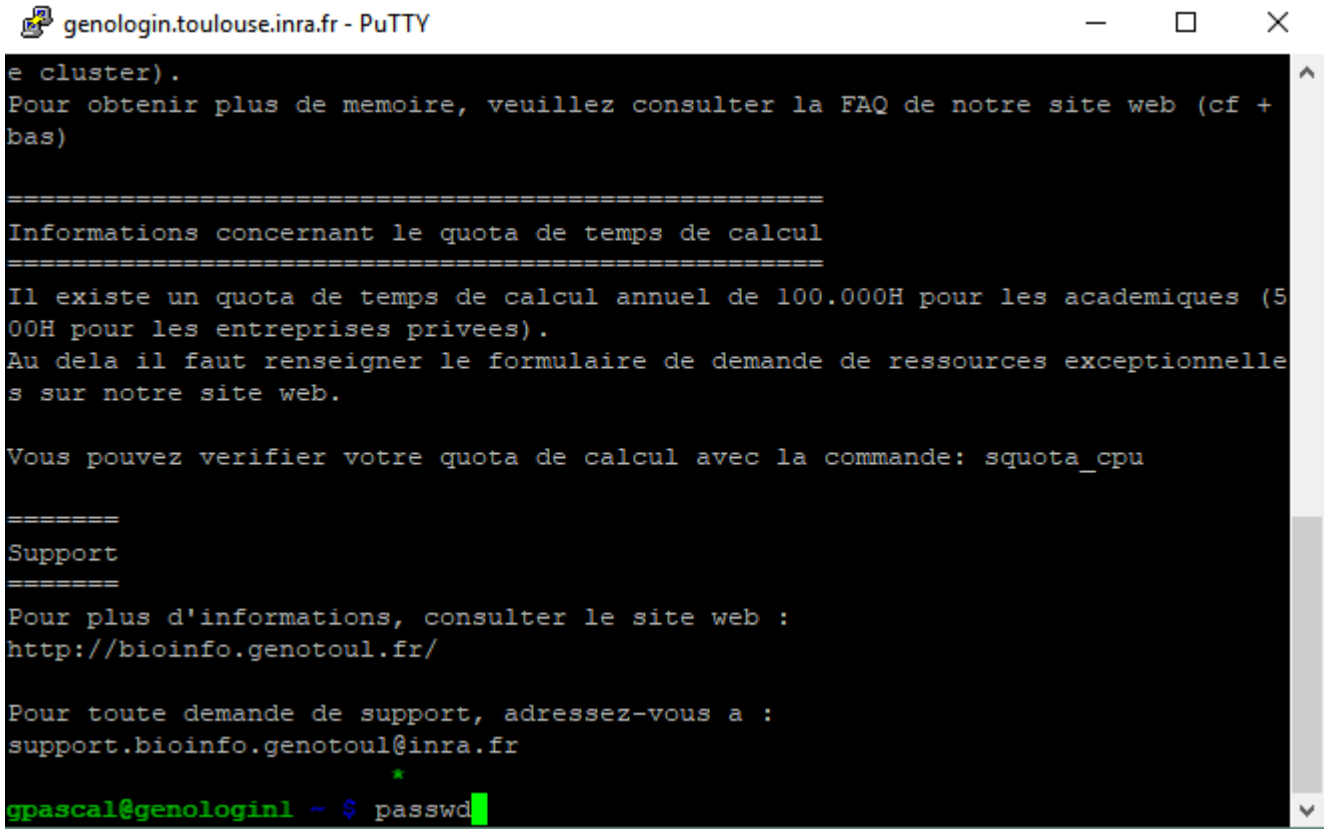
Change your password

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



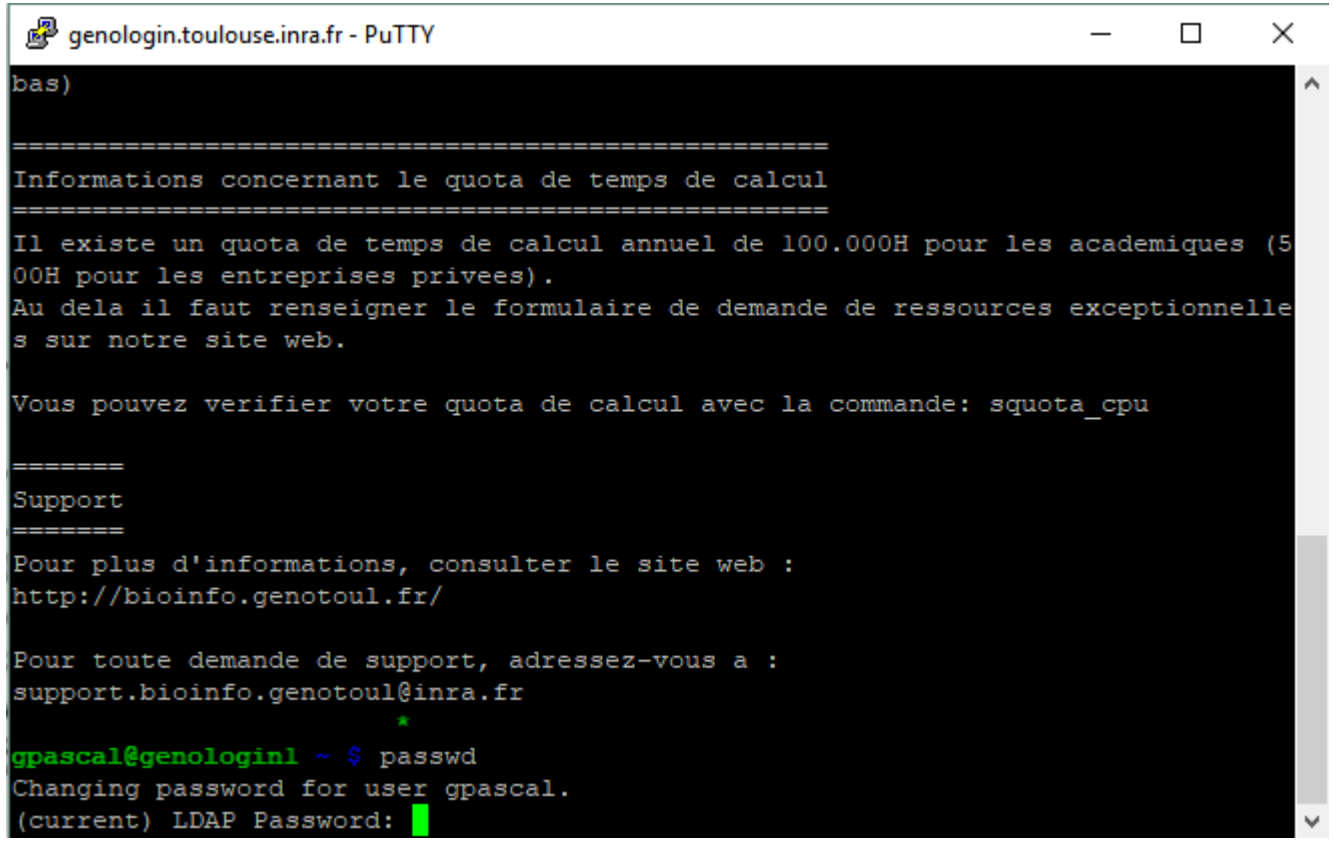
Change your password

- Type « passwd » and hit « Enter ».



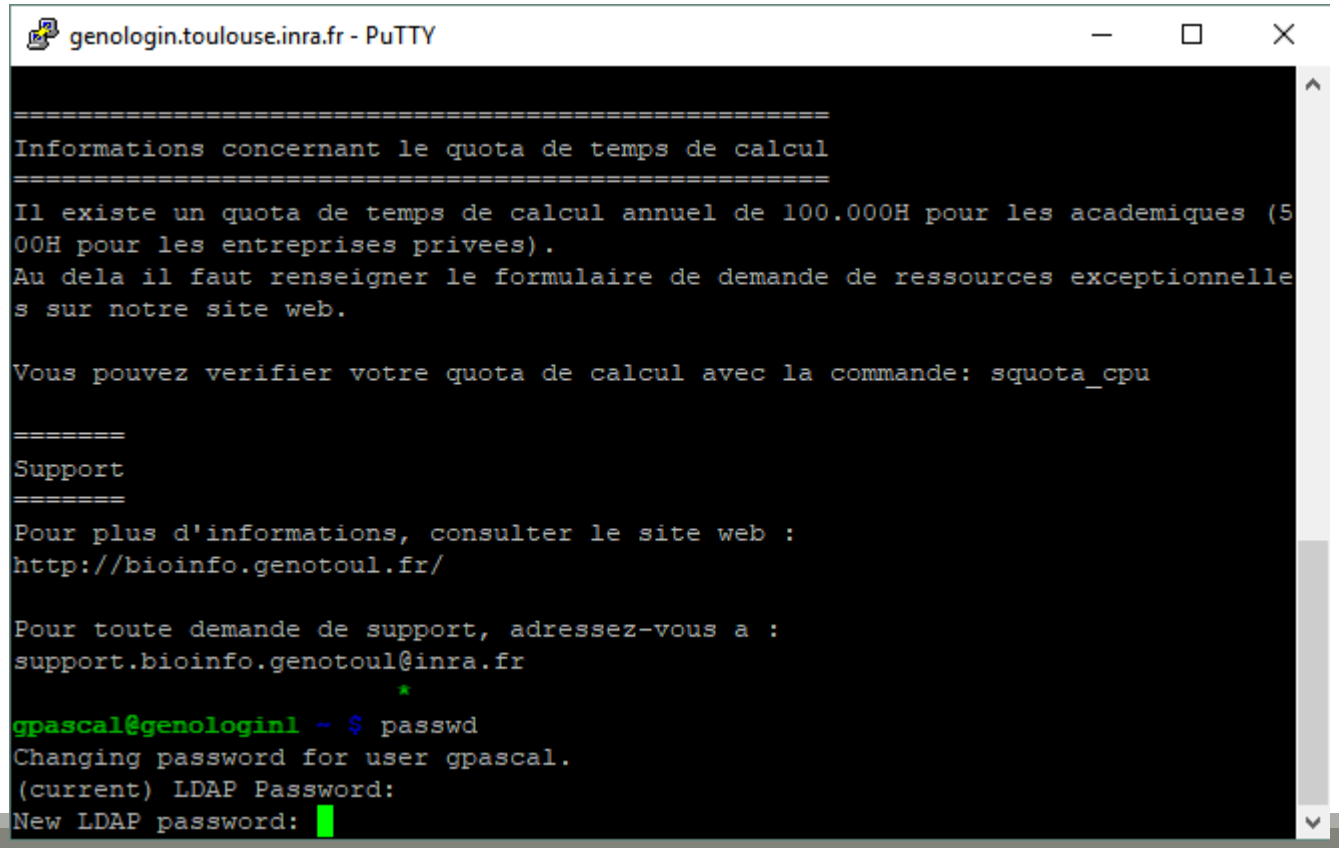
Change your password

- Enter your current password and hit « Enter ».



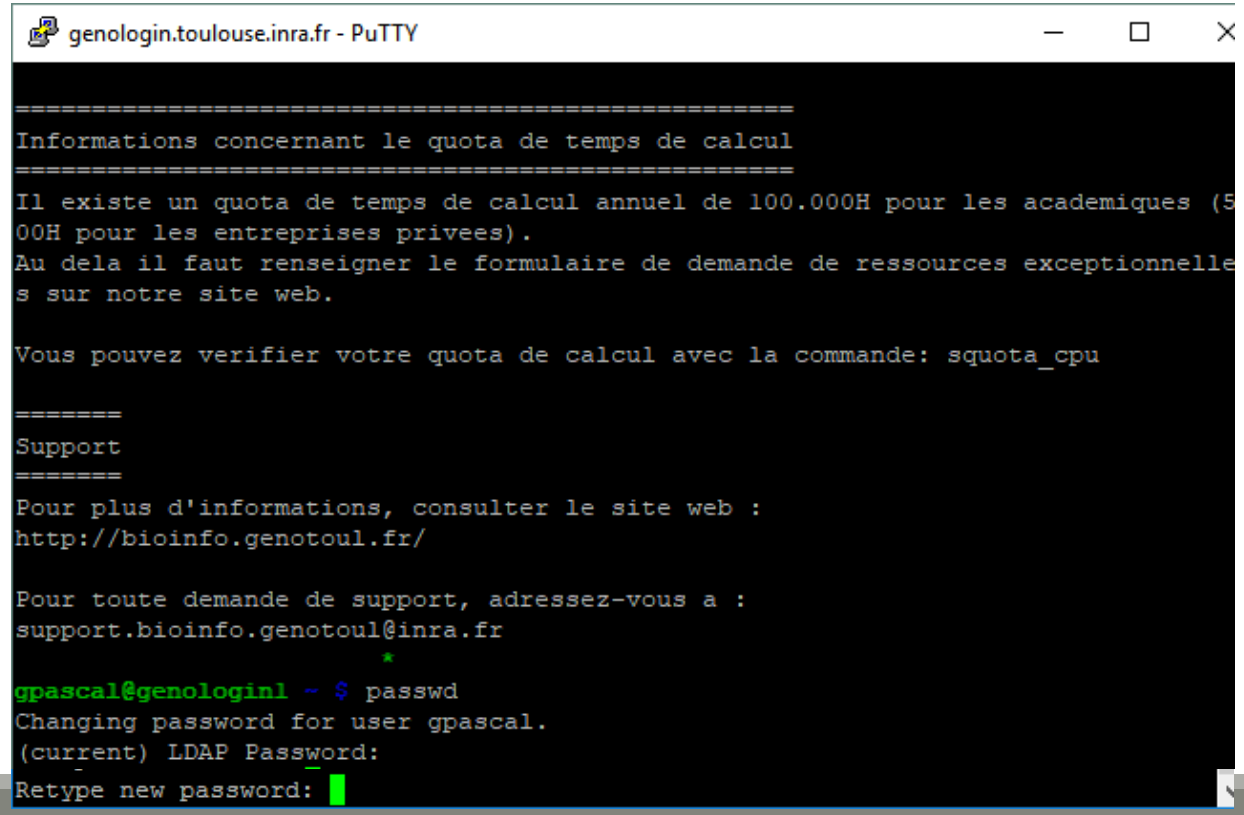
Change your password

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



Change your password

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



MAIN MENU

Galaxy
Analyze Data
Workflow
Shared Data
Visualization
Help
User
Using 0%
Tools
search tools
MANAGE YOUR DATA FILES
Get Data
Download Data
Jobs statistics
FILES MANIPULATION
Text Manipulation
Filter and Sort
Join, Subtract and Group
GFF
BED Tools
Convert Formats
Fetch Sequences
Statistics
SEQUENCES MANIPULATION
FASTA manipulation
FastQC: fastq/sam/bam
Illumina fastq
Generic FASTQ manipulation
FASTX-Toolkit for FASTQ data
SAM/BAM MANIPULATION : PICARD (BETA)
Conversion
QC/Metrics for sam/bam
BAM/SAM Cleaning
SAM/BAM manipulation: SAMtools
Sequences Queries
VCF Tools
SGS MAPPING
BWA - Bowtie
Indel Analysis
Variant calling
SNP annotation
RNAseq Alignement
RNAseq Raw Expression
RNAseq Cufflinks
AVAILABLE TOOLS
Genotoul Bioinfo
LA RÉGION OCCITANIE Pyrénées-Méditerranée
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/]
RESULTS VISUALISATION AND TOOL PARAMETER WINDOW
Ready to use Workflows
1/ FROGS - Find, Rapidly, Otus with Galaxy Solution
2/ Tax4Fun workflow
3/ Function Table for Tax4Fun matrix workflow
Outlook Web App
Nouveau message
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
History
search datasets
Data Collection
3 shown
33 b
3: seq1.fasta
2: seq2.fasta
1: seq3.fasta
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!

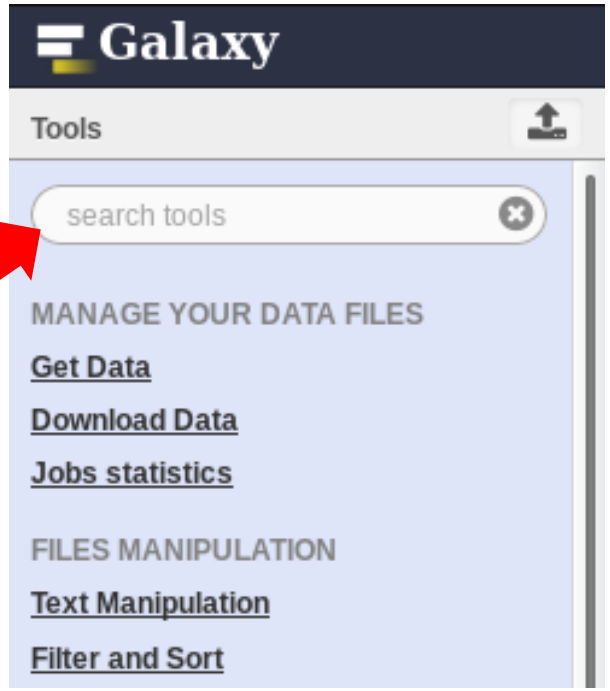
DISCOVER GALAXY

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.



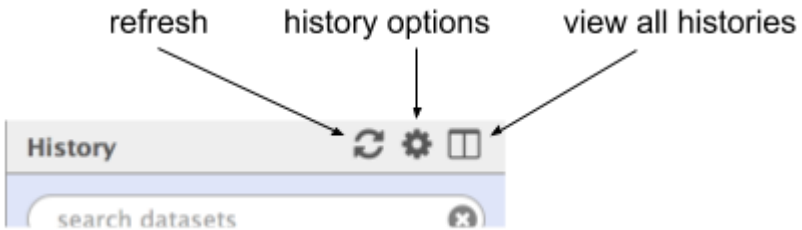
Manipulate Histories

Your Turn!

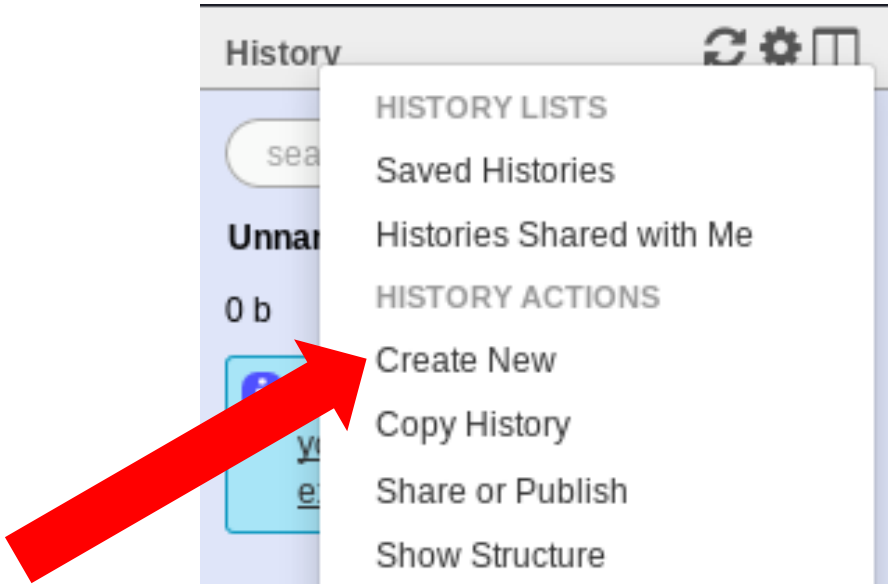
CREATE THE 4 HISTORIES NEEDED FOR THE FROGS FORMATION

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.

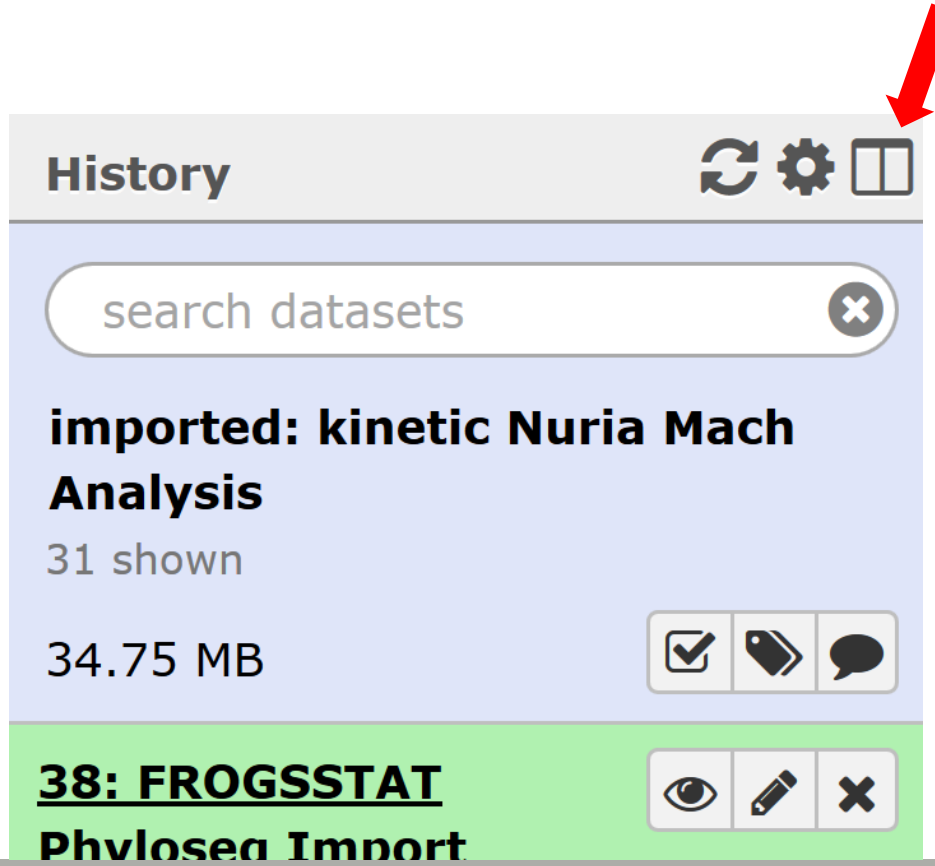
Don’t use special characters or accents!

Exercise

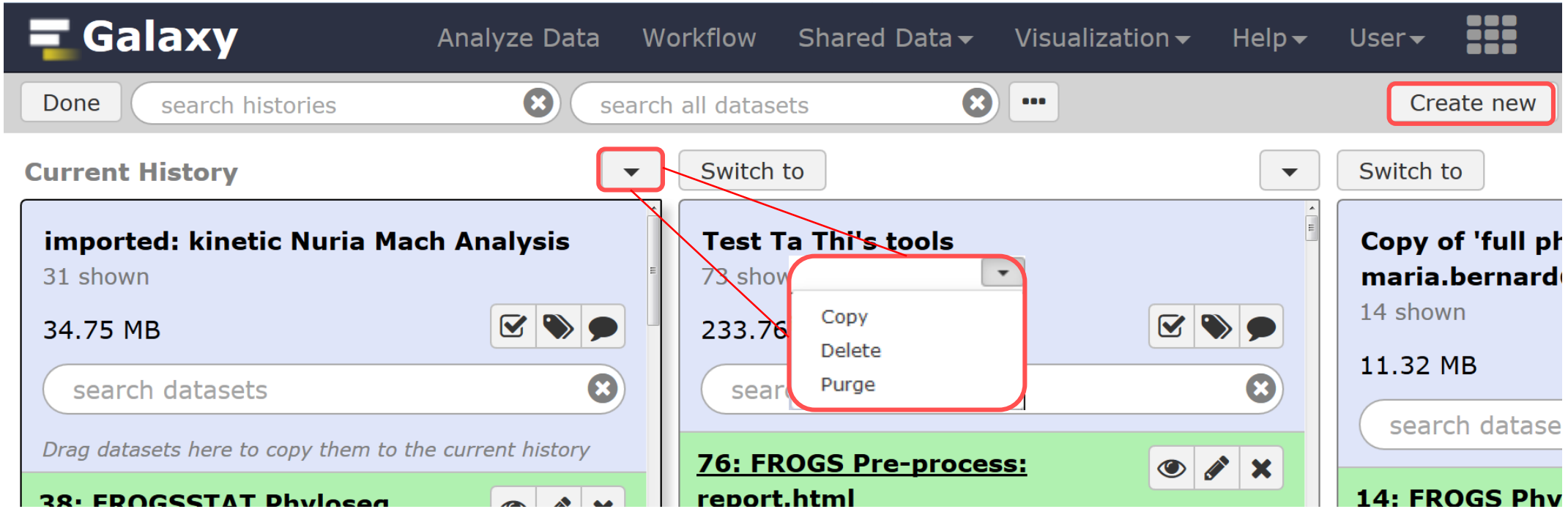
- Create histories named:
  - 454 or ITS
  - merged
  - temp
- Switch to the history named « multiplex » as current history.
- Go back to the main interface.

How to list all histories?

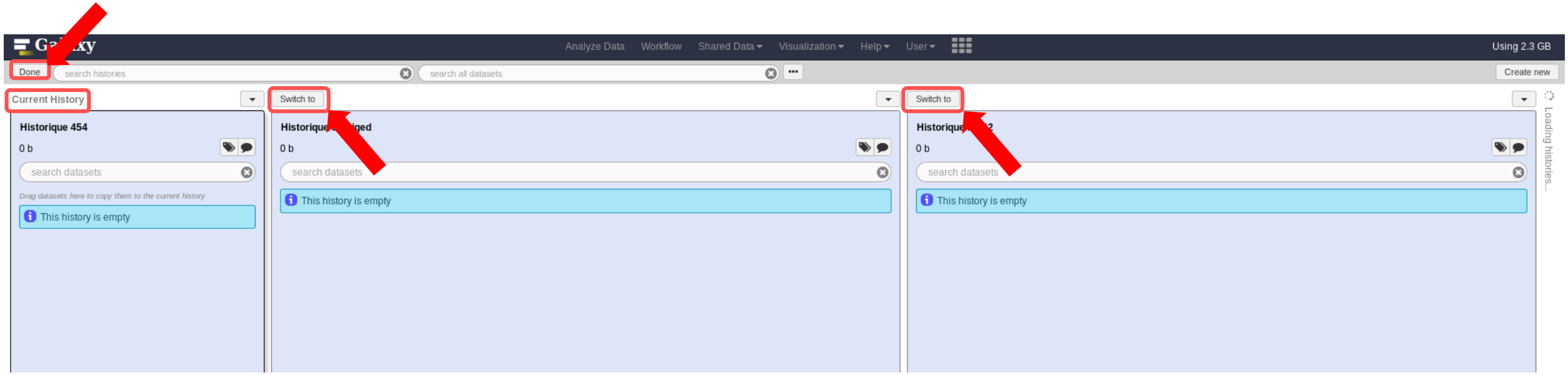
- To view all histories, click on this icon.



Explore the « View all histories » section



Switch current history



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

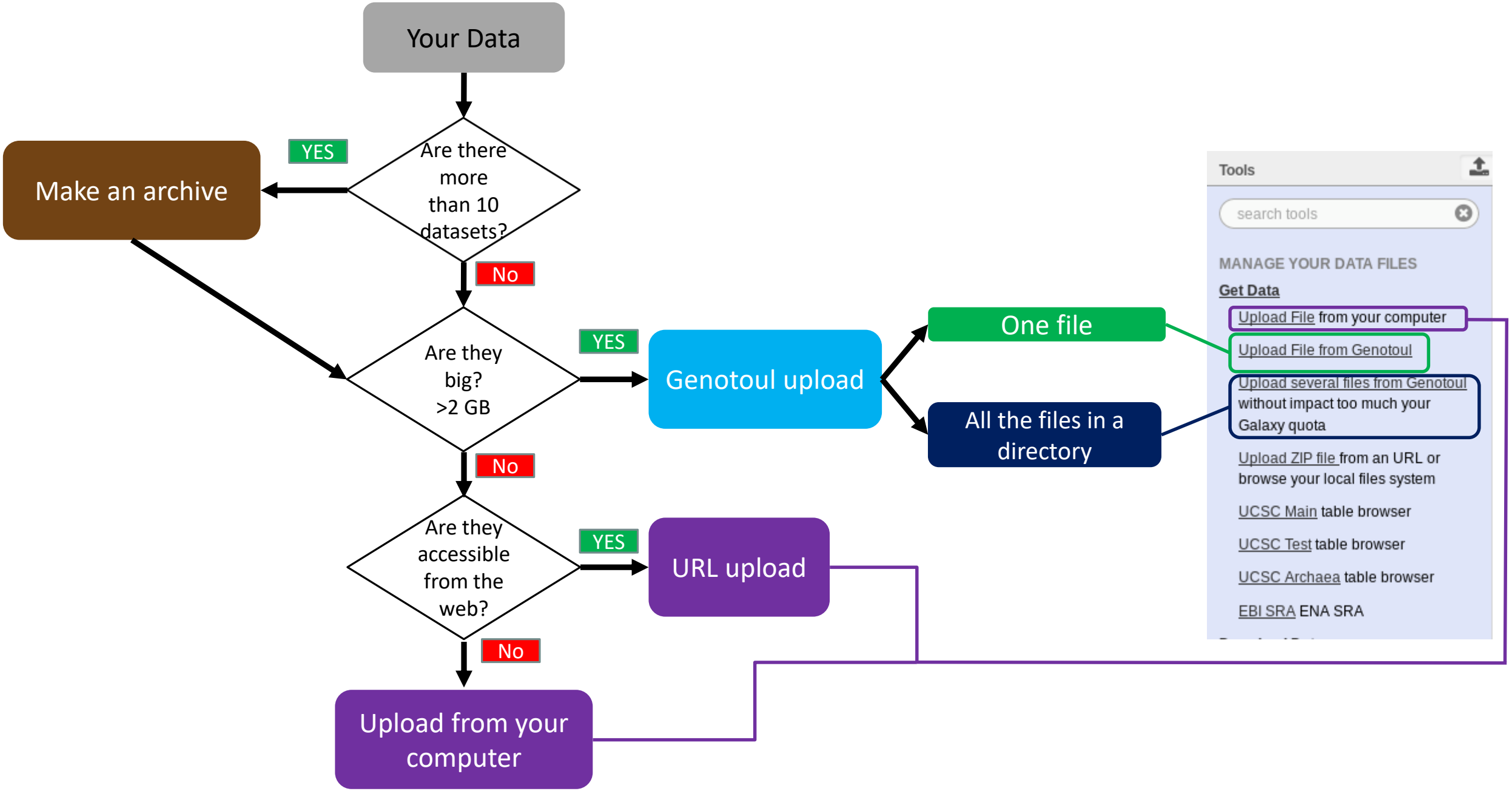
Data import

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?

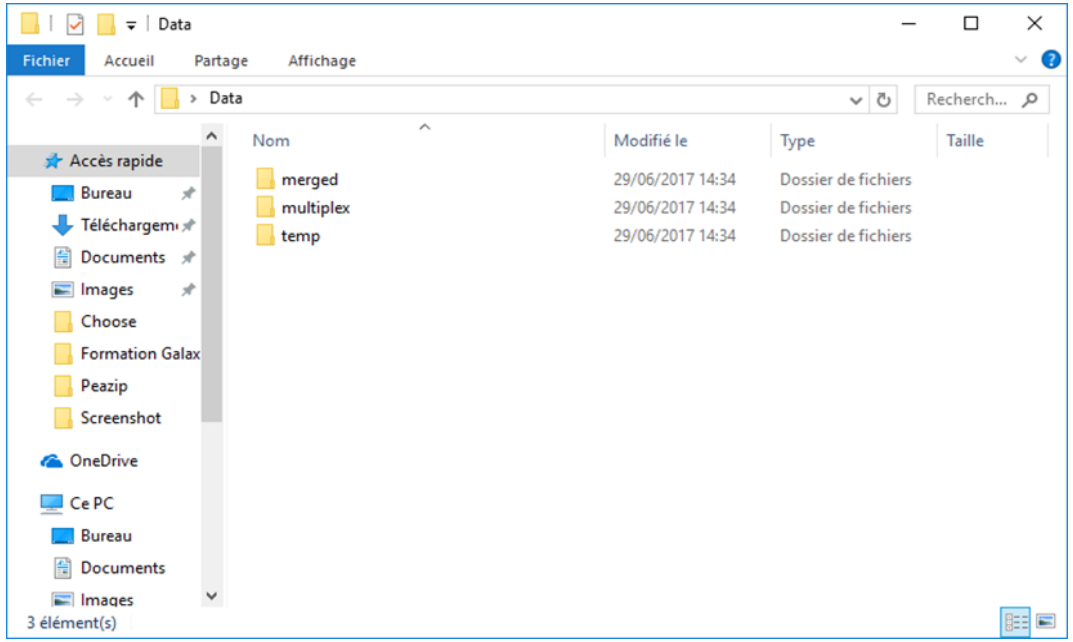


Your Turn!

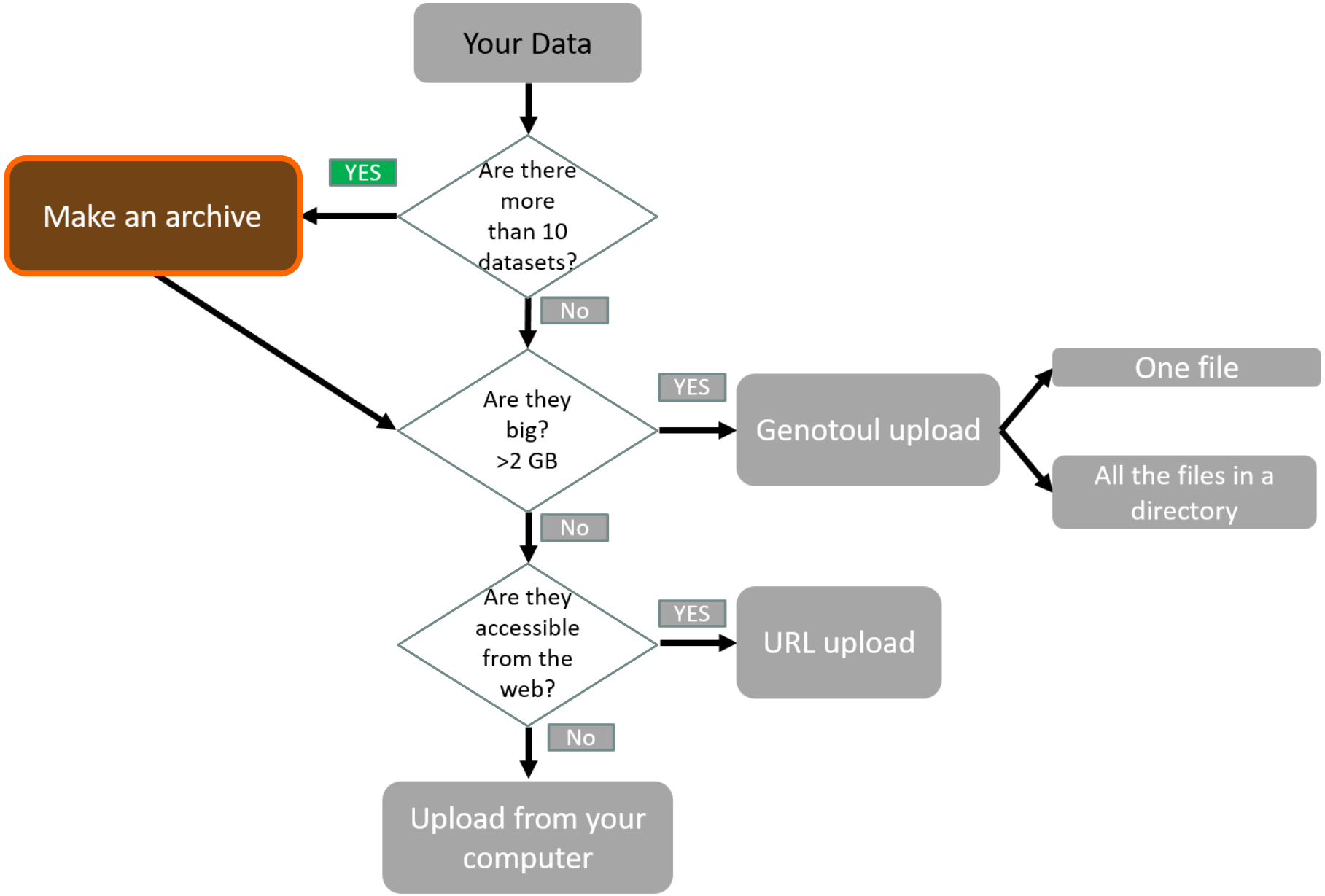
PREPARE FILES

Formation data

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

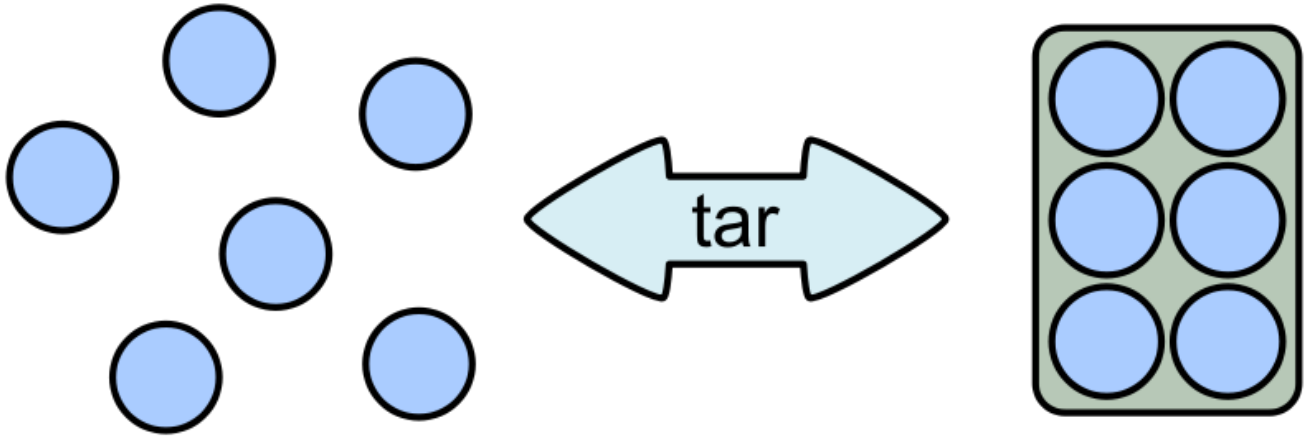


- If you don’t have this directory on your desktop, ask us.
- You can find all the formation data online on: http://genoweb.toulouse.inra.fr/~formation/15_FROGS/FROGS_ini/DATA/Dataset/


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

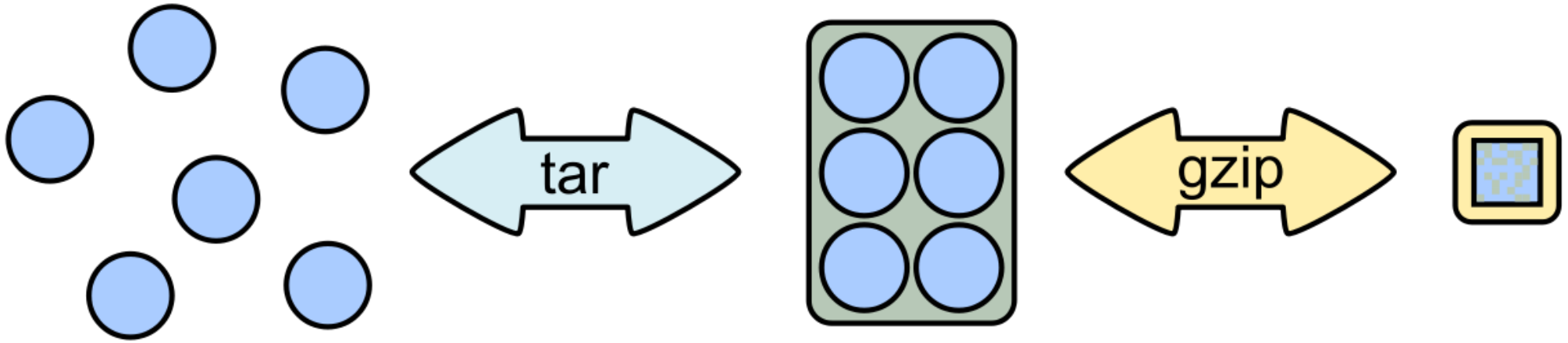
Create a Tar file

- 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!

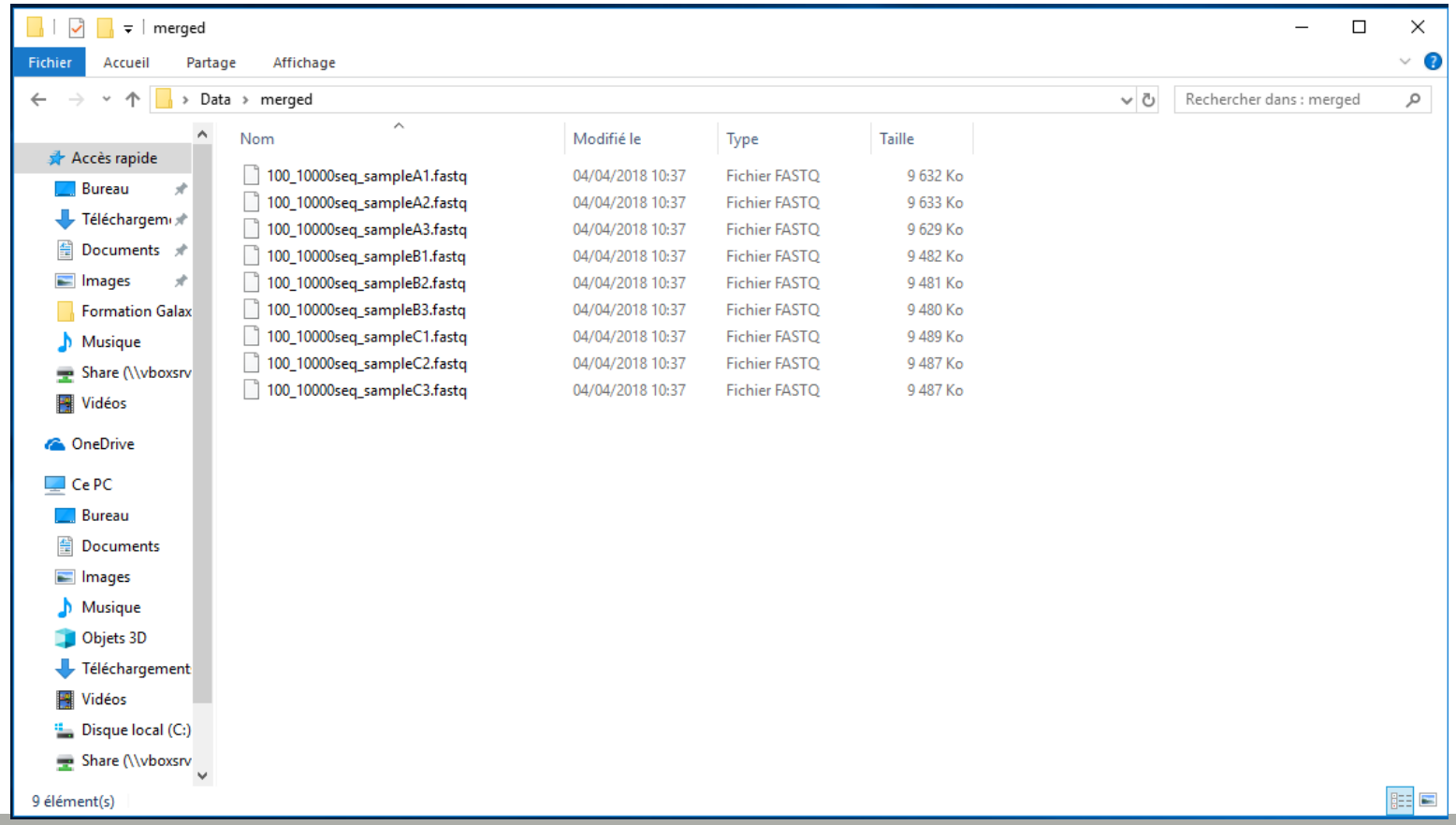
CREATE AN ARCHIVE WITH THE FILES IN MERGED

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/

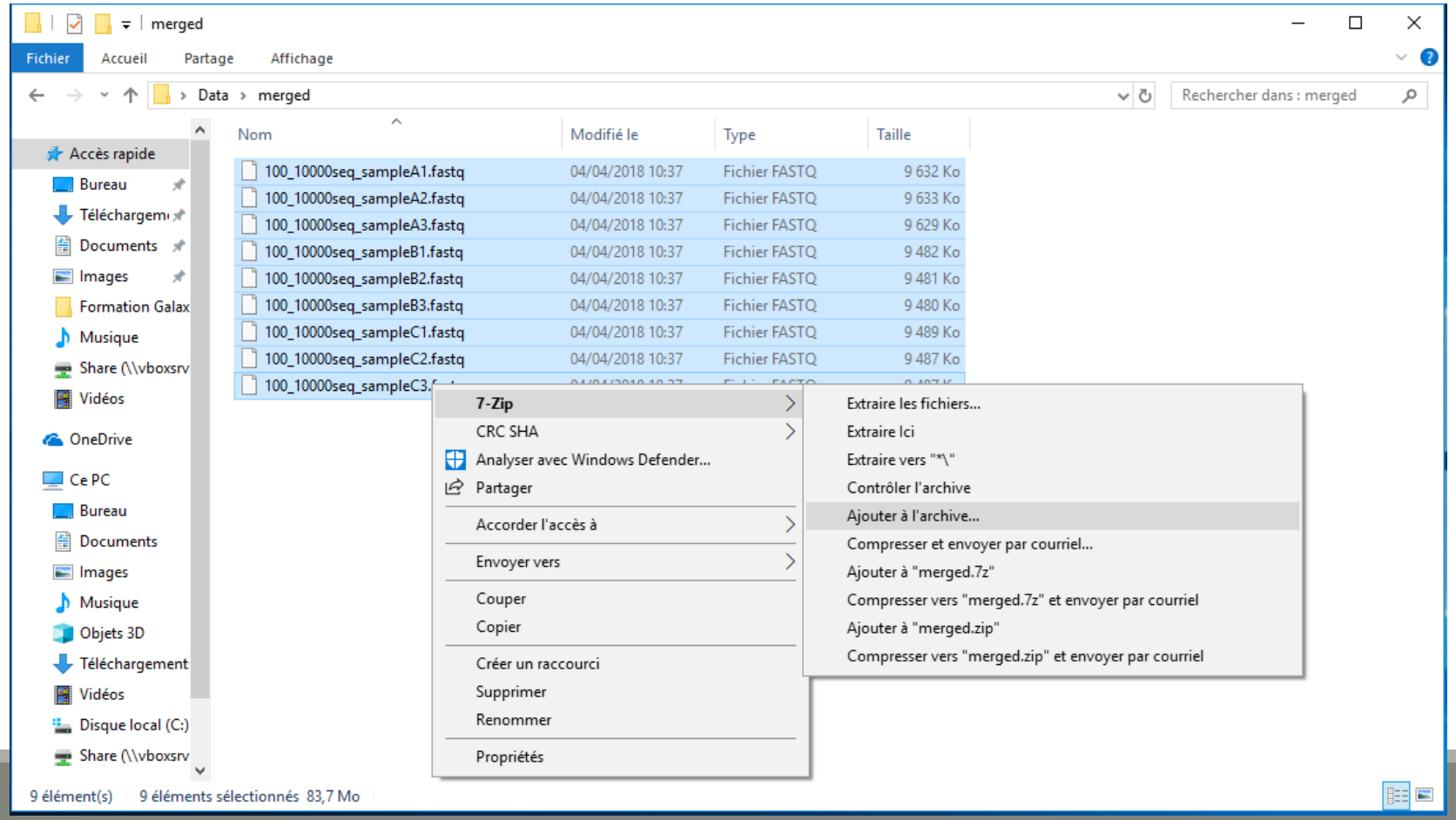
Create a Tar file with 7-Zip

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



Create a Tar file with 7-Zip

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

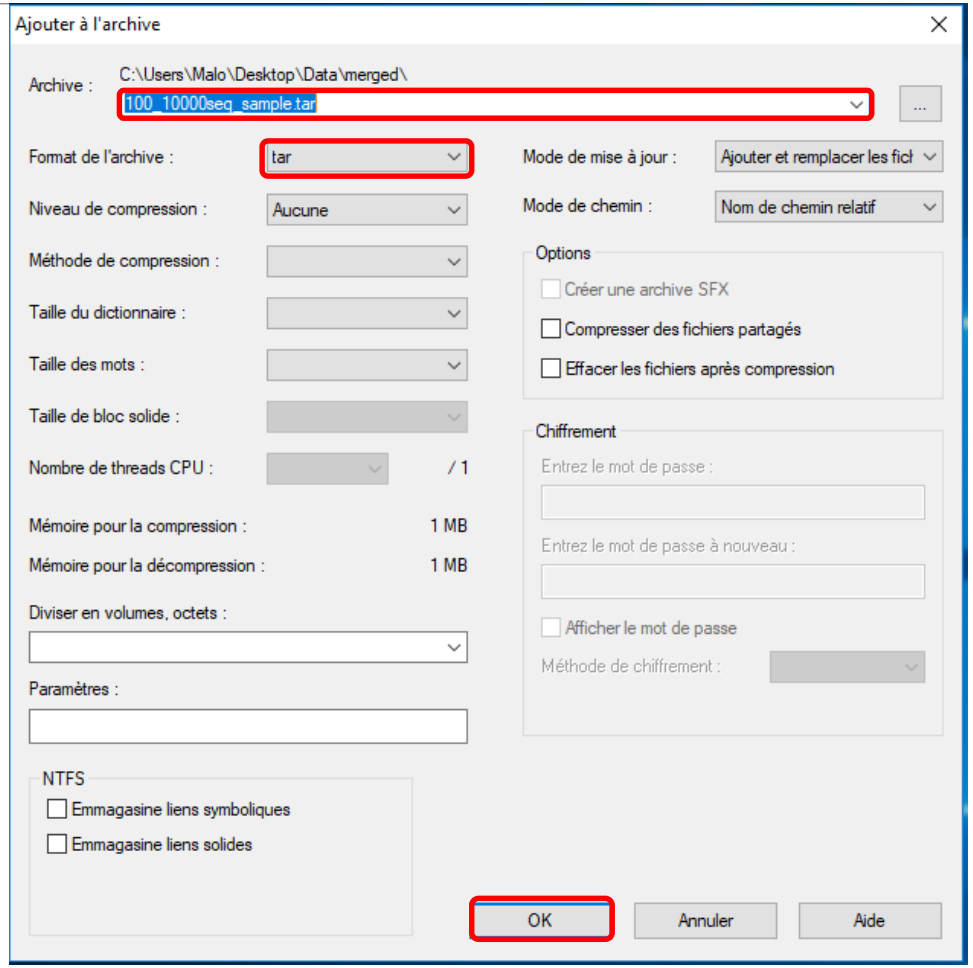


Create a Tar file with 7-Zip

- Change the name of the archive to:

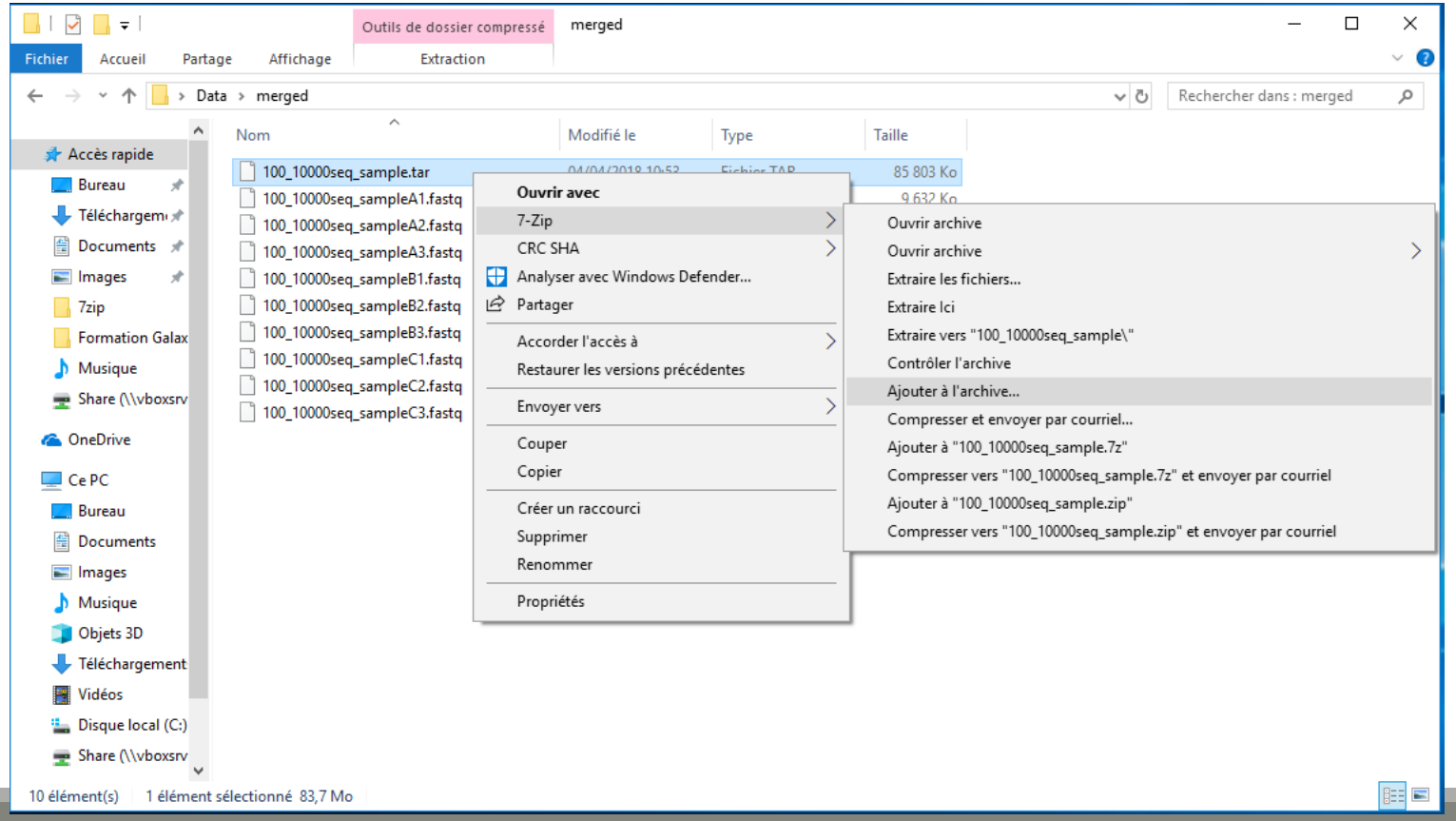
« 100_10000seq_sample.tar »

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



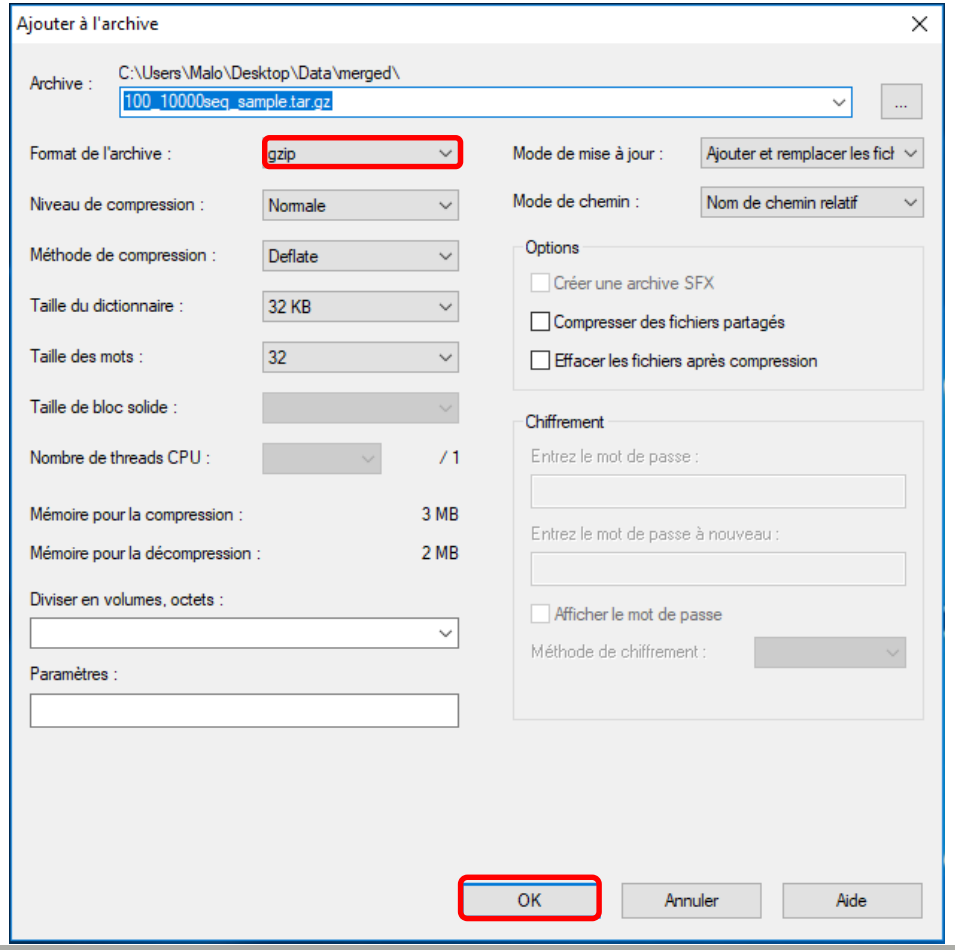
Create a Tar file with 7-Zip

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

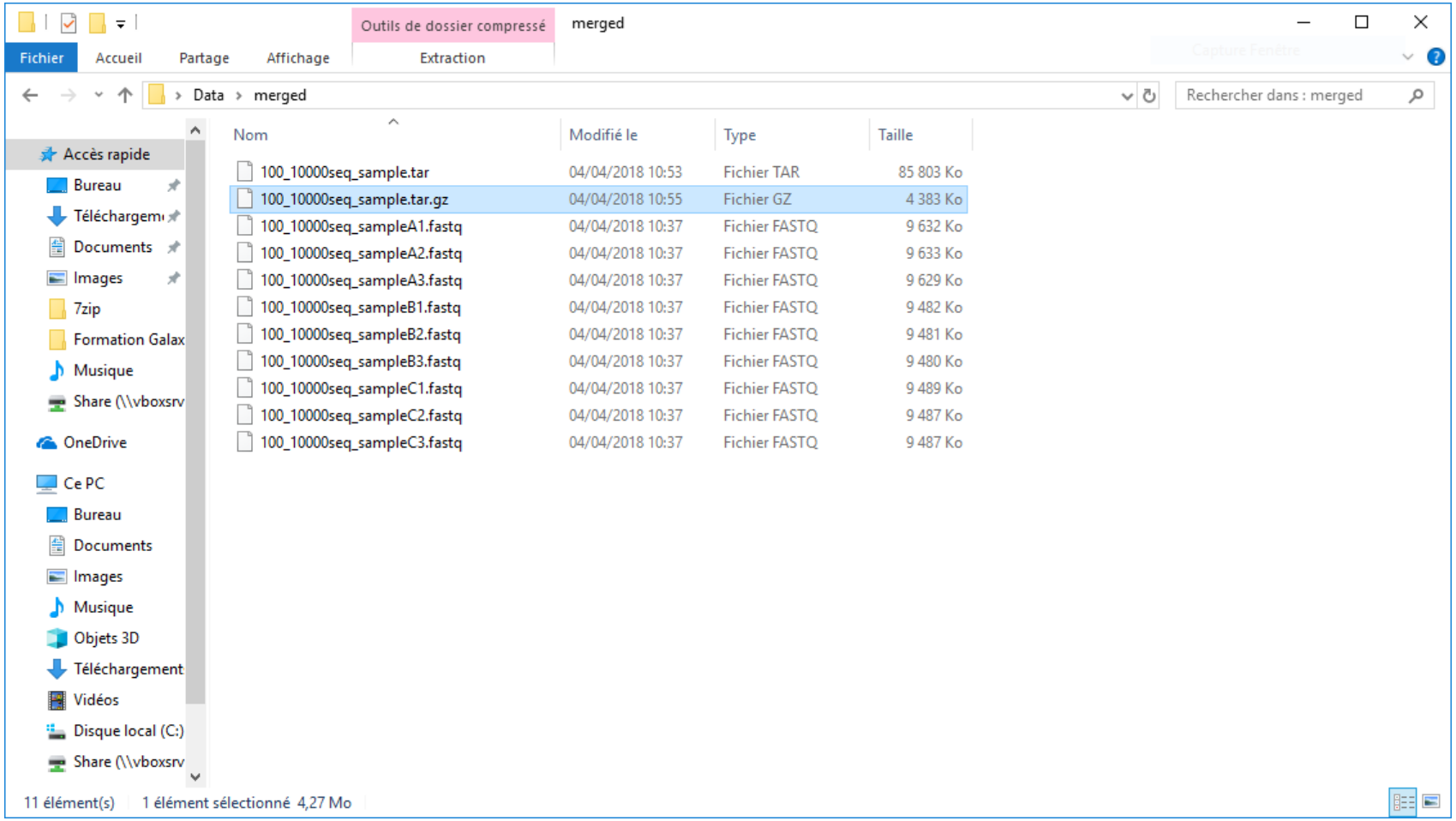


Create a Tar file with 7-Zip

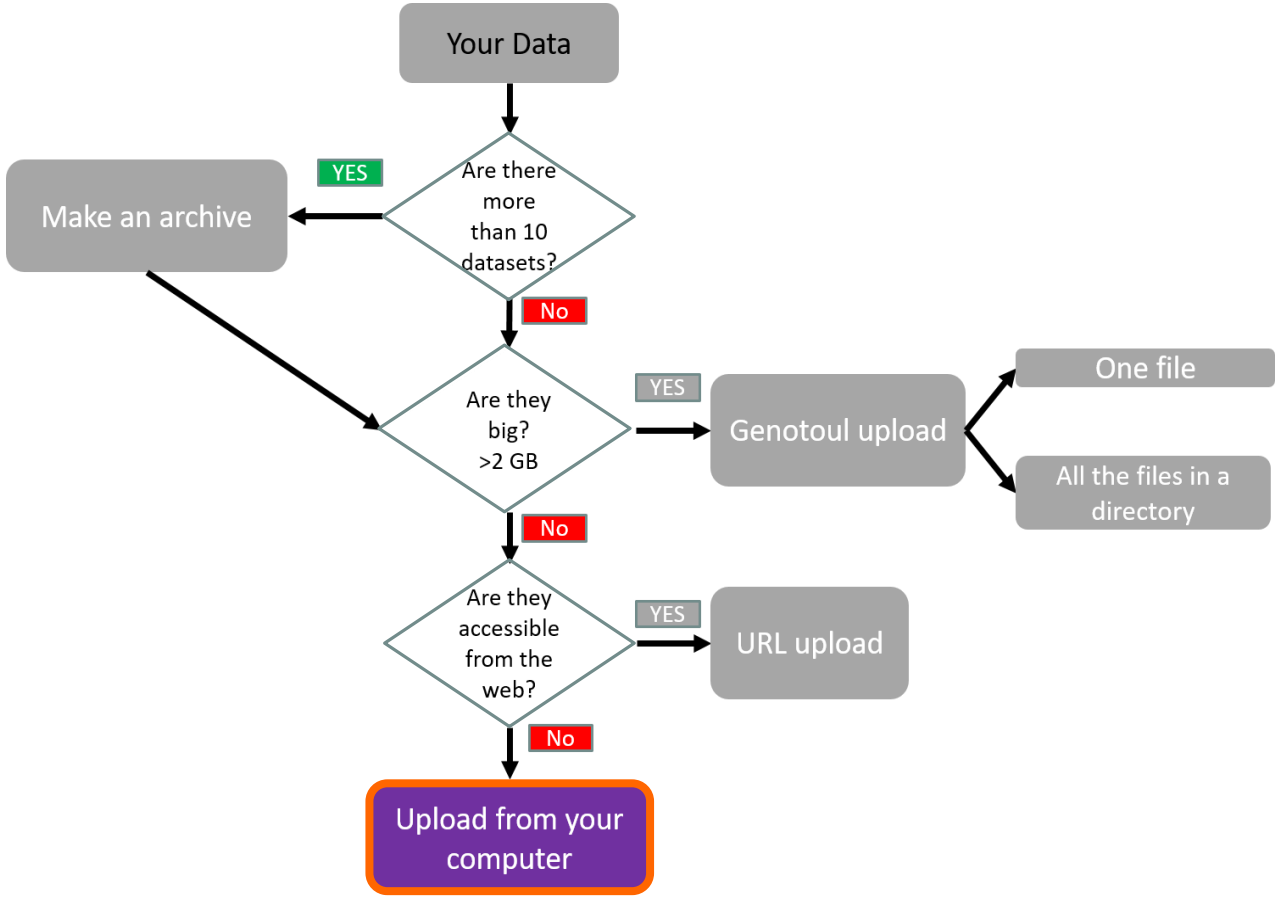
- Choose « gzip » as archive format .
- Click on OK .



Create a Tar file with 7-Zip



Your Turn!

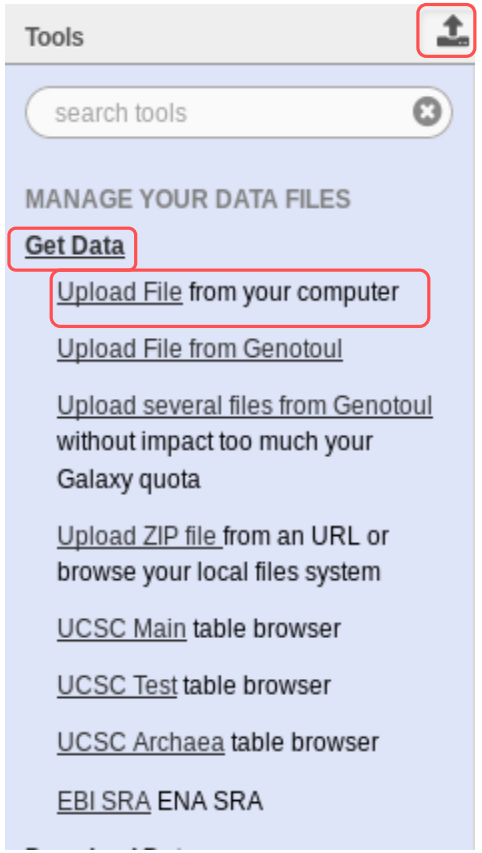


UPLOAD FILES FROM YOUR COMPUTER

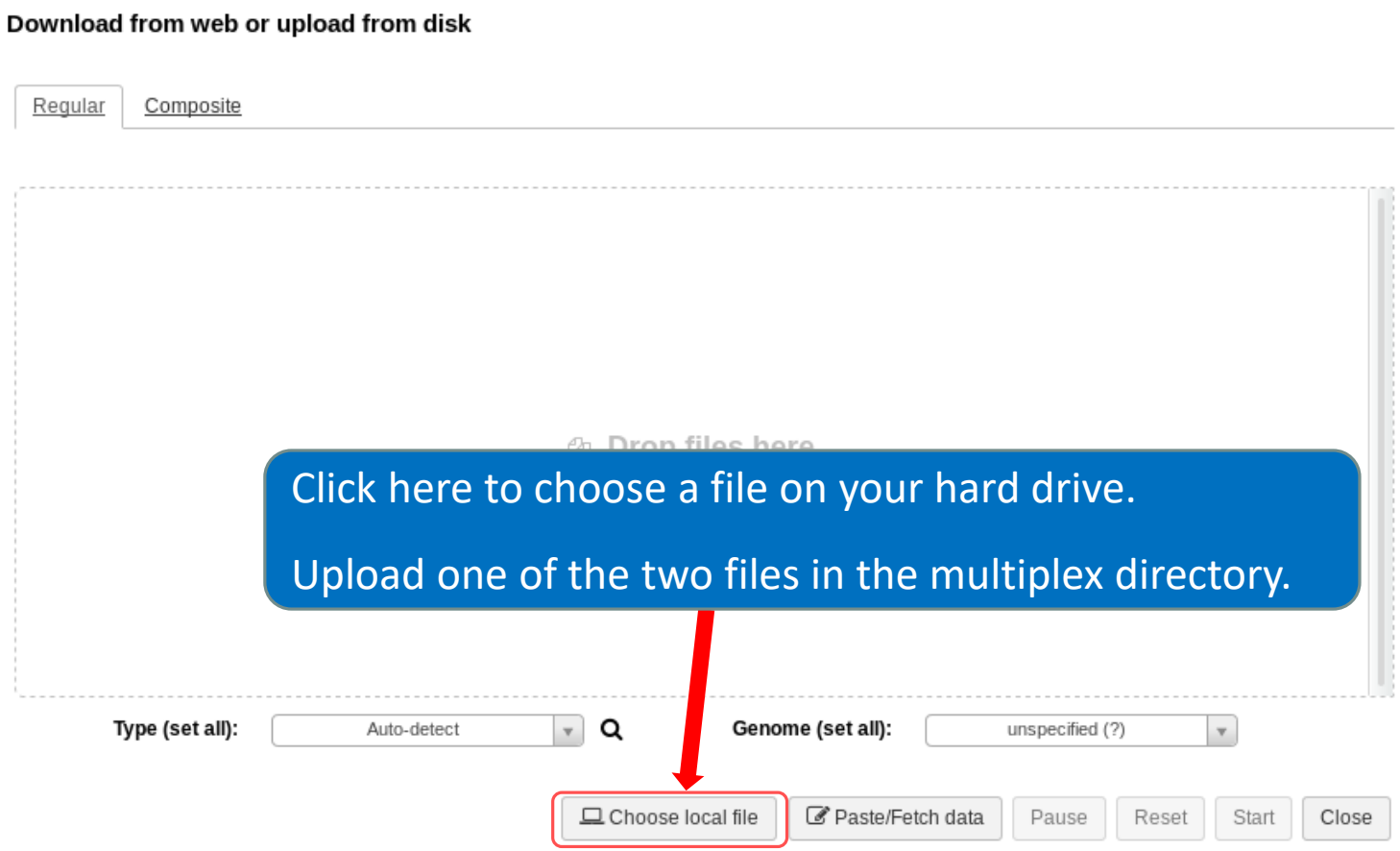
Upload local files

- 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



Upload local files



Upload local files

The datatype of a file indicates the content of it.

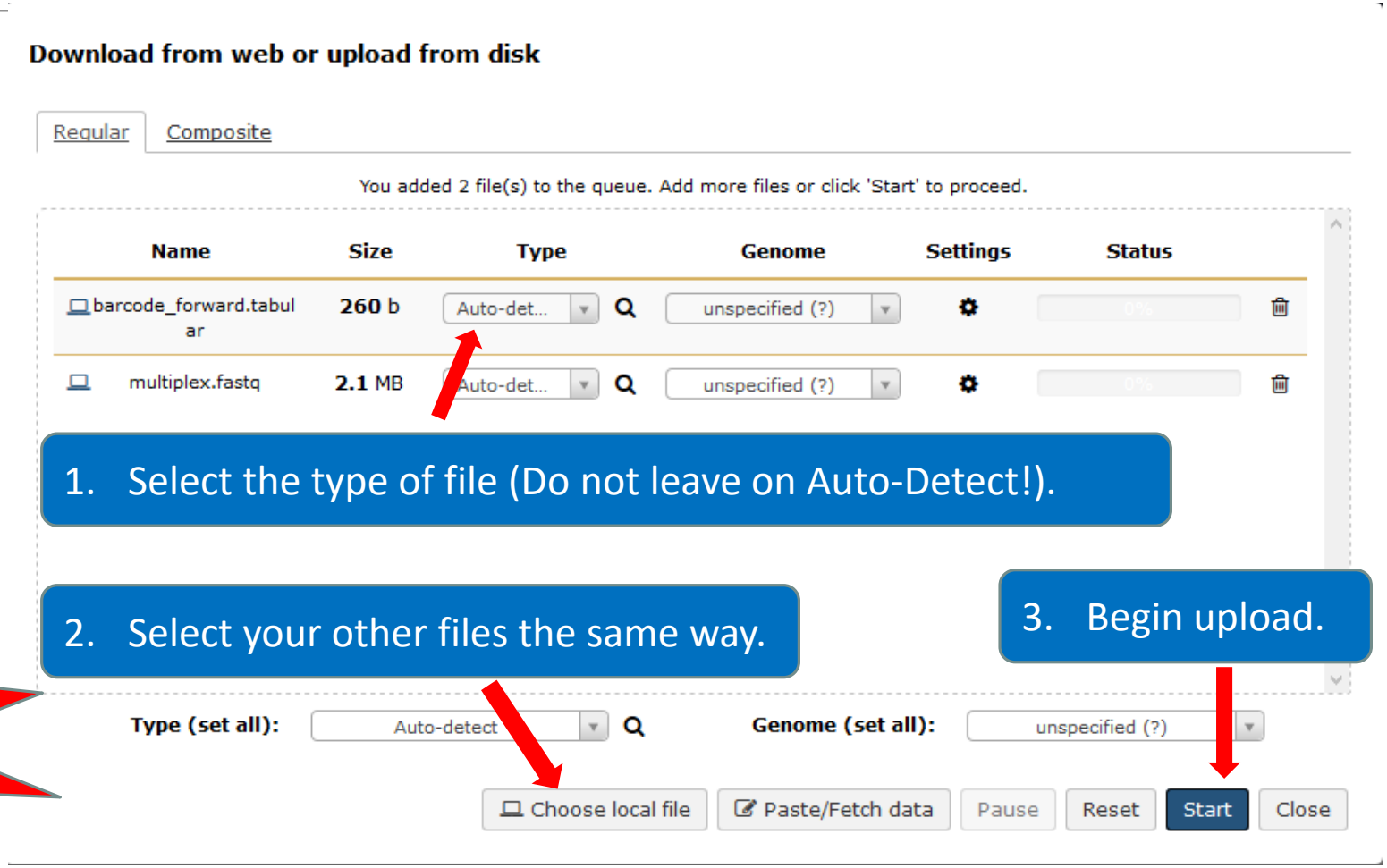
.fasta = fasta

.fastq = fastq

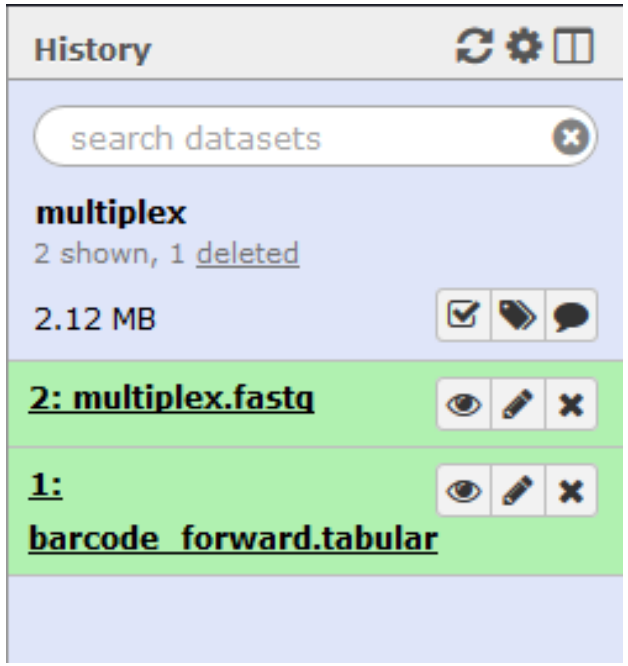
.tar = tar

Select the type corresponding to your file.

Do no not trust blindly the auto-detect!

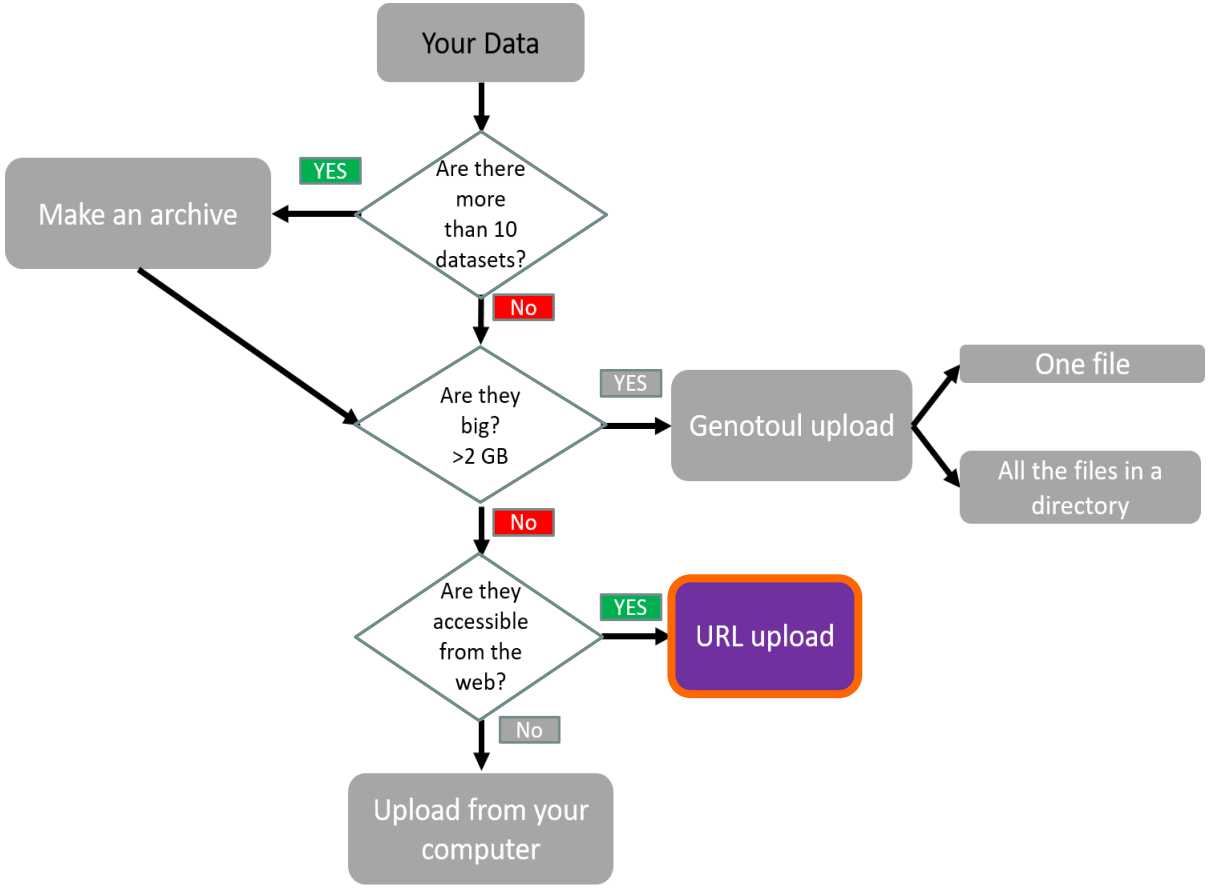


Upload local files



Your Turn!

UPLOAD FILE FROM AN URL



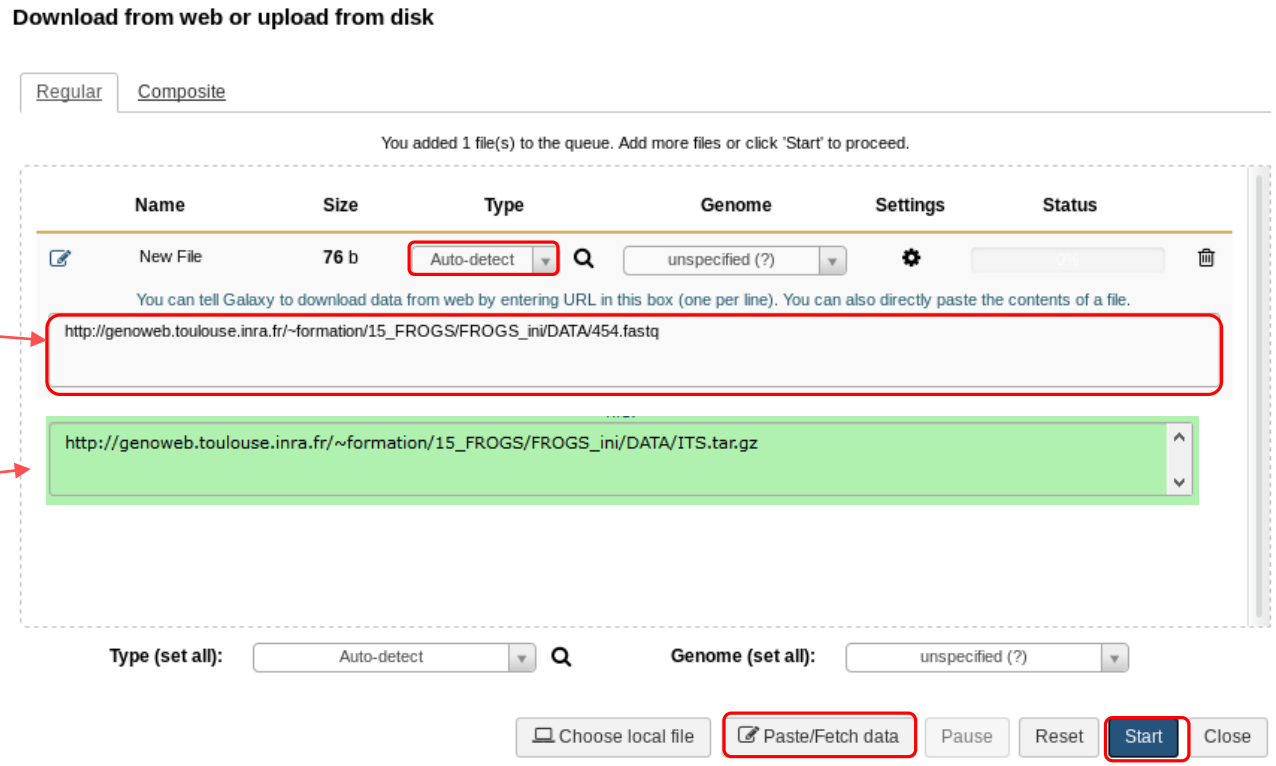
Upload file from URL

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

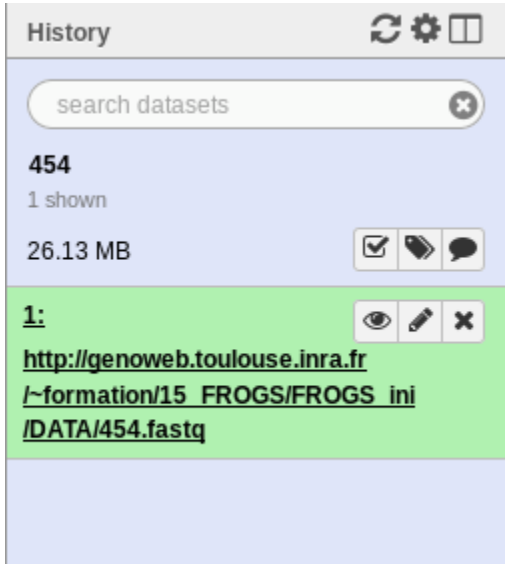
Or

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

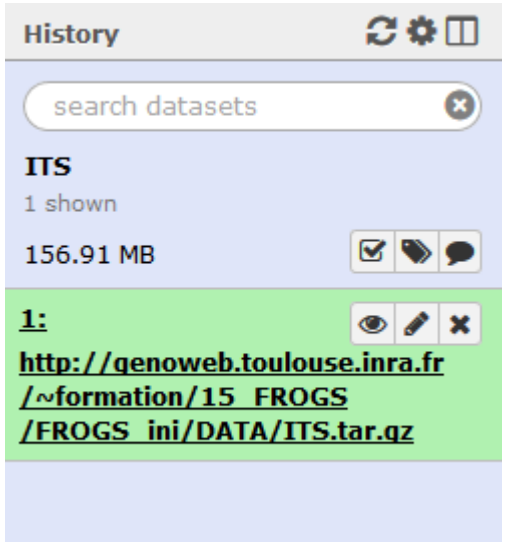
1. Change the type!
2. Click on Start.
3. You can put one address per line for multiple uploads.



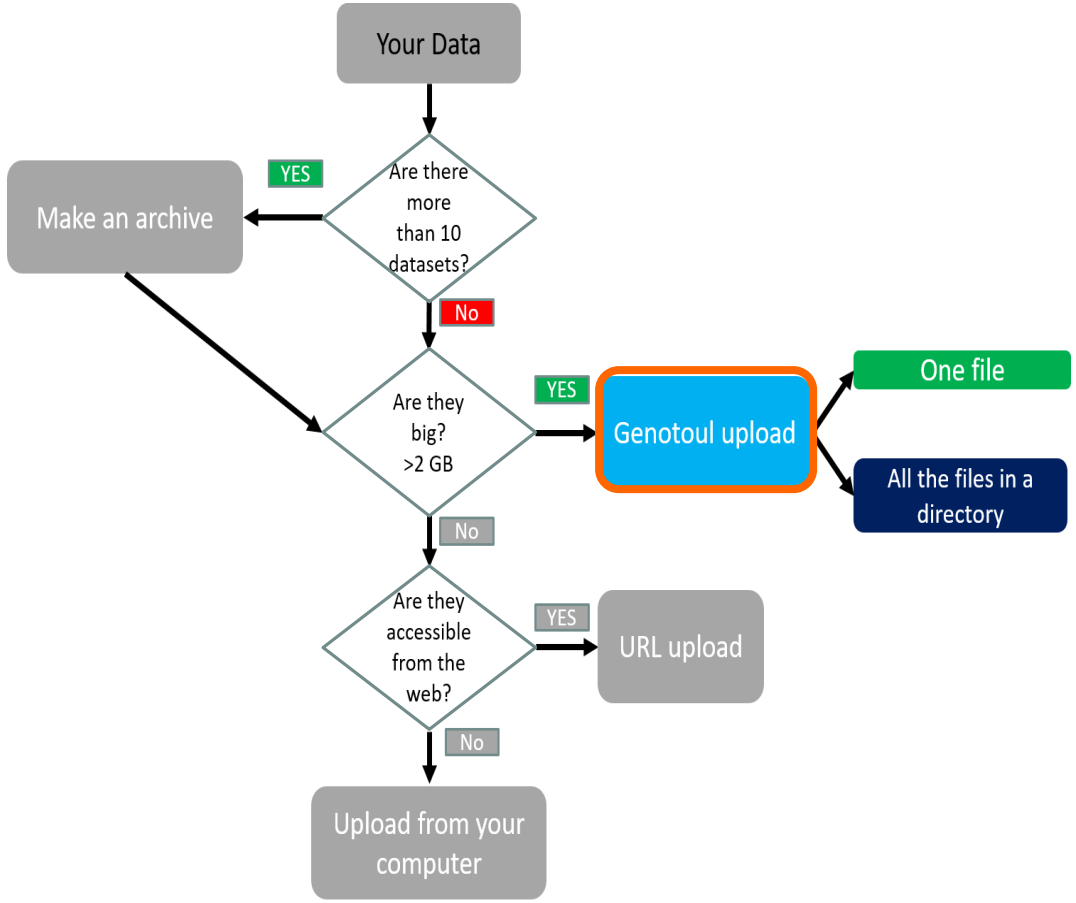
Upload file from URL



OR

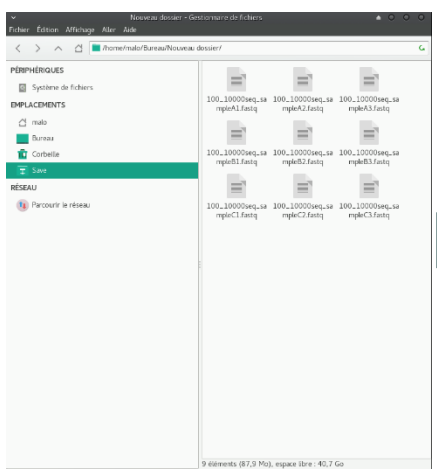


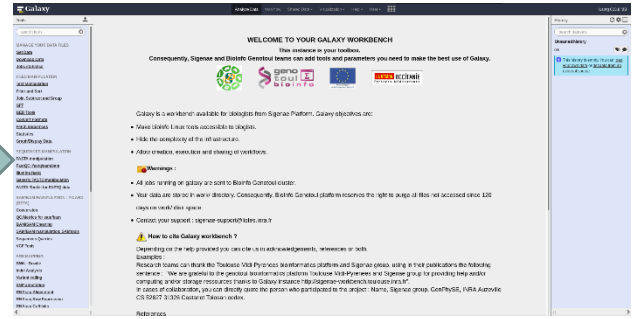
Your Turn!



UPLOAD FILES TO GENOTOUL AND LINK IT TO GALAXY

Objectives

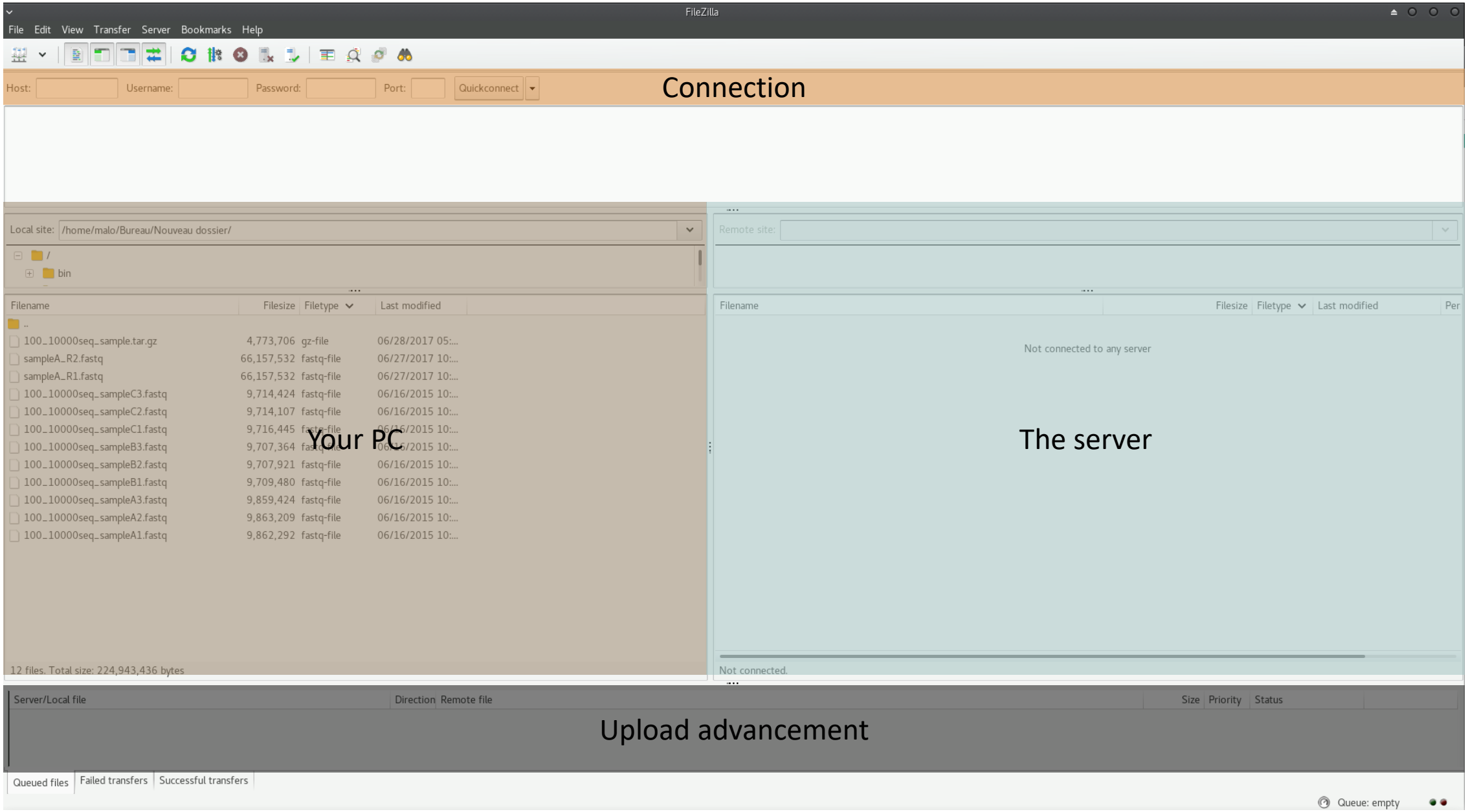


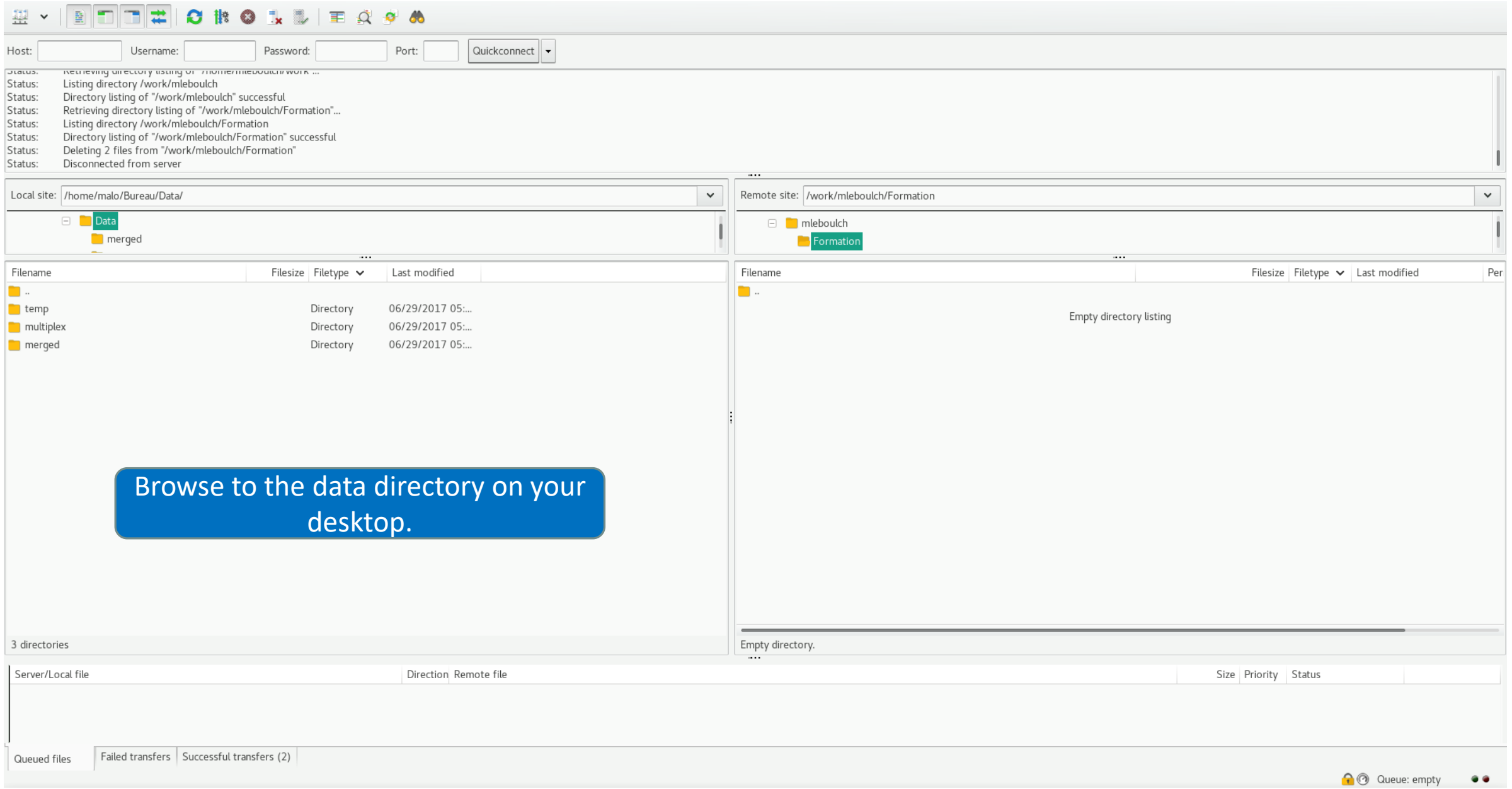


Transfer your files to your Genotoul account and link the file to Galaxy.

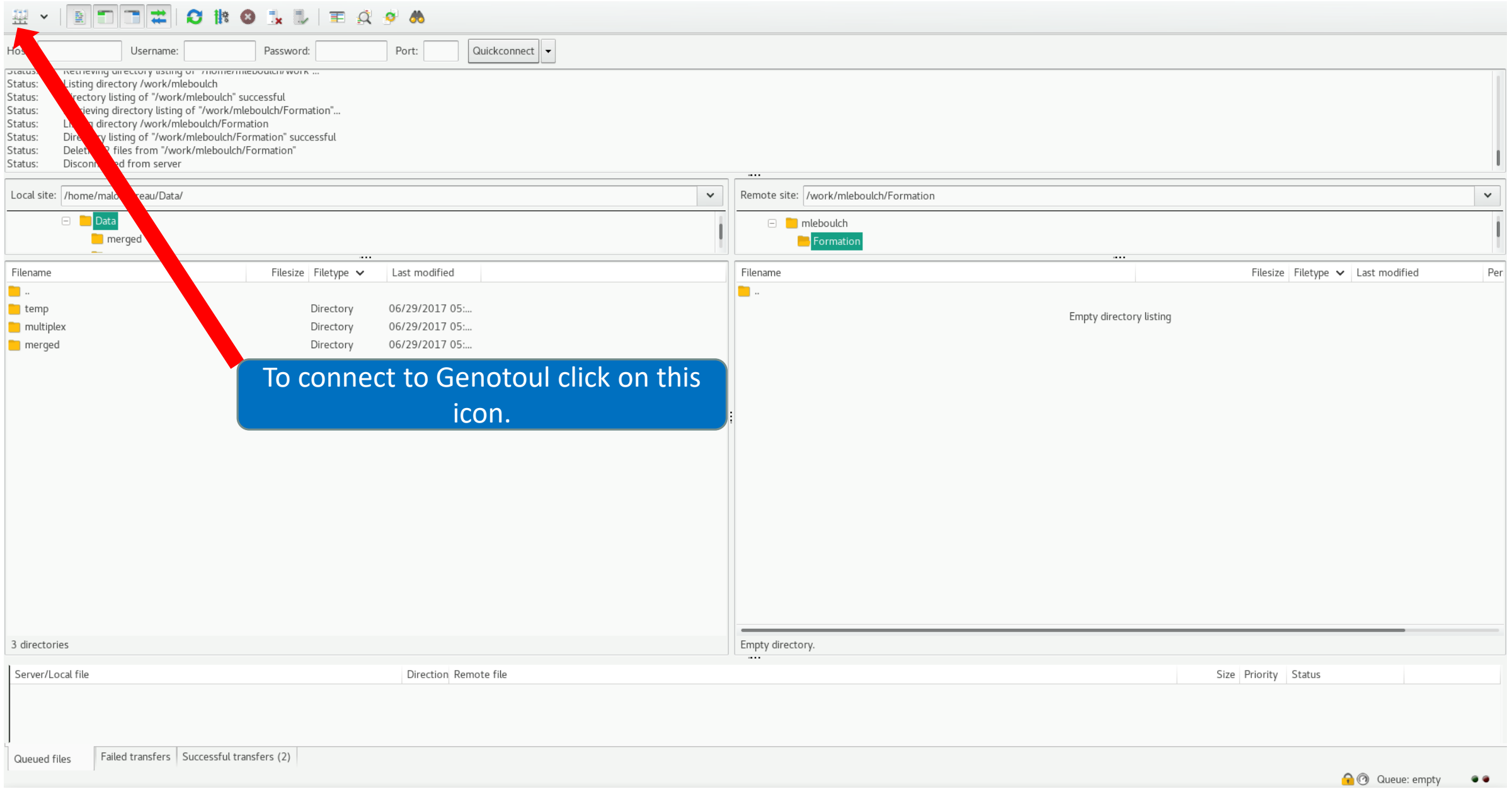
Preparation

- 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/


FileZilla
File Edit View Transfer Server Bookmarks Help
Host: Username: Password: Port: Quickconnect
Connection
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
Your PC
Remote site:
Filename Filesize Filetype Last modified Per
Not connected to any server
Not connected.
The server
Server/Local file Direction Remote file Size Priority Status
Upload advancement
Queued files Failed transfers Successful transfers
Queue: empty

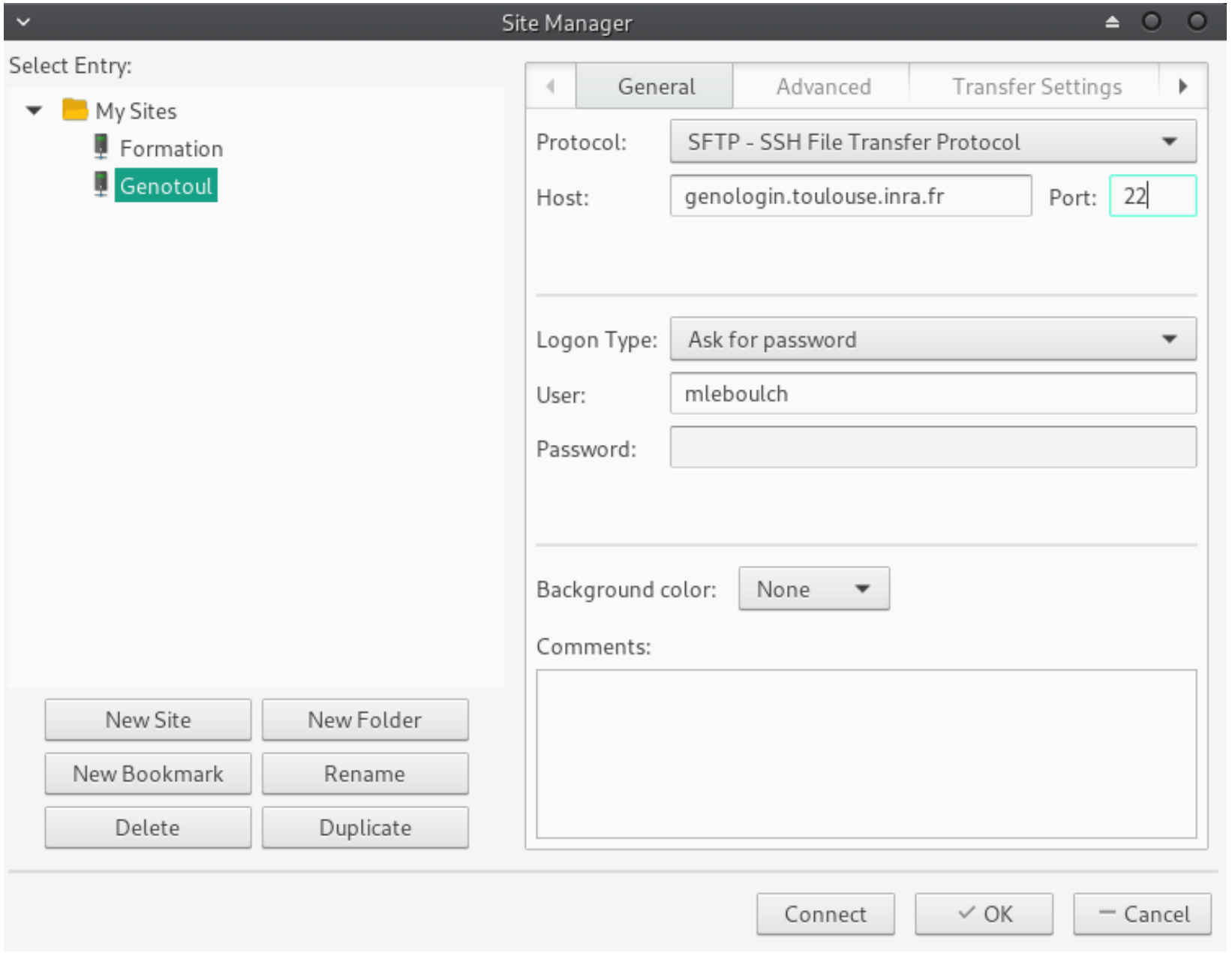


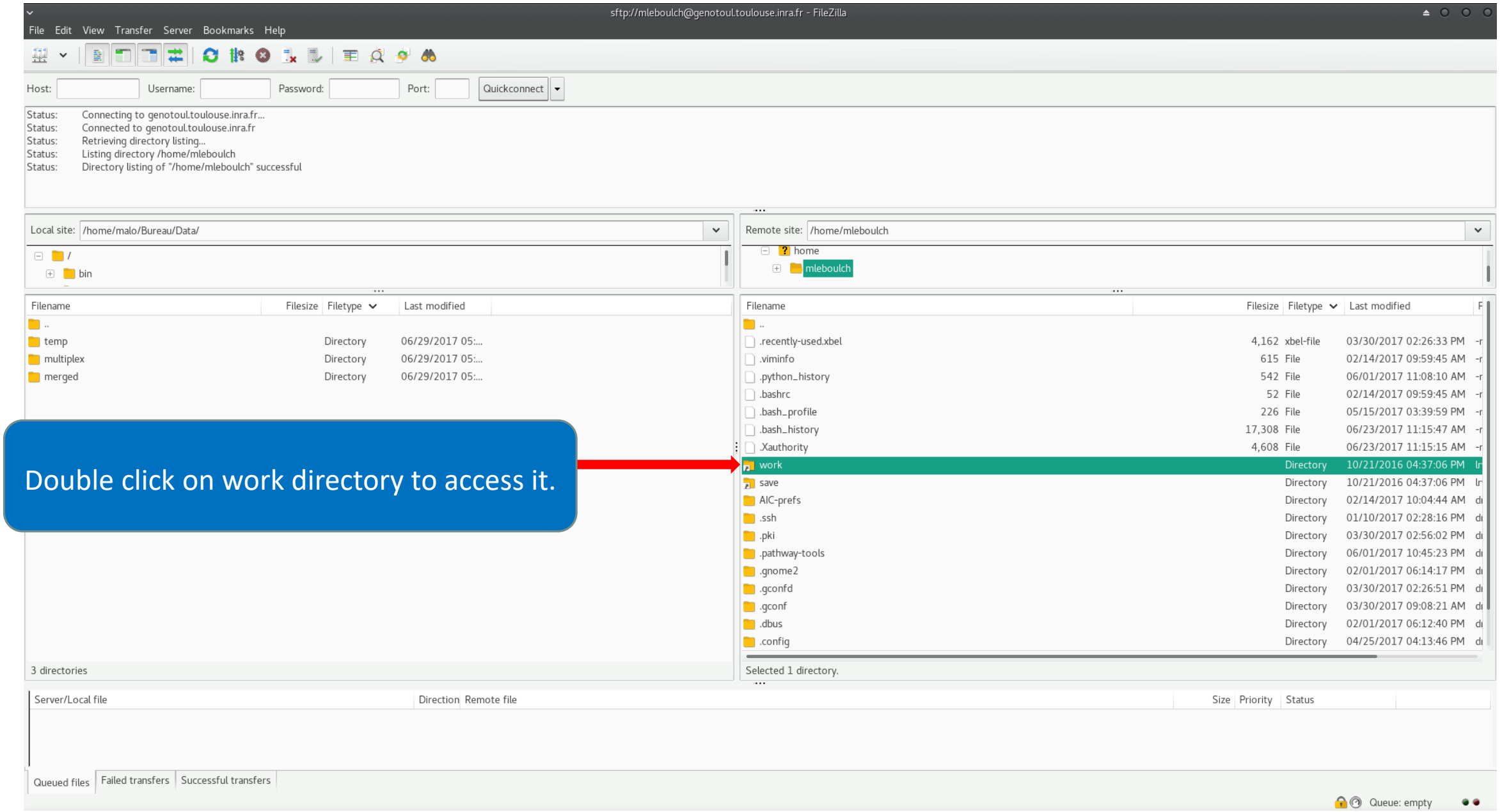
Browse to the data directory on your desktop.


Host:
Username:
Password:
Port:
Quickconnect
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...eau/Data/
Data
merged
Remote site: /work/mleboulch/Formation
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
Empty directory listing
Empty directory.
Server/Local file
Direction
Remote file
Size
Priority
Status
Queued files
Failed transfers
Successful transfers (2)
Queue: empty
To connect to Genotoul click on this icon.

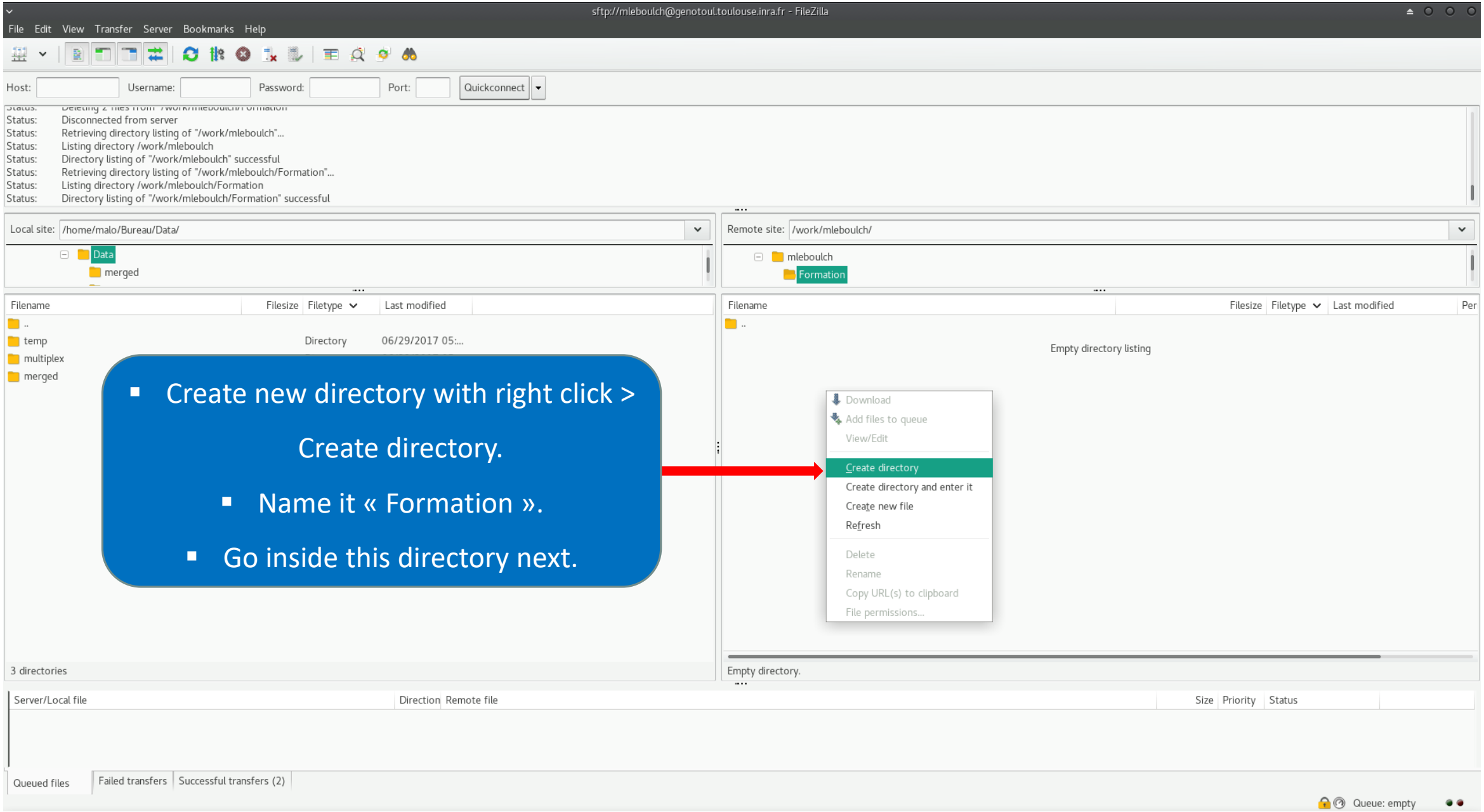
- Host: genologin.toulouse.inra.fr
- Port: 22
- Protocol: SFTP
- User: your Genotoul login
- Logon Type: Ask for password
- Click on connect.

- Password: your password

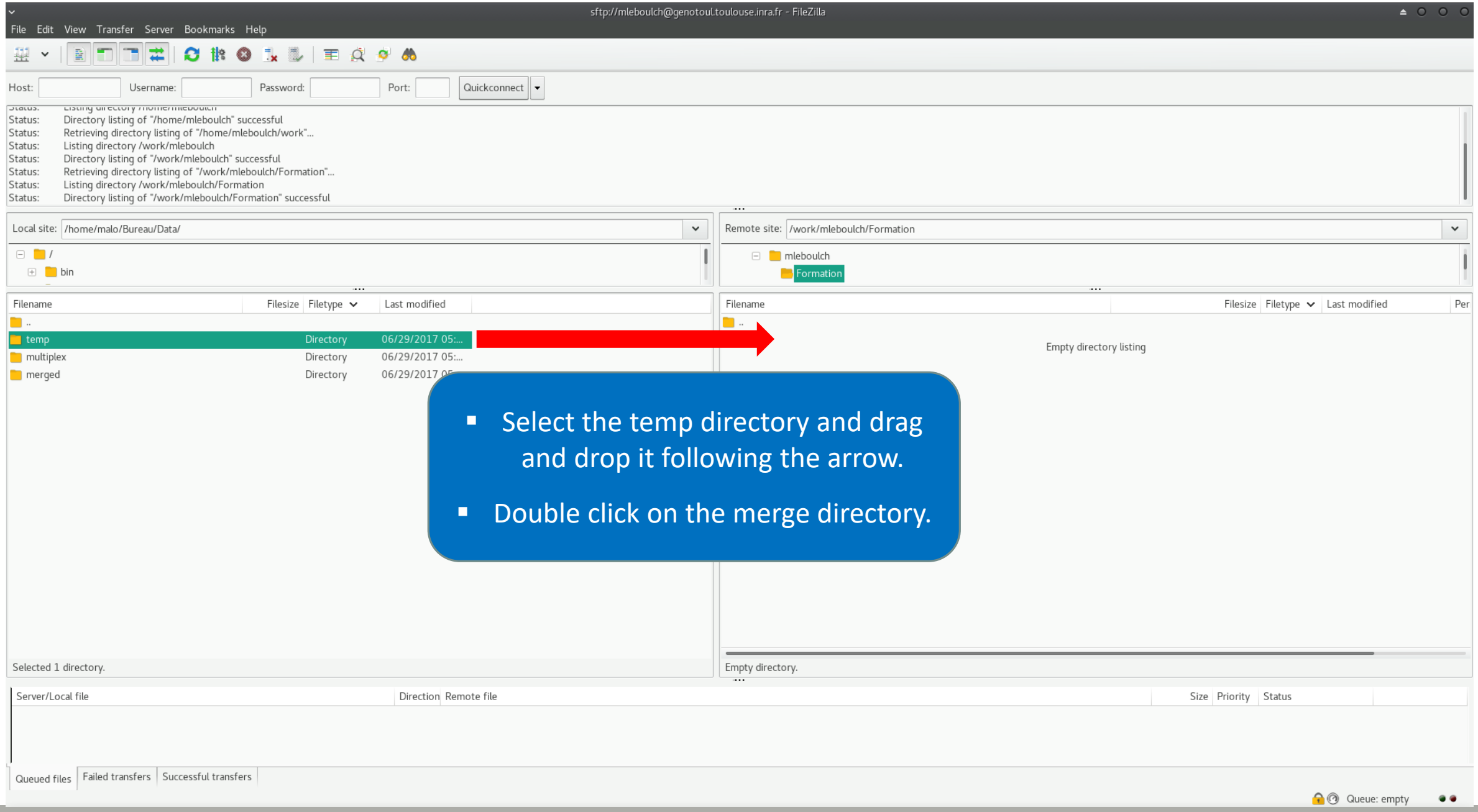


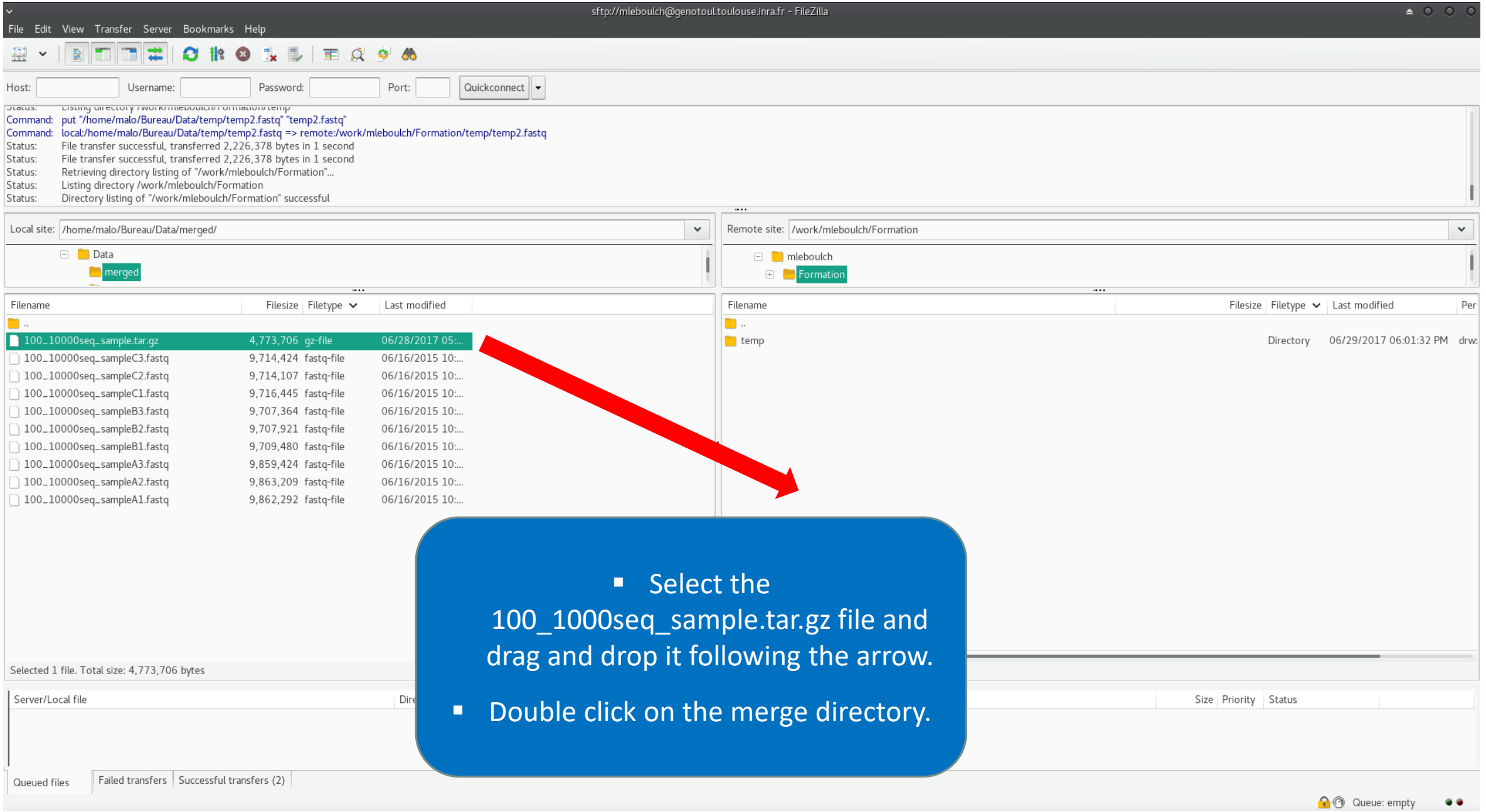


Double click on work directory to access it.

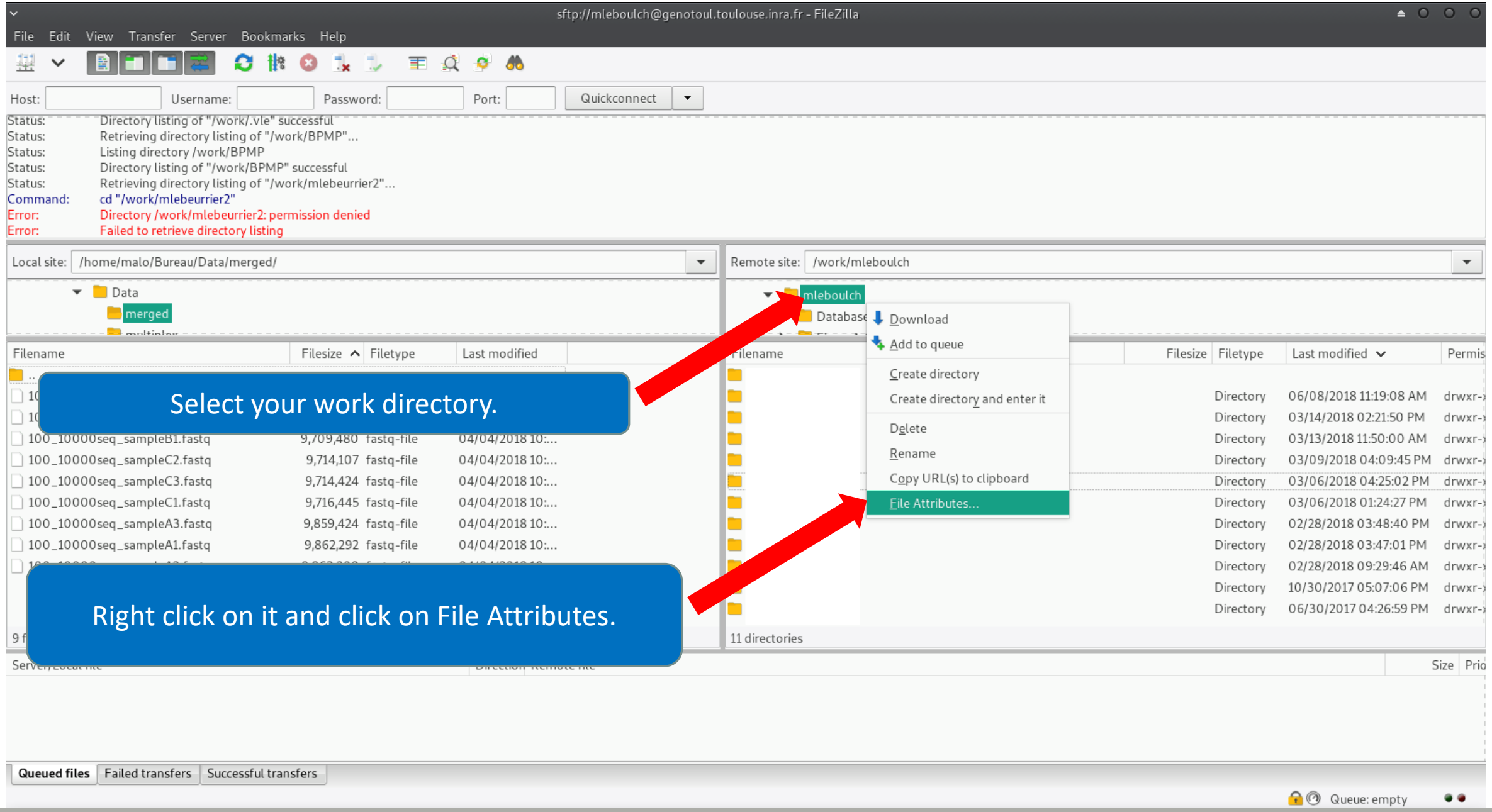


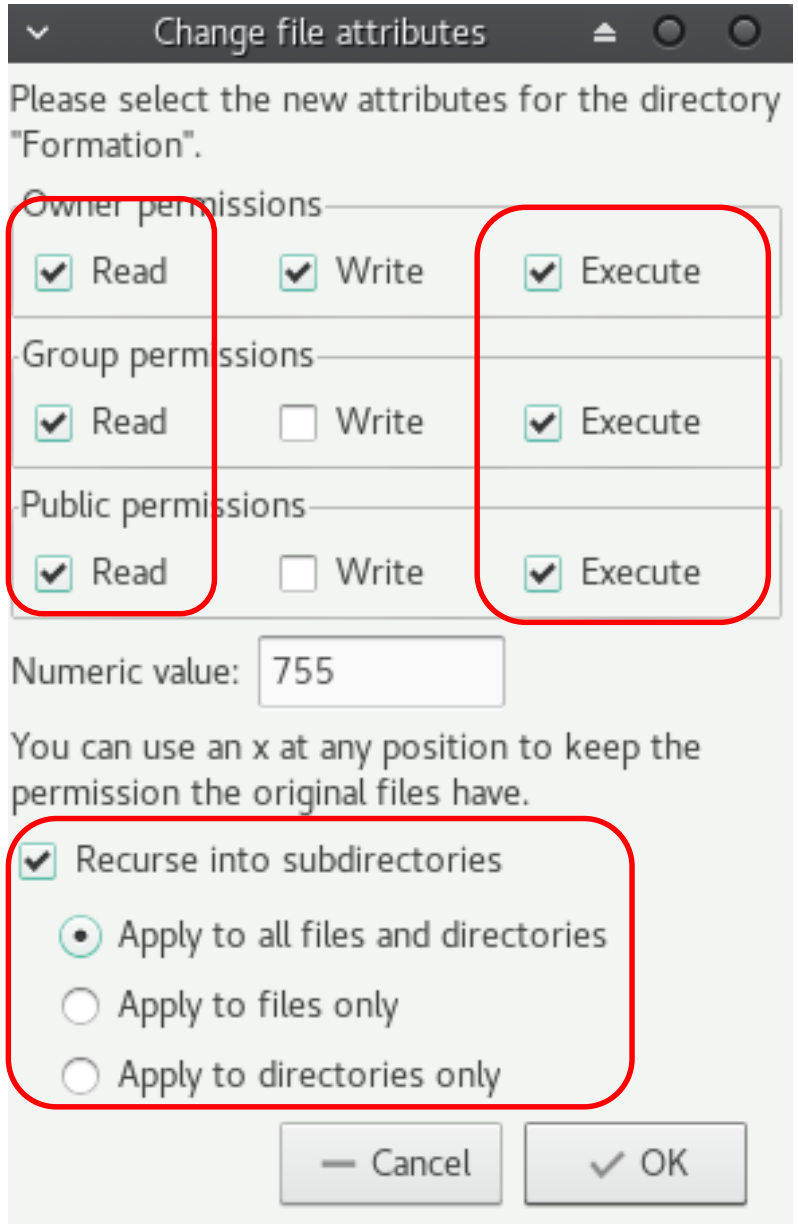
- Create new directory with right click > Create directory.
- Name it « Formation ».
- Go inside this directory next.


sftp://mleboulch@genotoul.toulouse.inra.fr - FileZilla
File Edit View Transfer Server Bookmarks Help
Host: Username: Password: Port: Quickconnect
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
Empty directory listing
Selected 1 directory.
Empty directory.
Server/Local file Direction Remote file Size Priority Status
Queued files Failed transfers Successful transfers
Queue: empty
Select the temp directory and drag and drop it following the arrow.
Double click on the merge directory.



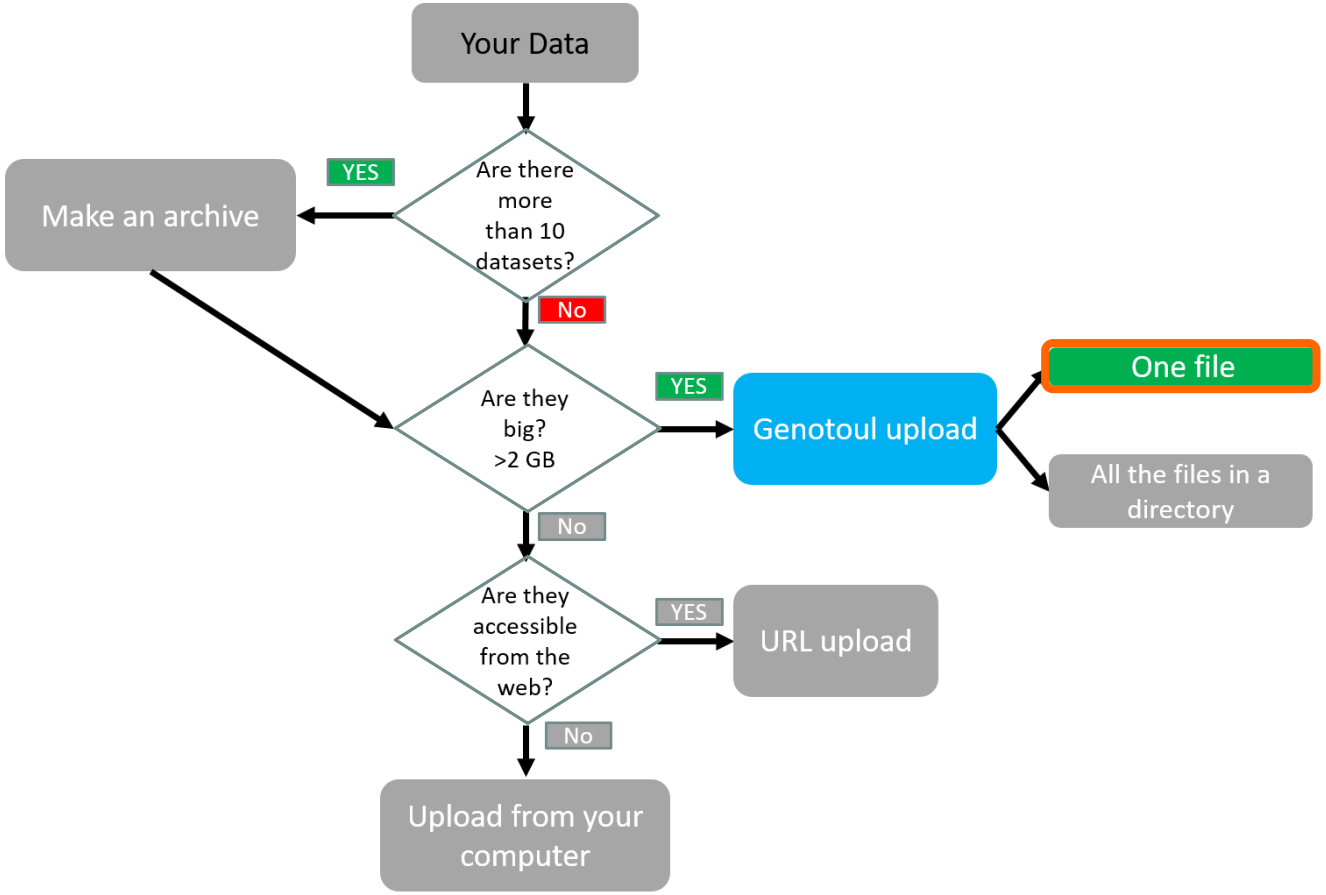
- Select the 100_1000seq_sample.tar.gz file and drag and drop it following the arrow.
- Double click on the merge directory.

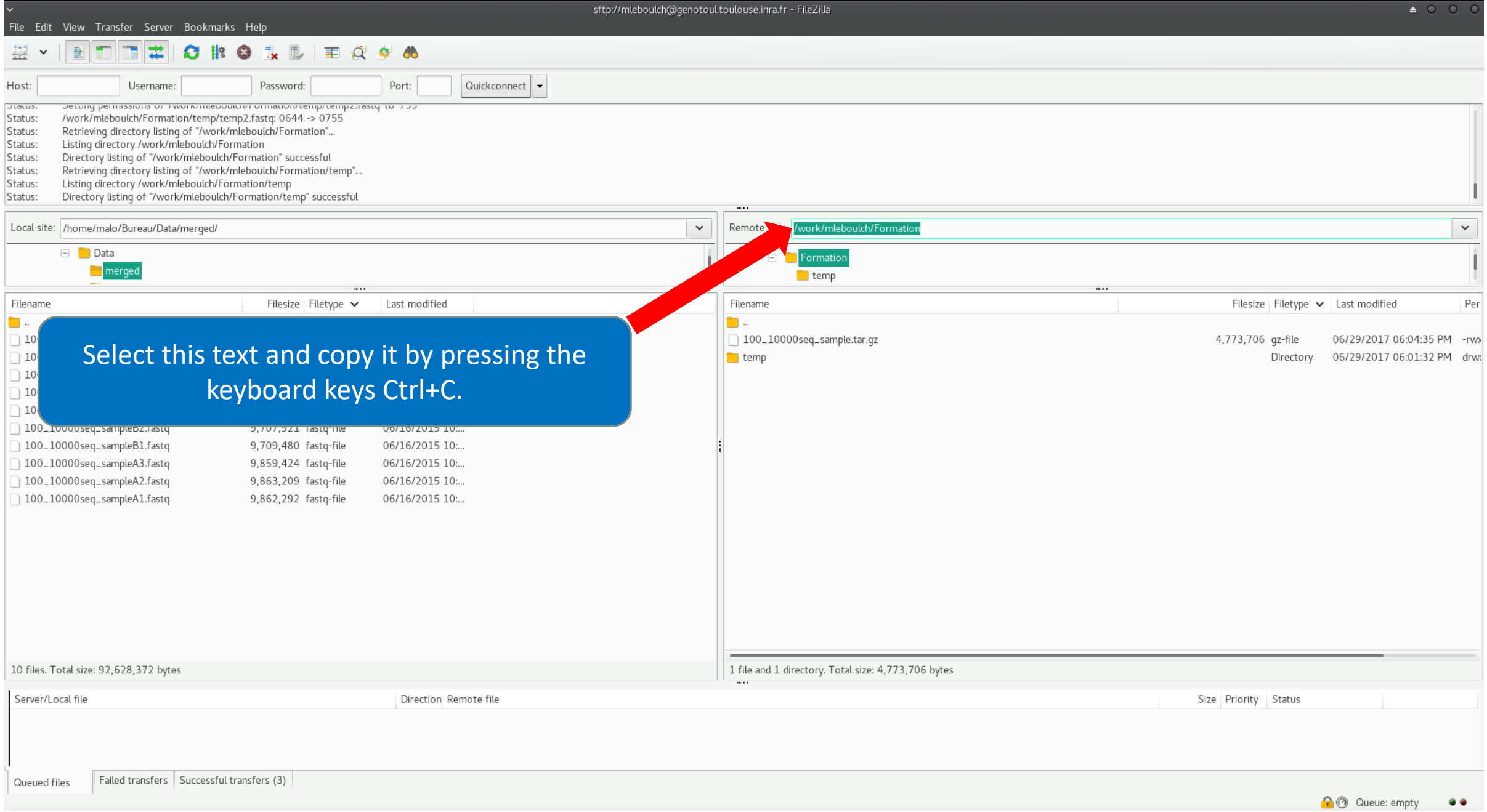

sftp://mleboulch@genotoul.toulouse.inra.fr - FileZilla
File Edit View Transfer Server Bookmarks Help
Host: Username: Password: Port: Quickconnect
Status: Directory listing of "/work/.vle" successful
Status: Retrieving directory listing of "/work/BPMP"...
Status: Listing directory /work/BPMP
Status: Directory 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/
Data
merged
Remote site: /work/mleboulch
mleboulch
Database
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:...
Filename Filesize Filetype Last modified Permis
Directory 06/08/2018 11:19:08 AM drwxr-x
Directory 03/14/2018 02:21:50 PM drwxr-x
Directory 03/13/2018 11:50:00 AM drwxr-x
Directory 03/09/2018 04:09:45 PM drwxr-x
Directory 03/06/2018 04:25:02 PM drwxr-x
Directory 03/06/2018 01:24:27 PM drwxr-x
Directory 02/28/2018 03:48:40 PM drwxr-x
Directory 02/28/2018 03:47:01 PM drwxr-x
Directory 02/28/2018 09:29:46 AM drwxr-x
Directory 10/30/2017 05:07:06 PM drwxr-x
Directory 06/30/2017 04:26:59 PM drwxr-x
11 directories
Download
Add to queue
Create directory
Create directory and enter it
Delete
Rename
Copy URL(s) to clipboard
File Attributes...
Select your work directory.
Right click on it and click on File Attributes.
Size Prio
Queued files Failed transfers Successful transfers
Queue: empty



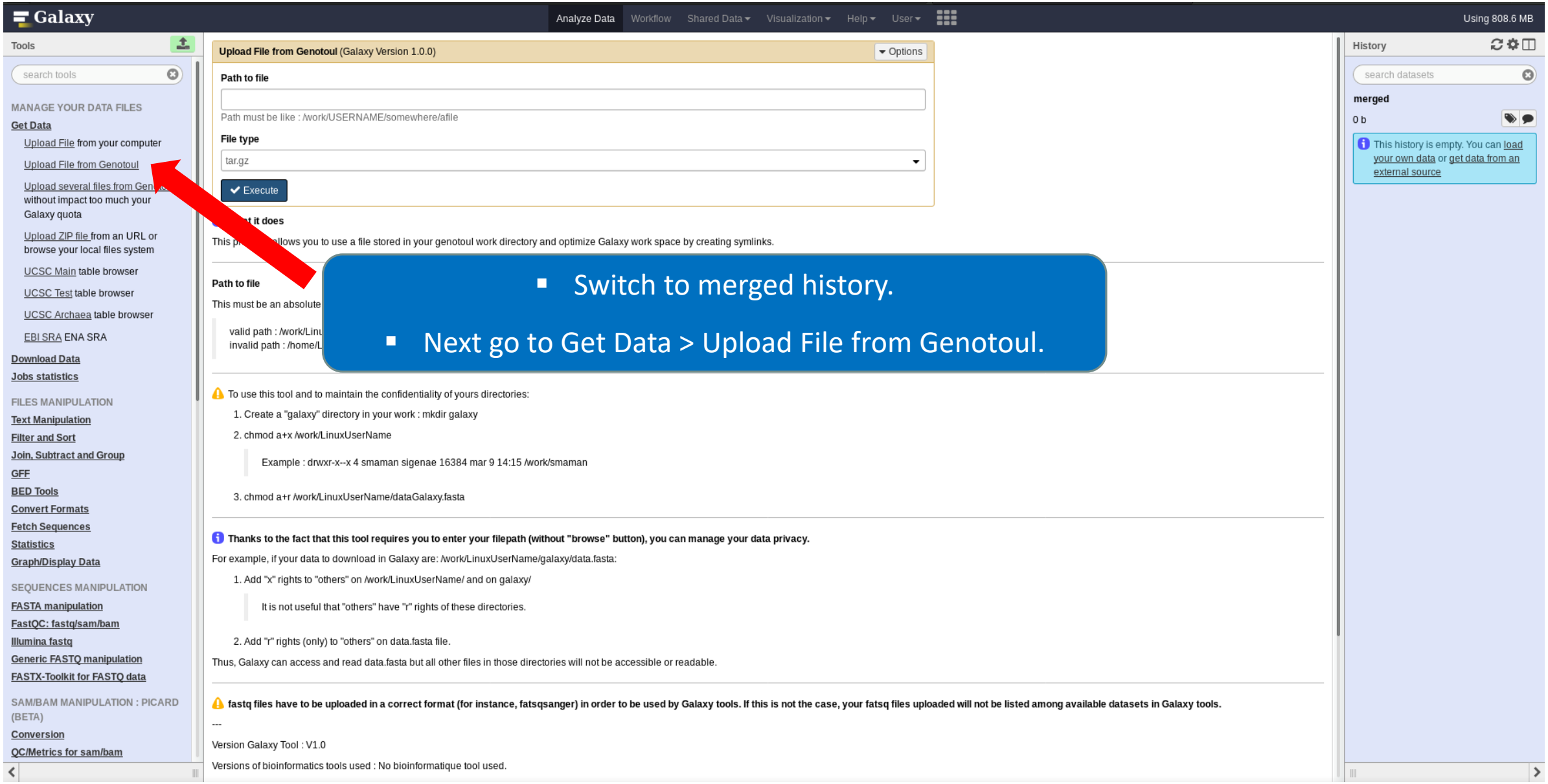
- Check that all boxes for execute and read are checked.
- Recurse this action to all files and subdirectories.
- That allows Galaxy to access your files on Genotoul.
- Click on Ok.

Upload files from Genotoul

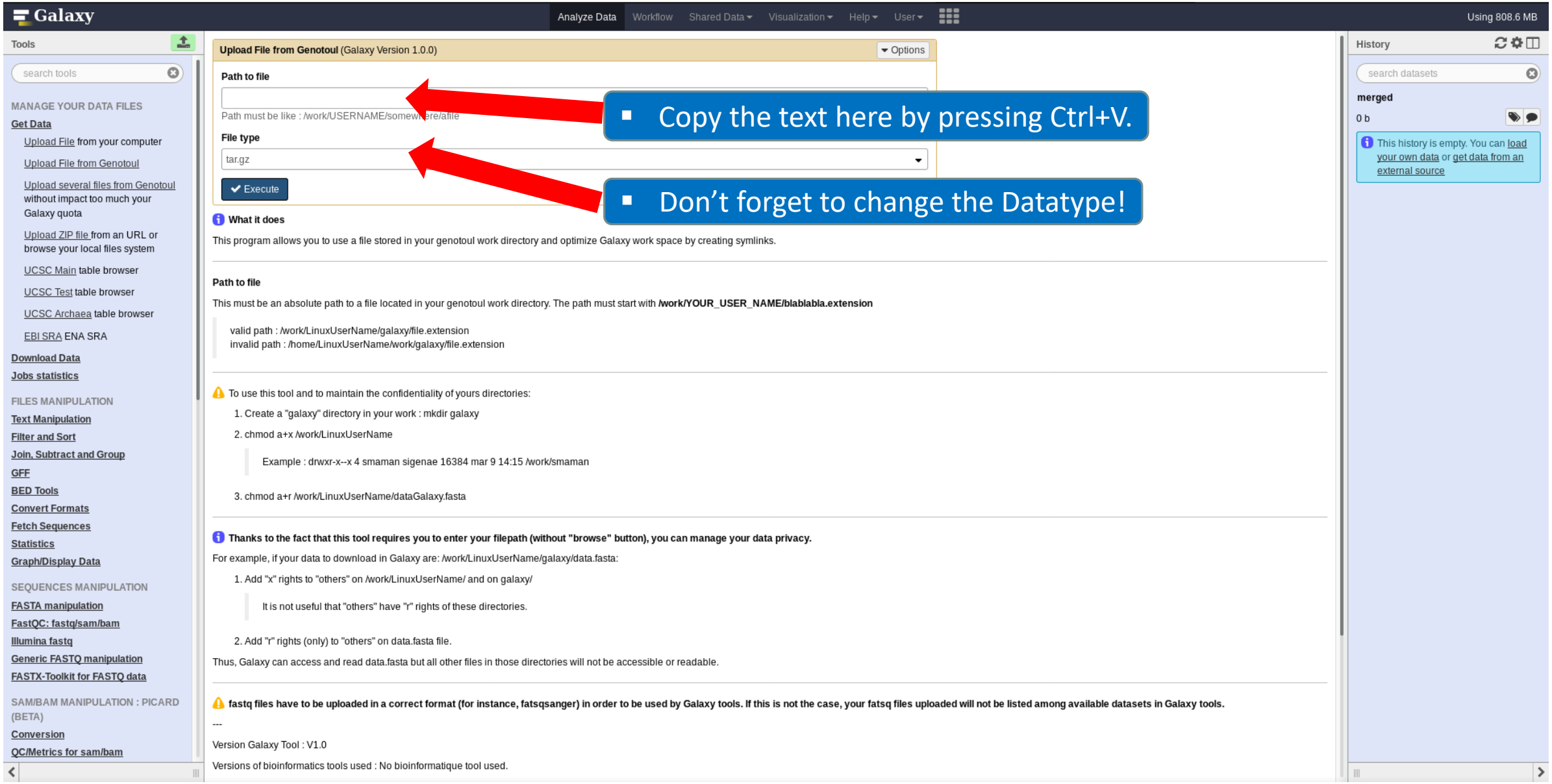




Select this text and copy it by pressing the keyboard keys Ctrl+C.

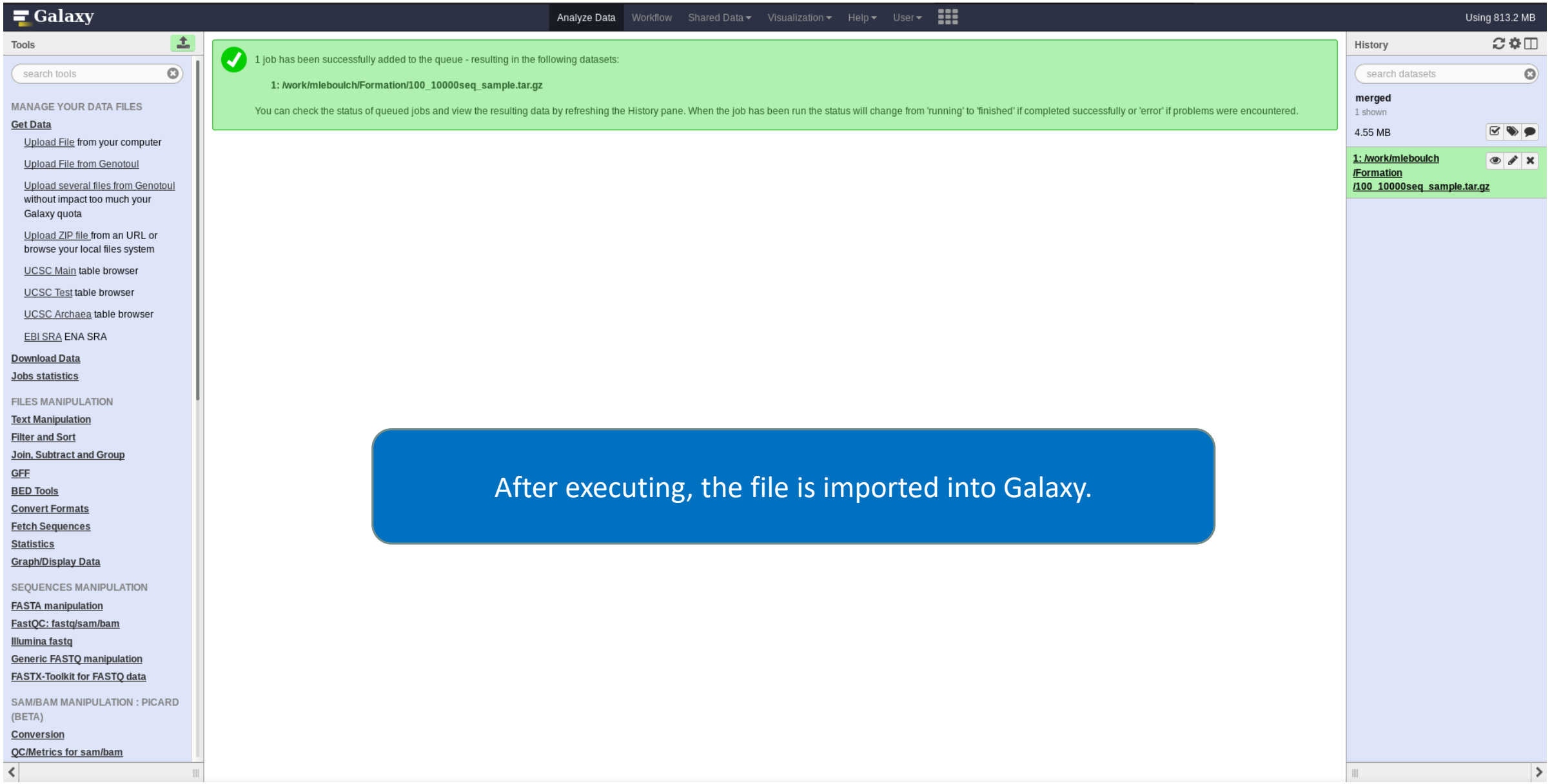

Galaxy
Analyze Data
Workflow
Shared Data
Visualization
Help
User
Using 808.6 MB
Tools
search tools
MANAGE YOUR DATA FILES
Get Data
Upload File from your computer
Upload File from Genotoul
Upload several files from Gen
without impact too much your Galaxy quota
Upload ZIP file from an URL or browse your local files system
UCSC Main table browser
UCSC Test table browser
UCSC Archaea table browser
EBI SRA ENA SRA
Download Data
Jobs statistics
FILES MANIPULATION
Text Manipulation
Filter and Sort
Join, Subtract and Group
GFF
BED Tools
Convert Formats
Fetch Sequences
Statistics
Graph/Display Data
SEQUENCES MANIPULATION
FASTA manipulation
FastQC: fastq/sam/bam
Illumina fastq
Generic FASTQ manipulation
FASTX-Toolkit for FASTQ data
SAM/BAM MANIPULATION : PICARD (BETA)
Conversion
QC/Metrics for sam/bam
Upload File from Genotoul (Galaxy Version 1.0.0)
Options
Path to file
Path must be like : /work/USERNAME/somewhere/afile
File type
tar.gz
Execute
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
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.
History
search datasets
merged
0 b
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.

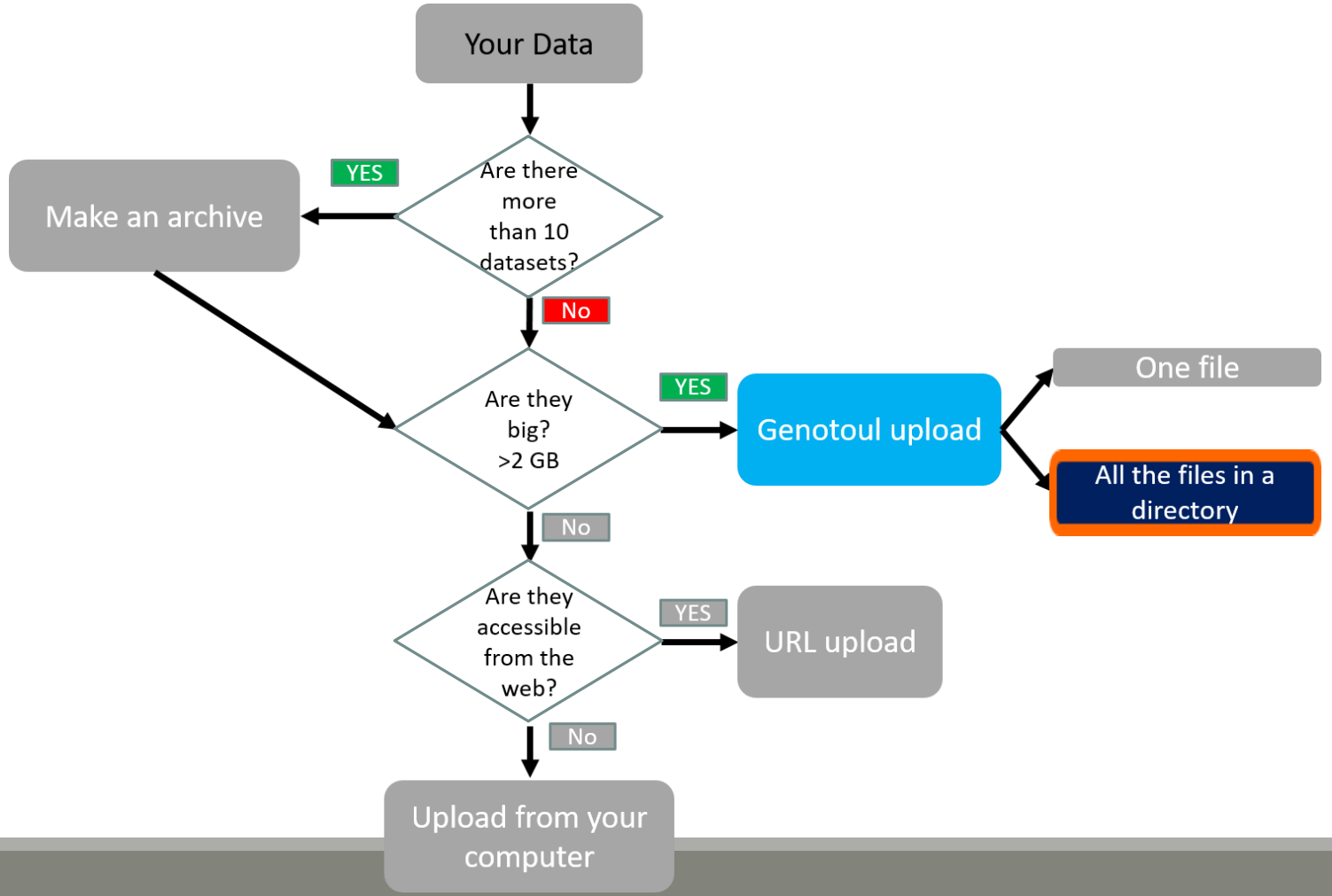

Galaxy
Analyze Data
Workflow
Shared Data
Visualization
Help
User
Using 808.6 MB
Tools
search tools
MANAGE YOUR DATA FILES
Get Data
Upload File from your computer
Upload File from Genotoul
Upload several files from Genotoul without impact too much your Galaxy quota
Upload ZIP file from an URL or browse your local files system
UCSC Main table browser
UCSC Test table browser
UCSC Archaea table browser
EBI SRA ENA SRA
Download Data
Jobs statistics
FILES MANIPULATION
Text Manipulation
Filter and Sort
Join, Subtract and Group
GFF
BED Tools
Convert Formats
Fetch Sequences
Statistics
Graph/Display Data
SEQUENCES MANIPULATION
FASTA manipulation
FastQC: fastq/sam/bam
Illumina fastq
Generic FASTQ manipulation
FASTX-Toolkit for FASTQ data
SAM/BAM MANIPULATION : PICARD (BETA)
Conversion
QC/Metrics for sam/bam
Upload File from Genotoul (Galaxy Version 1.0.0)
Options
Path to file
Path must be like : /work/USERNAME/somewhere/afile
File type
tar.gz
Execute
Copy the text here by pressing Ctrl+V.
Don’t forget to change the Datatype!
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.

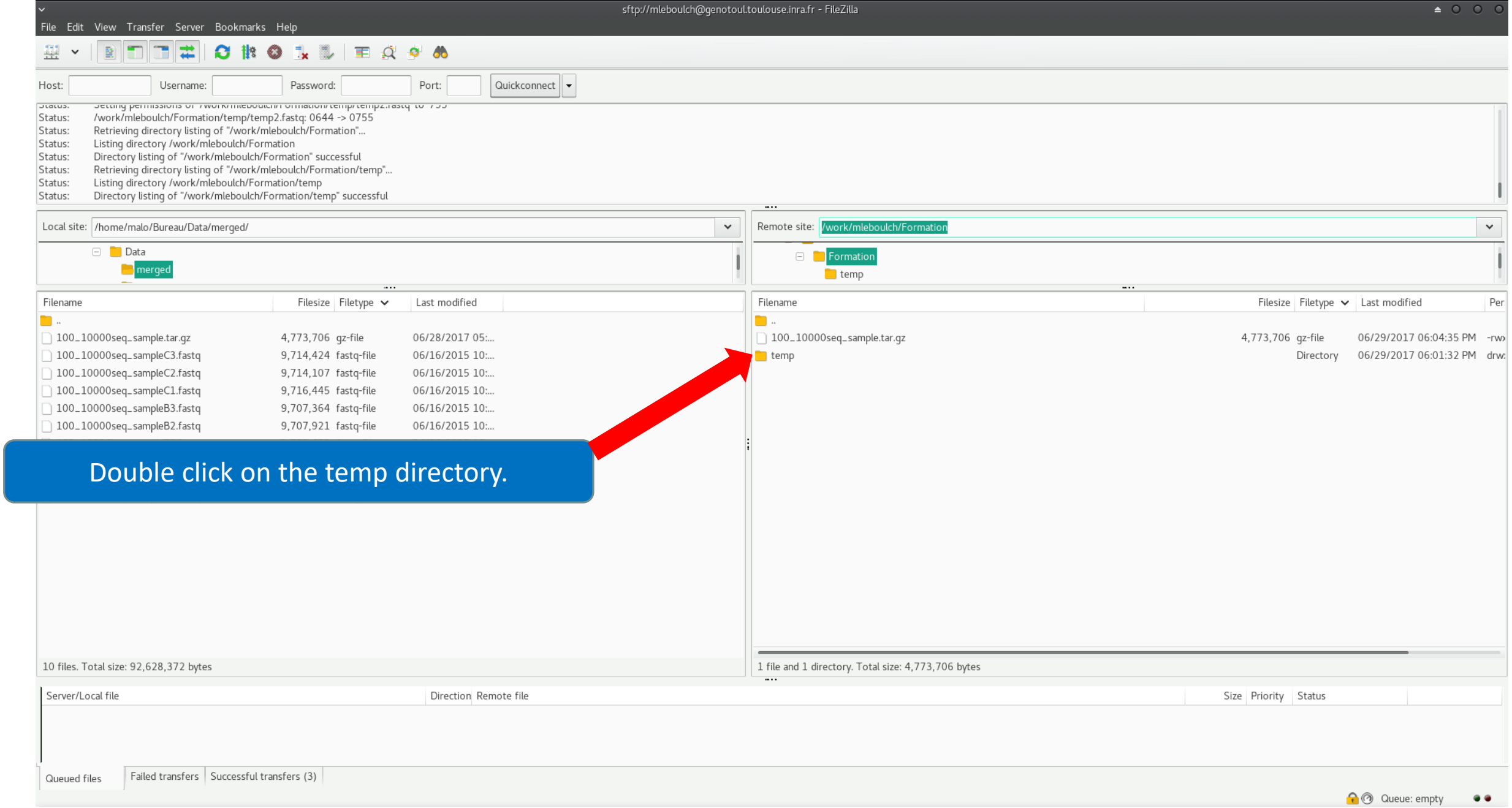
Version Galaxy Tool : V1.0
Versions of bioinformatics tools used : No bioinformatique tool used.
History
search datasets
merged
0 b
This history is empty. You can load your own data or get data from an external source

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

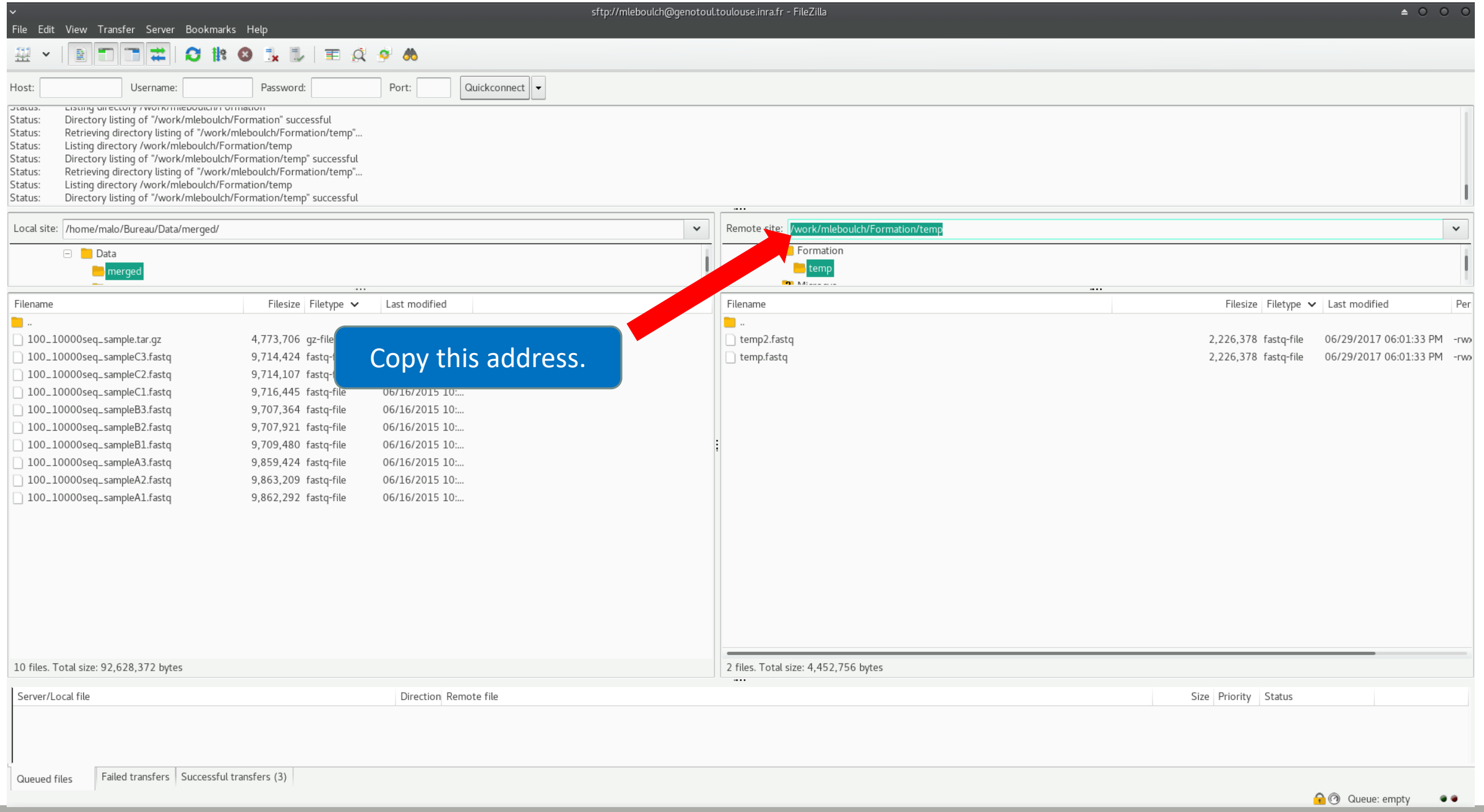

Galaxy
Analyze Data
Workflow
Shared Data
Visualization
Help
User
Using 813.2 MB
Tools
search tools
MANAGE YOUR DATA FILES
Get Data
Upload File from your computer
Upload File from Genotoul
Upload several files from Genotoul without impact too much your Galaxy quota
Upload ZIP file from an URL or browse your local files system
UCSC Main table browser
UCSC Test table browser
UCSC Archaea table browser
EBI SRA ENA SRA
Download Data
Jobs statistics
FILES MANIPULATION
Text Manipulation
Filter and Sort
Join, Subtract and Group
GFF
BED Tools
Convert Formats
Fetch Sequences
Statistics
Graph/Display Data
SEQUENCES MANIPULATION
FASTA manipulation
FastQC: fastq/sam/bam
Illumina fastq
Generic FASTQ manipulation
FASTX-Toolkit for FASTQ data
SAM/BAM MANIPULATION : PICARD (BETA)
Conversion
QC/Metrics for sam/bam
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.
History
search datasets
merged
1 shown
4.55 MB
1: /work/mleboulch /Formation /100_10000seq_sample.tar.gz
After executing, the file is imported into Galaxy.

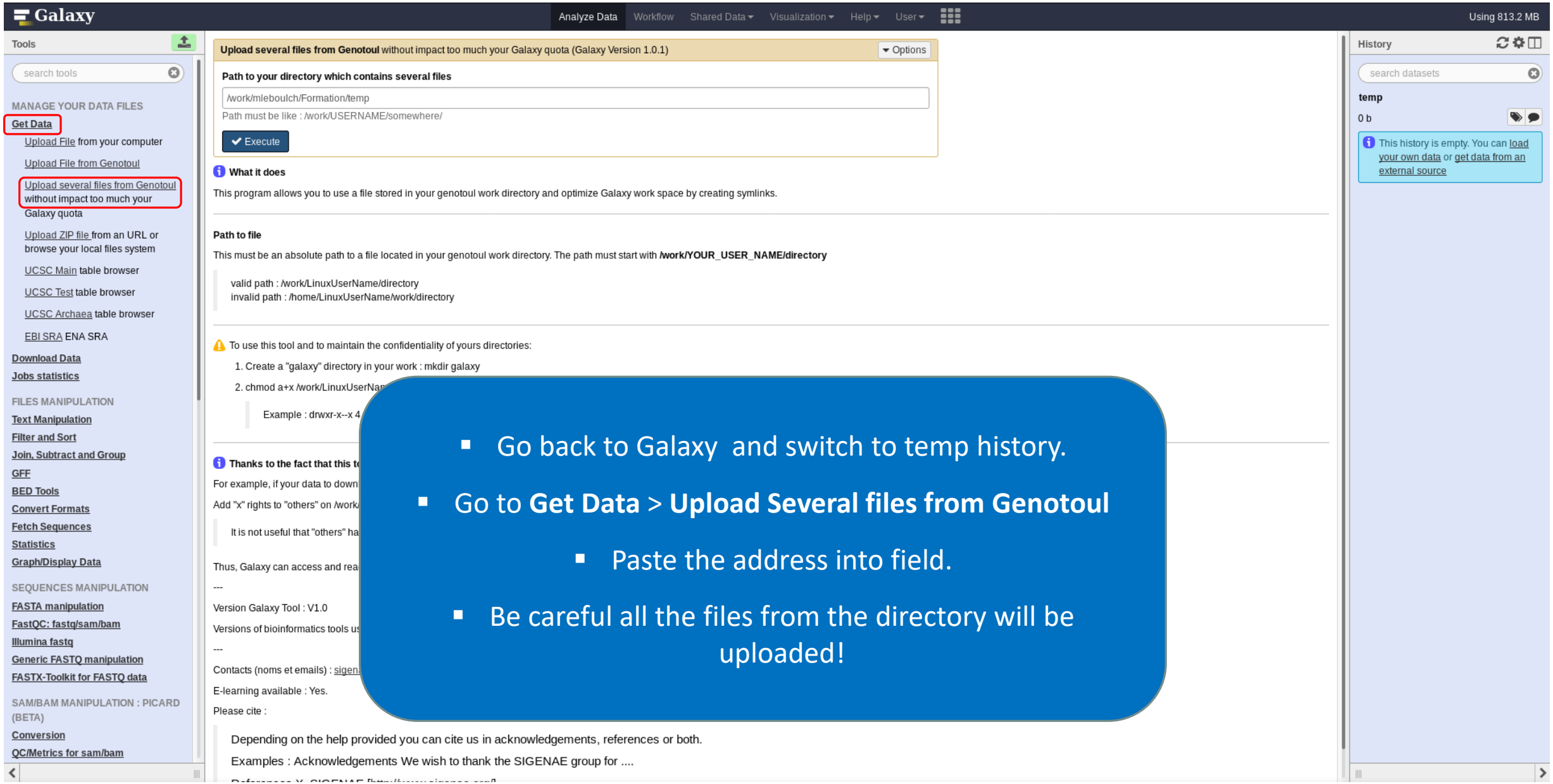
Upload files from Genotoul



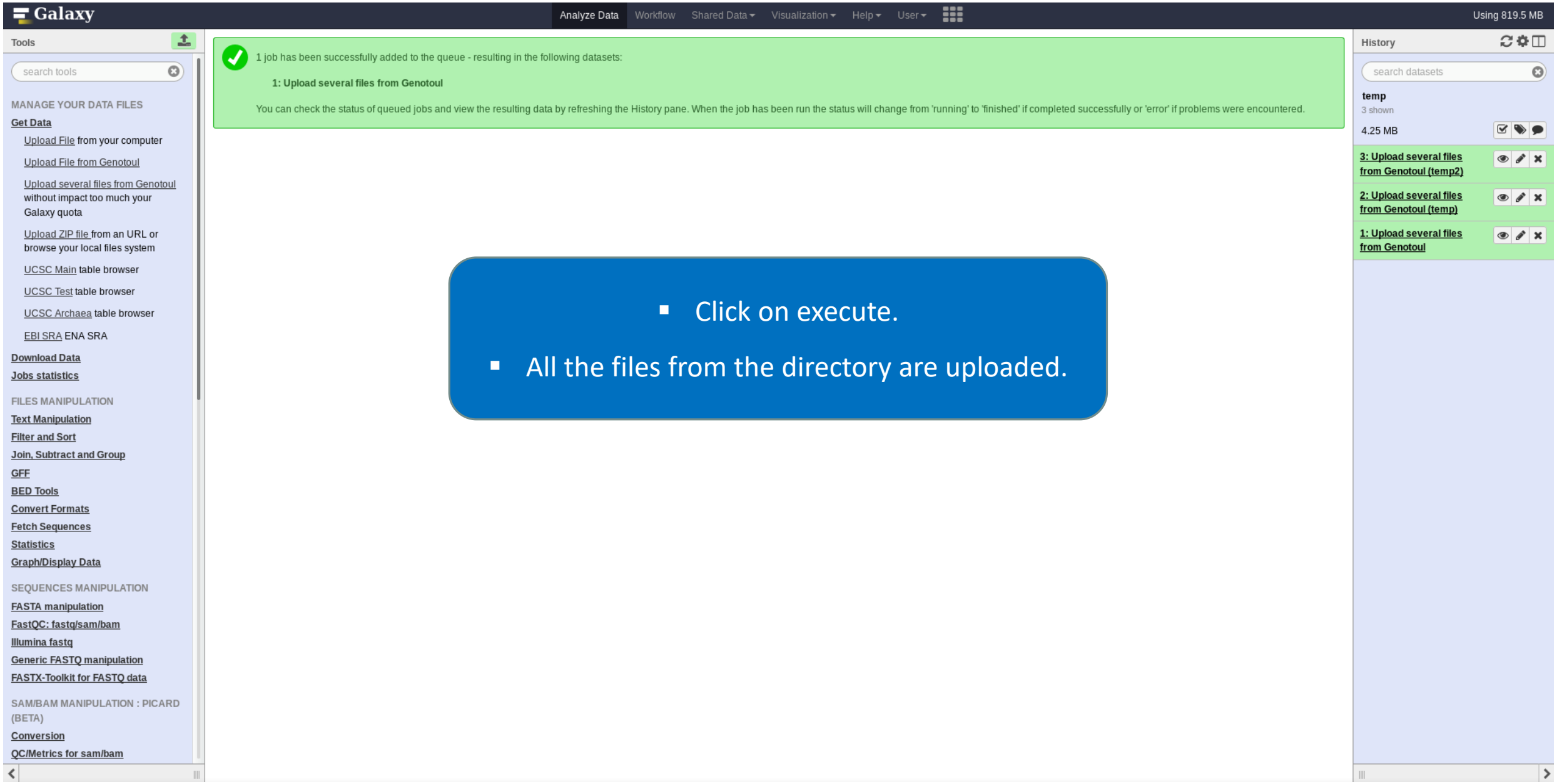


Double click on the temp directory.


sftp://mleboulch@genotoul.toulouse.inra.fr - FileZilla
File Edit View Transfer Server Bookmarks Help
Host: Username: Password: Port: Quickconnect
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
Remote site: /work/mleboulch/Formation/temp
Formation
temp
Filename Filesize Filetype Last modified
100_10000seq_sample.tar.gz 4,773,706 gz-file
100_10000seq_sampleC3.fastq 9,714,424 fastq-
100_10000seq_sampleC2.fastq 9,714,107 fastq-
100_10000seq_sampleC1.fastq 9,716,445 fastq-file
100_10000seq_sampleB3.fastq 9,707,364 fastq-file
100_10000seq_sampleB2.fastq 9,707,921 fastq-file
100_10000seq_sampleB1.fastq 9,709,480 fastq-file
100_10000seq_sampleA3.fastq 9,859,424 fastq-file
100_10000seq_sampleA2.fastq 9,863,209 fastq-file
100_10000seq_sampleA1.fastq 9,862,292 fastq-file
10 files. Total size: 92,628,372 bytes
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
Copy this address.
Server/Local file Direction Remote file Size Priority Status
Queued files Failed transfers Successful transfers (3)
Queue: empty

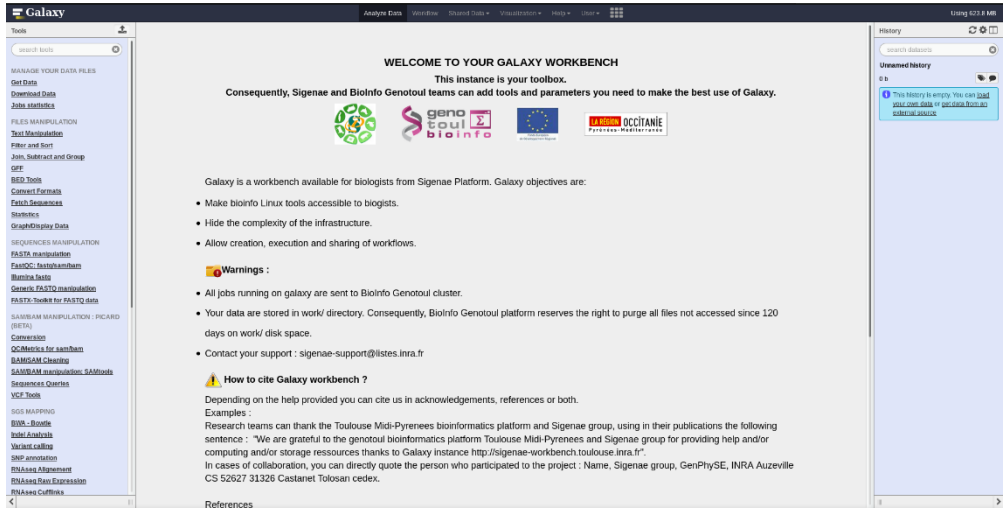


- Go back to Galaxy and switch to temp history.
- Go to **Get Data** > **Upload Several files from Genotoul**
- Paste the address into field.
- Be careful all the files from the directory will be uploaded!



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

Upload by Genotoul



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

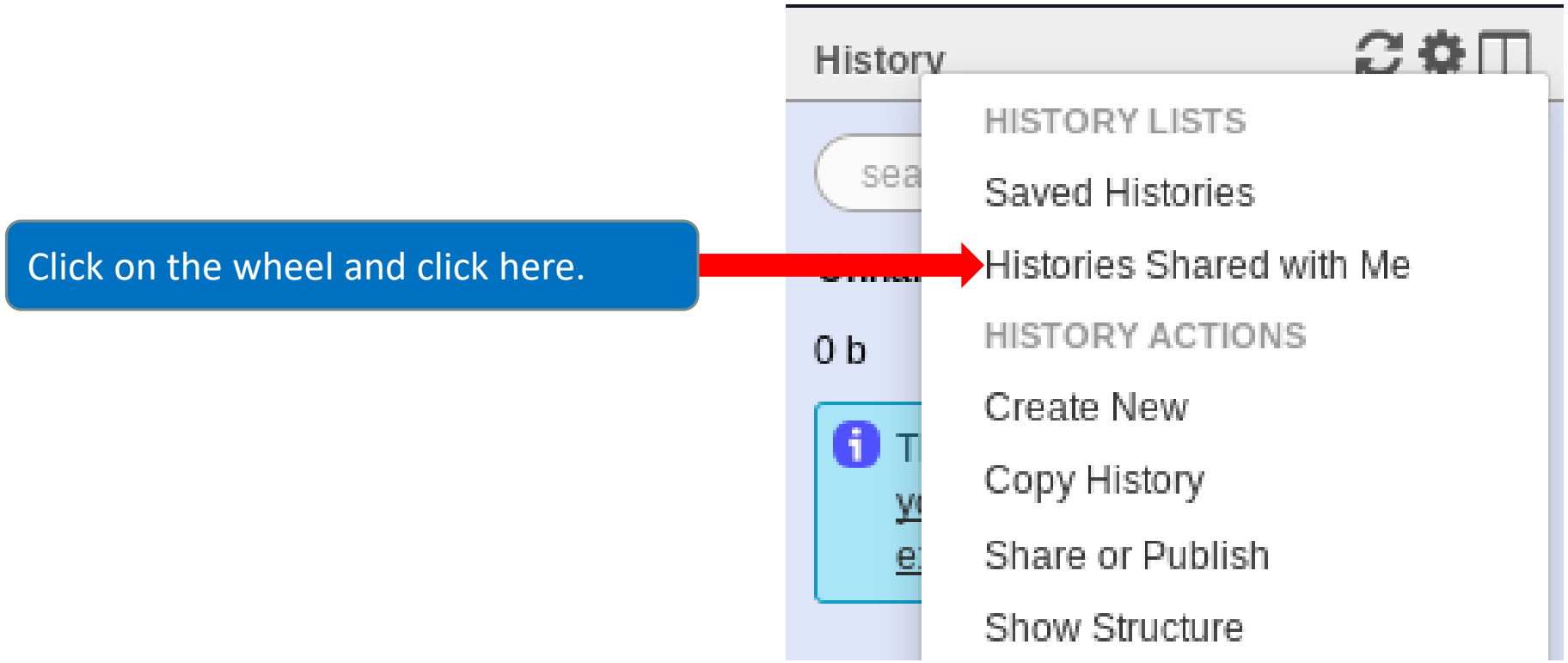
Share a history

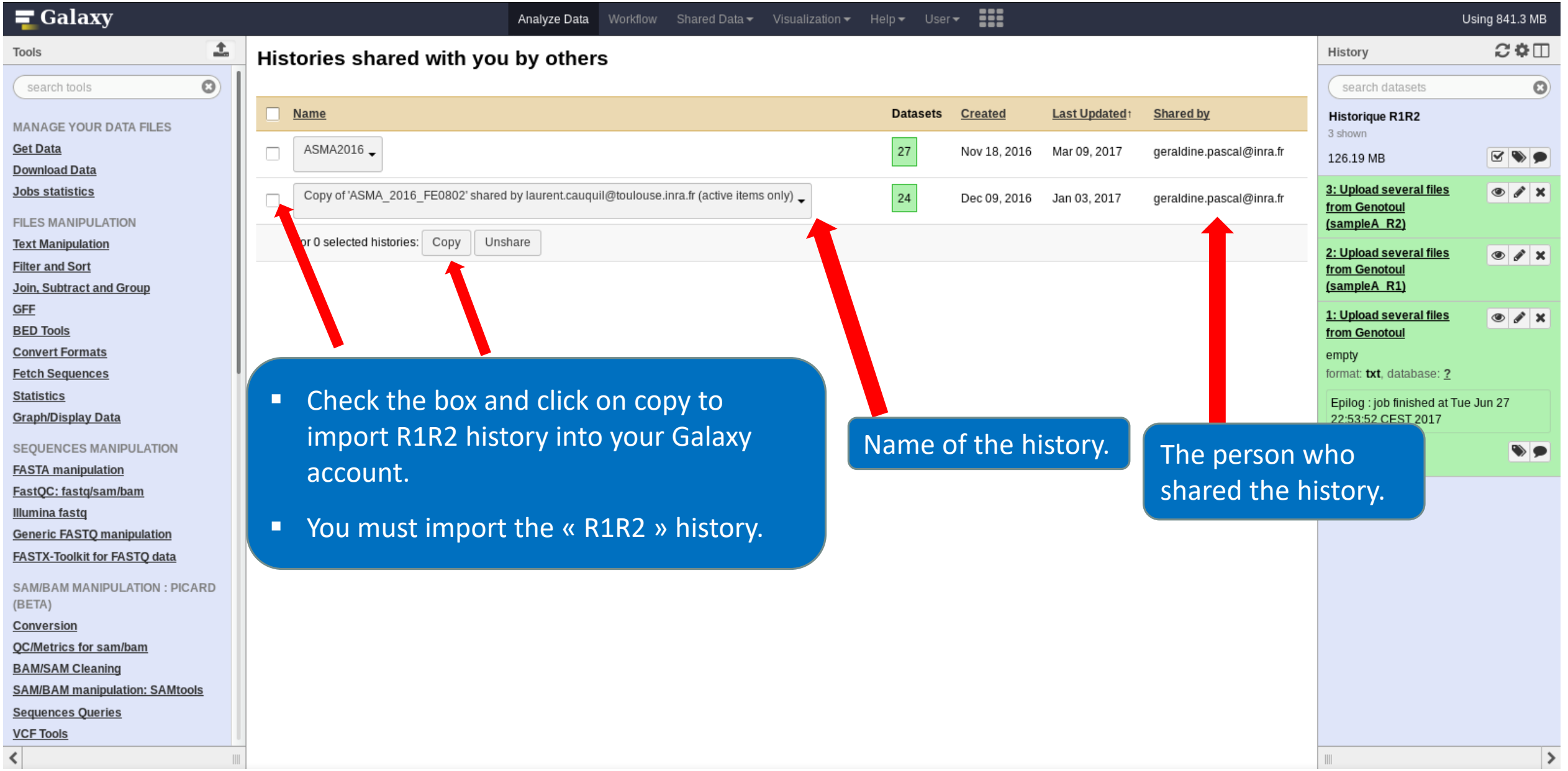
- You can share a history with another Galaxy user:
  - For working with your colleague.
  - For support, so we can help you better and faster.
- You can import shared history to your account too.

Your Turn!

IMPORT A SHARED HISTORY TO YOUR ACCOUNT

Import a shared history



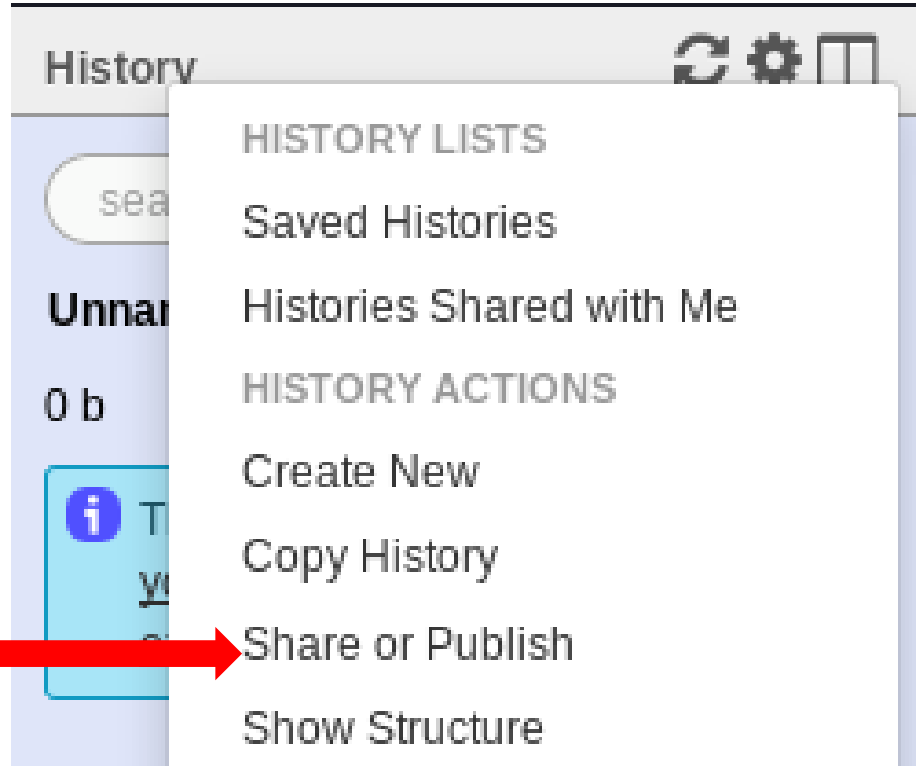

Galaxy
Analyze Data
Workflow
Shared Data
Visualization
Help
User
Using 841.3 MB
Tools
search tools
MANAGE YOUR DATA FILES
Get Data
Download Data
Jobs statistics
FILES MANIPULATION
Text Manipulation
Filter and Sort
Join, Subtract and Group
GFF
BED Tools
Convert Formats
Fetch Sequences
Statistics
Graph/Display Data
SEQUENCES MANIPULATION
FASTA manipulation
FastQC: fastq/sam/bam
Illumina fastq
Generic FASTQ manipulation
FASTX-Toolkit for FASTQ data
SAM/BAM MANIPULATION : PICARD (BETA)
Conversion
QC/Metrics for sam/bam
BAM/SAM Cleaning
SAM/BAM manipulation: SAMtools
Sequences Queries
VCF Tools
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
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.
History
search datasets
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

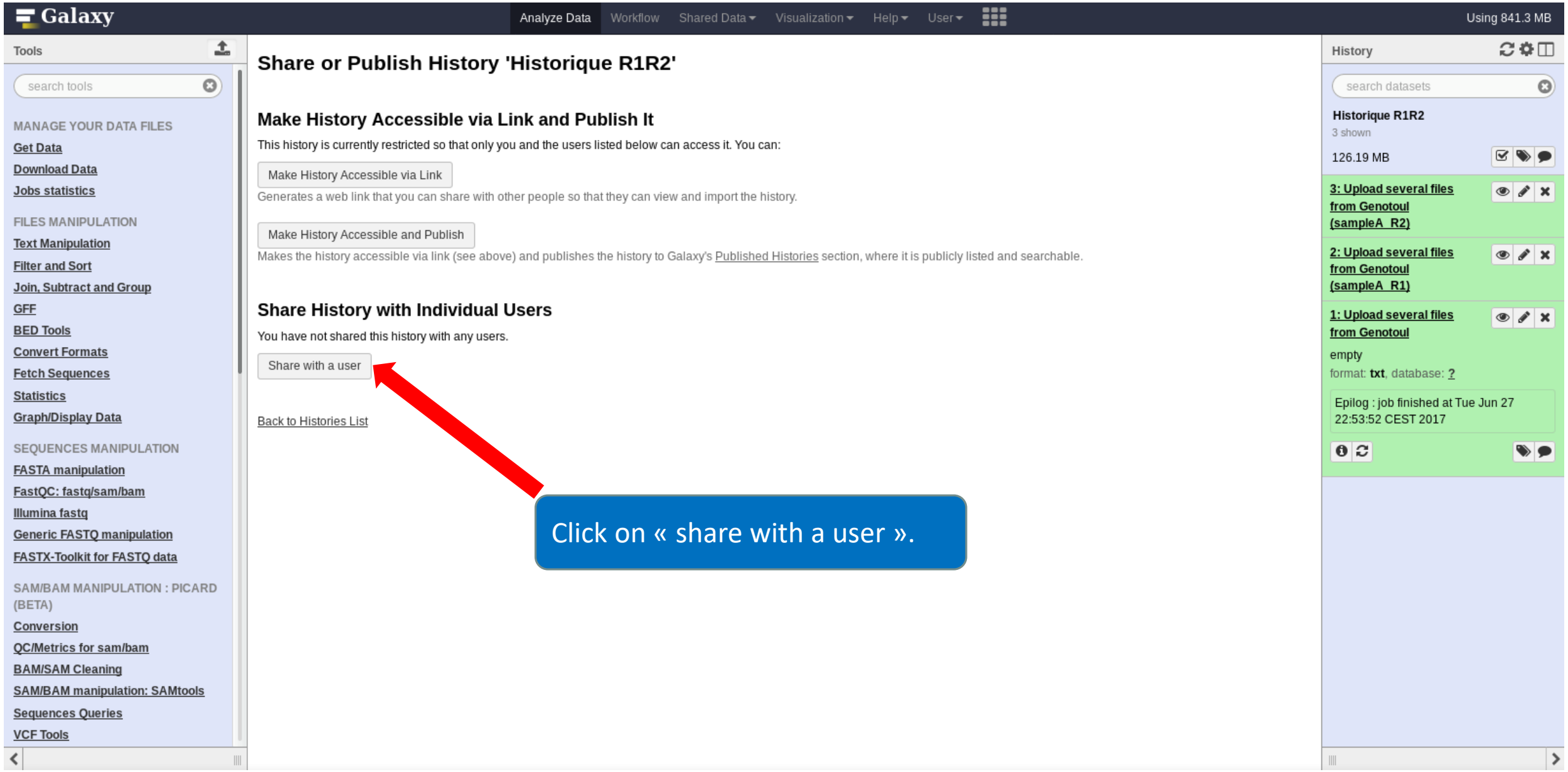
Your Turn!

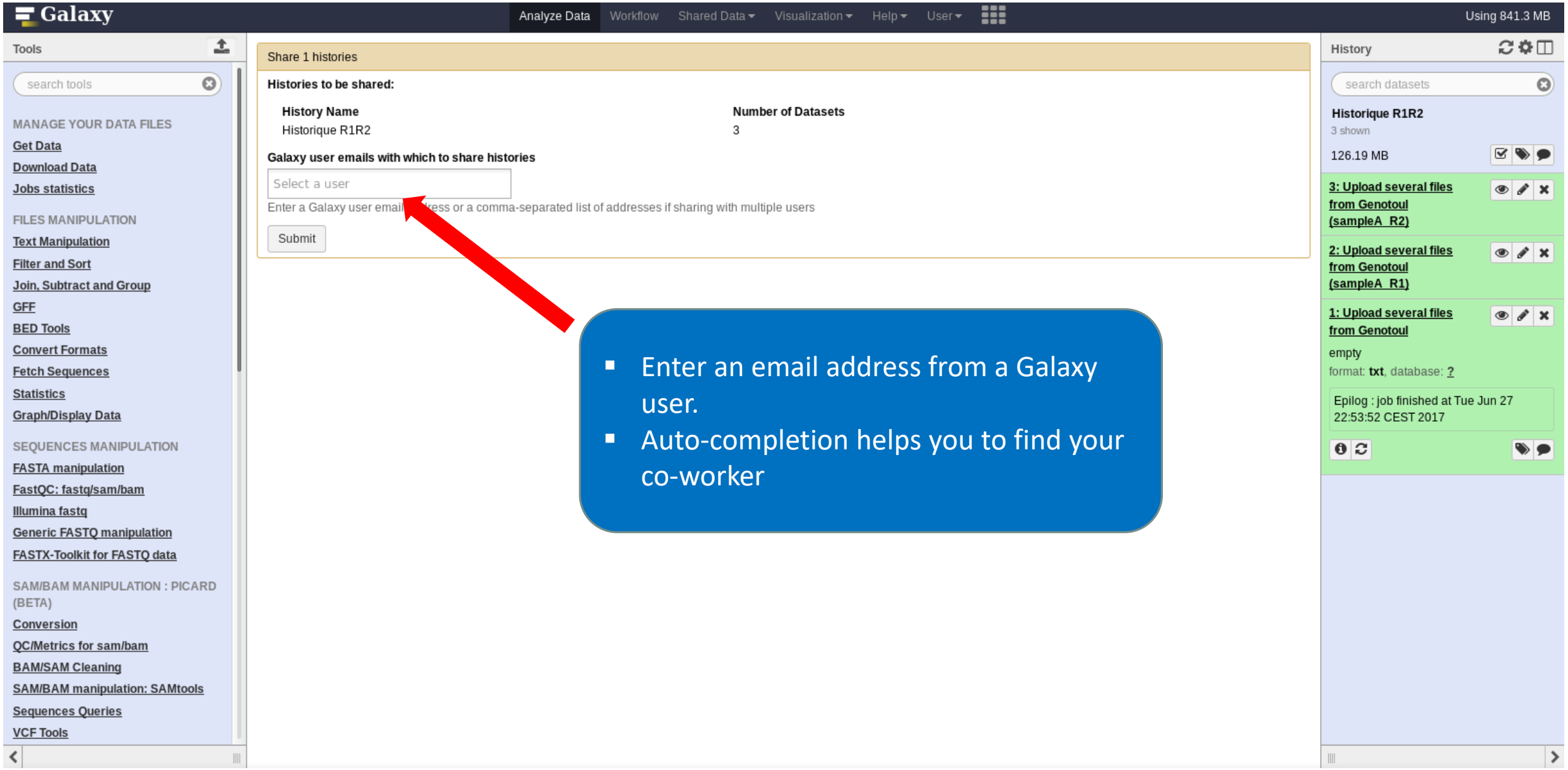
SHARE A HISTORY WITH YOUR NEIGHBOUR

Share a history

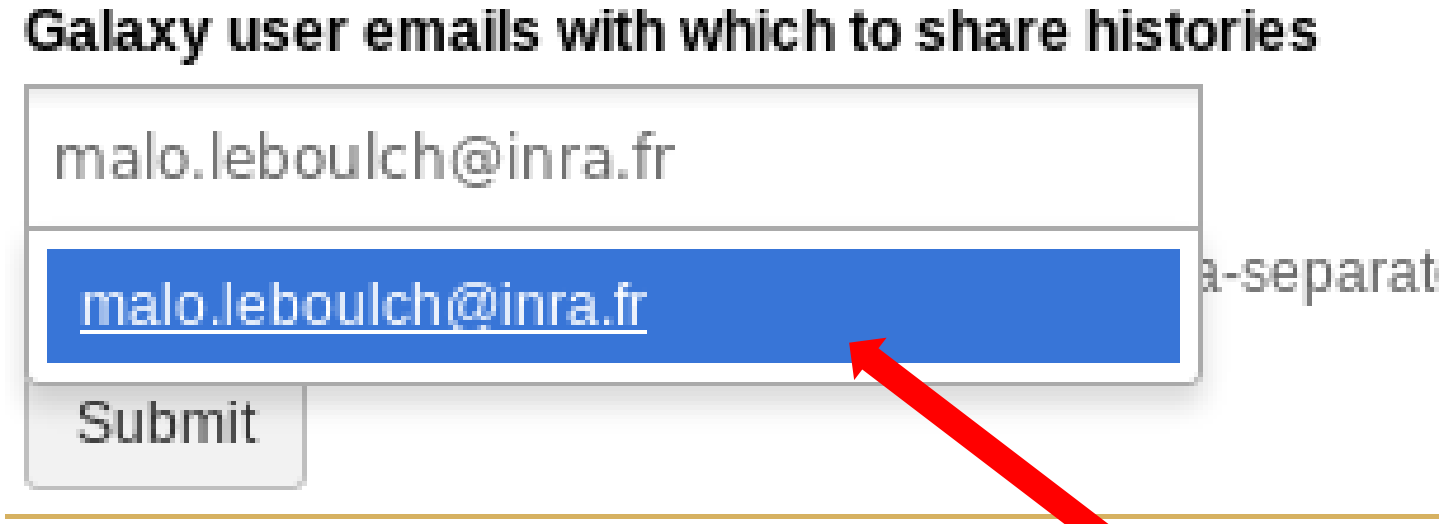
- Switch to the history you want to share.
- Click on the wheel and click here.




Galaxy
Analyze Data
Workflow
Shared Data
Visualization
Help
User
Using 841.3 MB
Tools
search tools
MANAGE YOUR DATA FILES
Get Data
Download Data
Jobs statistics
FILES MANIPULATION
Text Manipulation
Filter and Sort
Join, Subtract and Group
GFF
BED Tools
Convert Formats
Fetch Sequences
Statistics
Graph/Display Data
SEQUENCES MANIPULATION
FASTA manipulation
FastQC: fastq/sam/bam
Illumina fastq
Generic FASTQ manipulation
FASTX-Toolkit for FASTQ data
SAM/BAM MANIPULATION : PICARD (BETA)
Conversion
QC/Metrics for sam/bam
BAM/SAM Cleaning
SAM/BAM manipulation: SAMtools
Sequences Queries
VCF Tools
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
Click on « share with a user ».
History
search datasets
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



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

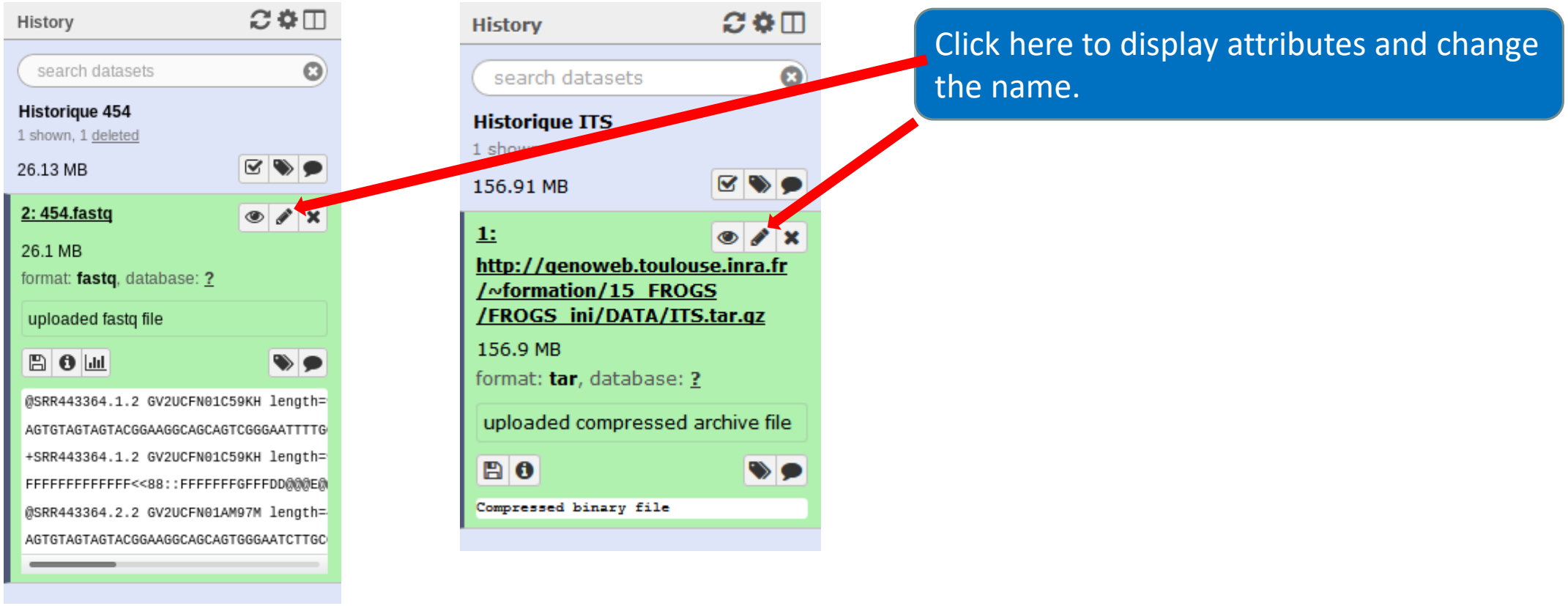
Manipulate datasets

Your Turn!

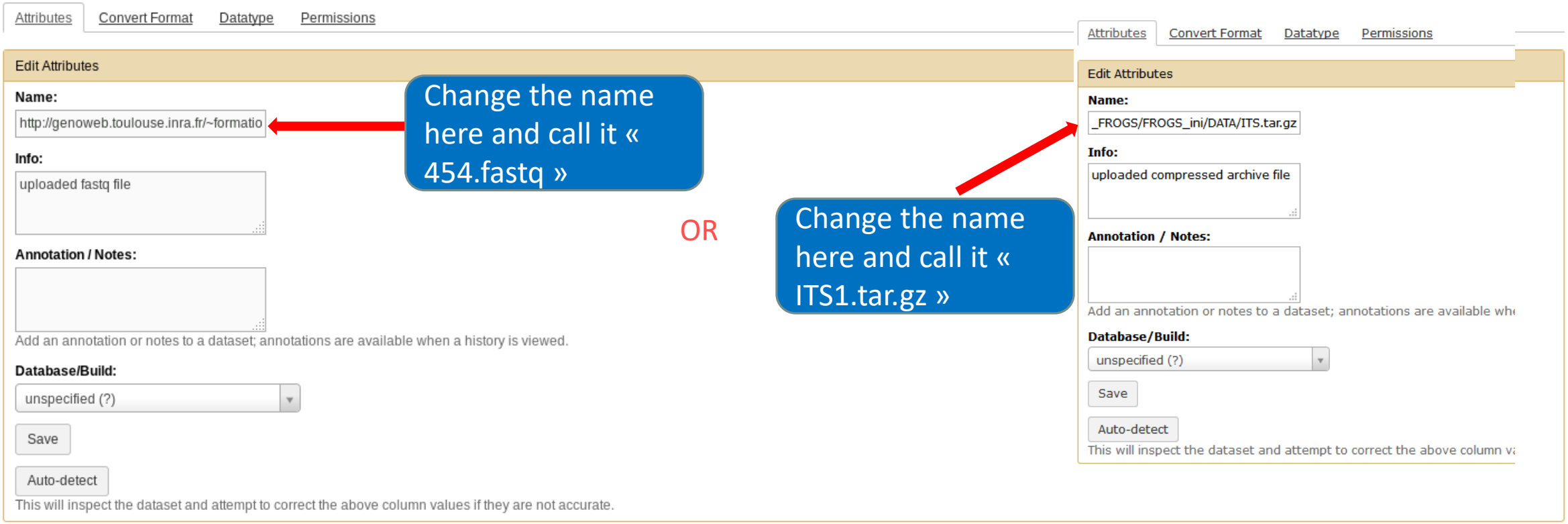
RENAME A DATASET

Rename a dataset

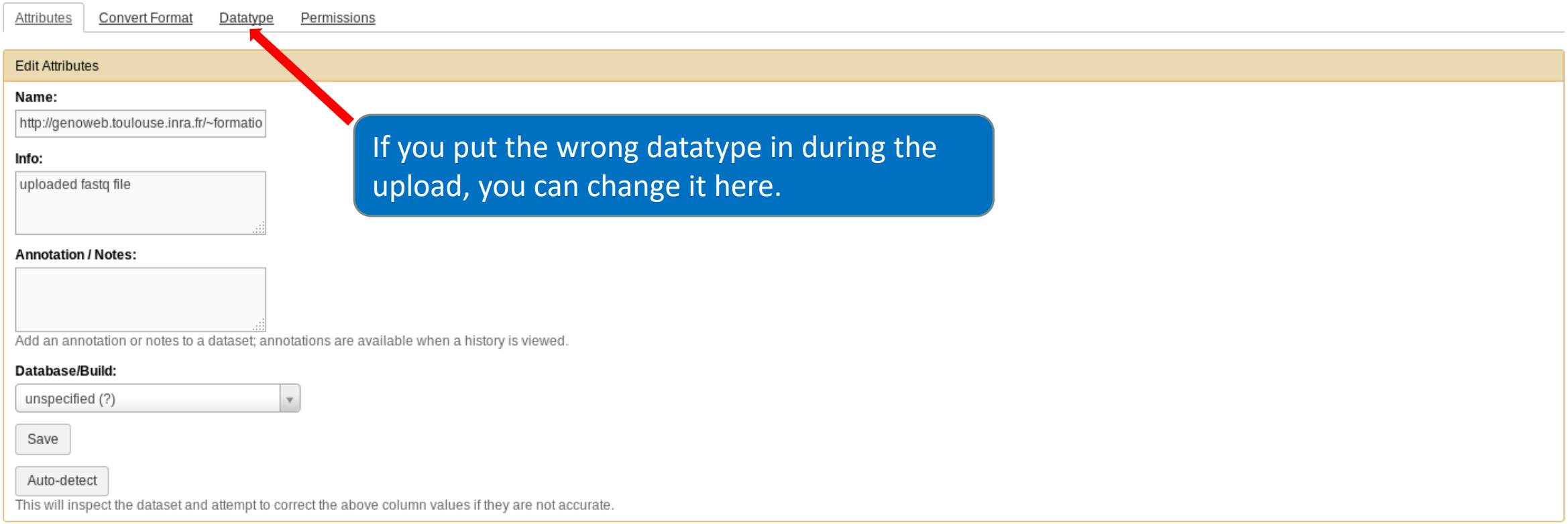
- Switch to 454 history or ITS history



Rename a dataset



Rename a dataset



Rename a dataset

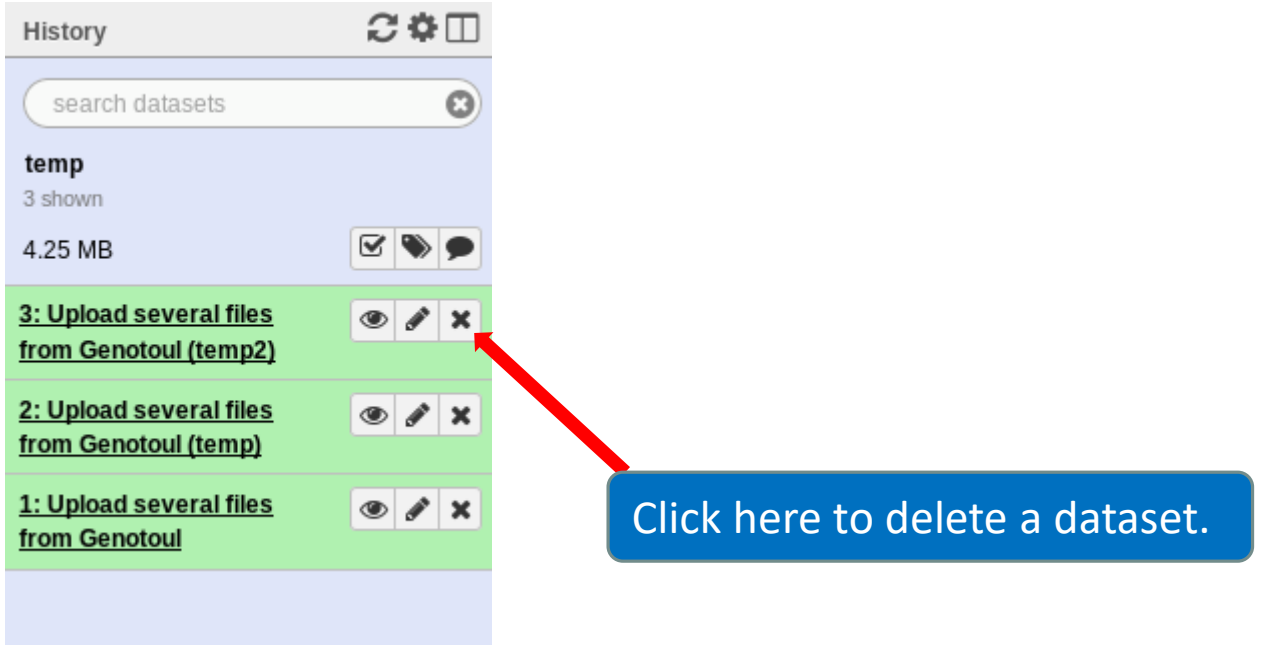
- Do the same with the merged history:
  - Switch to the merged history.
  - Change the name of the file to «100_10000seq_sample.tar.gz ».

Your Turn!

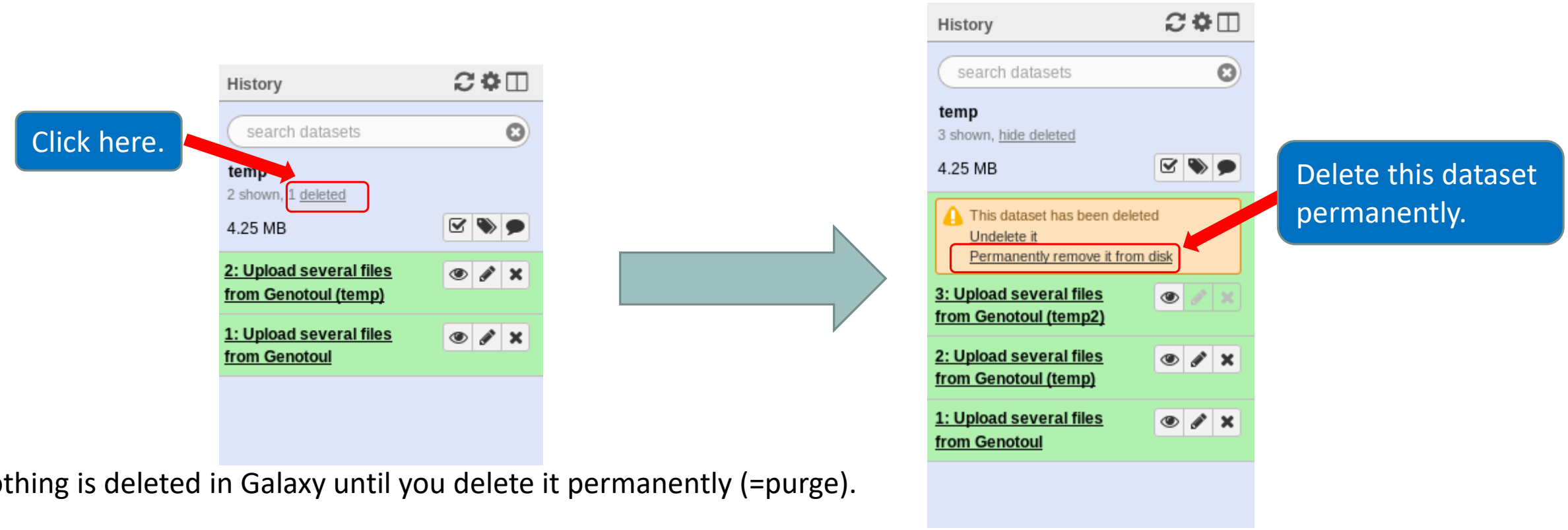
DELETE A DATASET

Delete a dataset

- Switch to temp history.



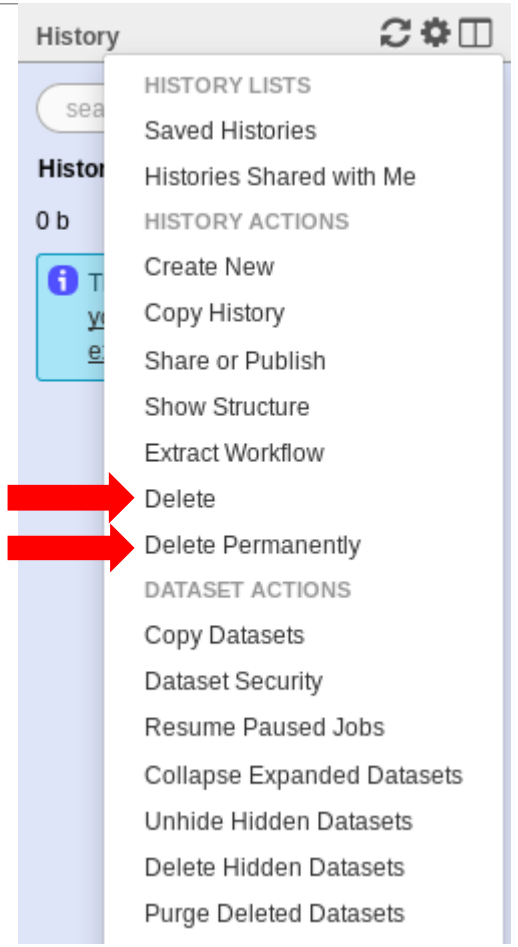
Delete a dataset

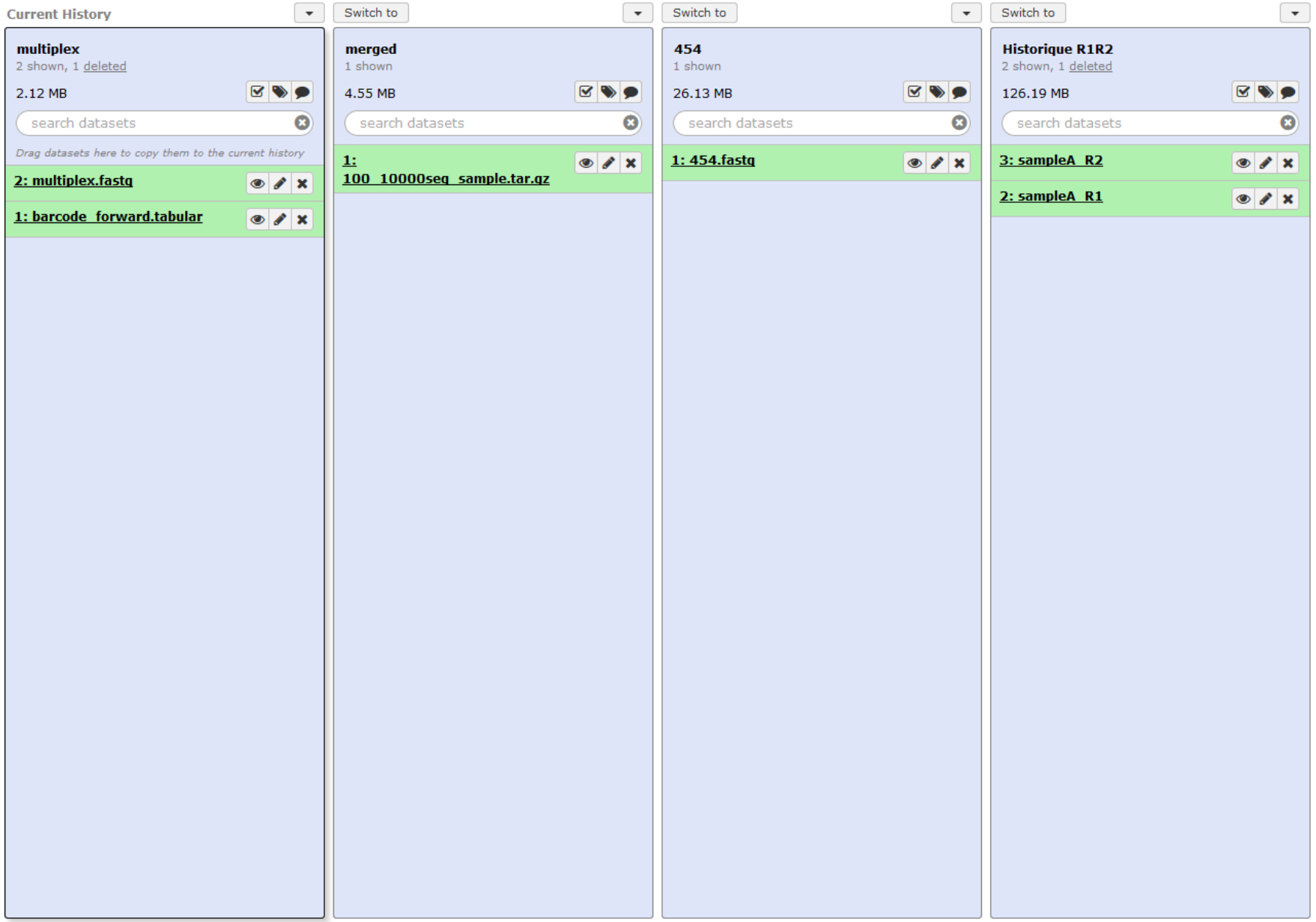


Nothing is deleted in Galaxy until you delete it permanently (=purge).

How to delete a history?

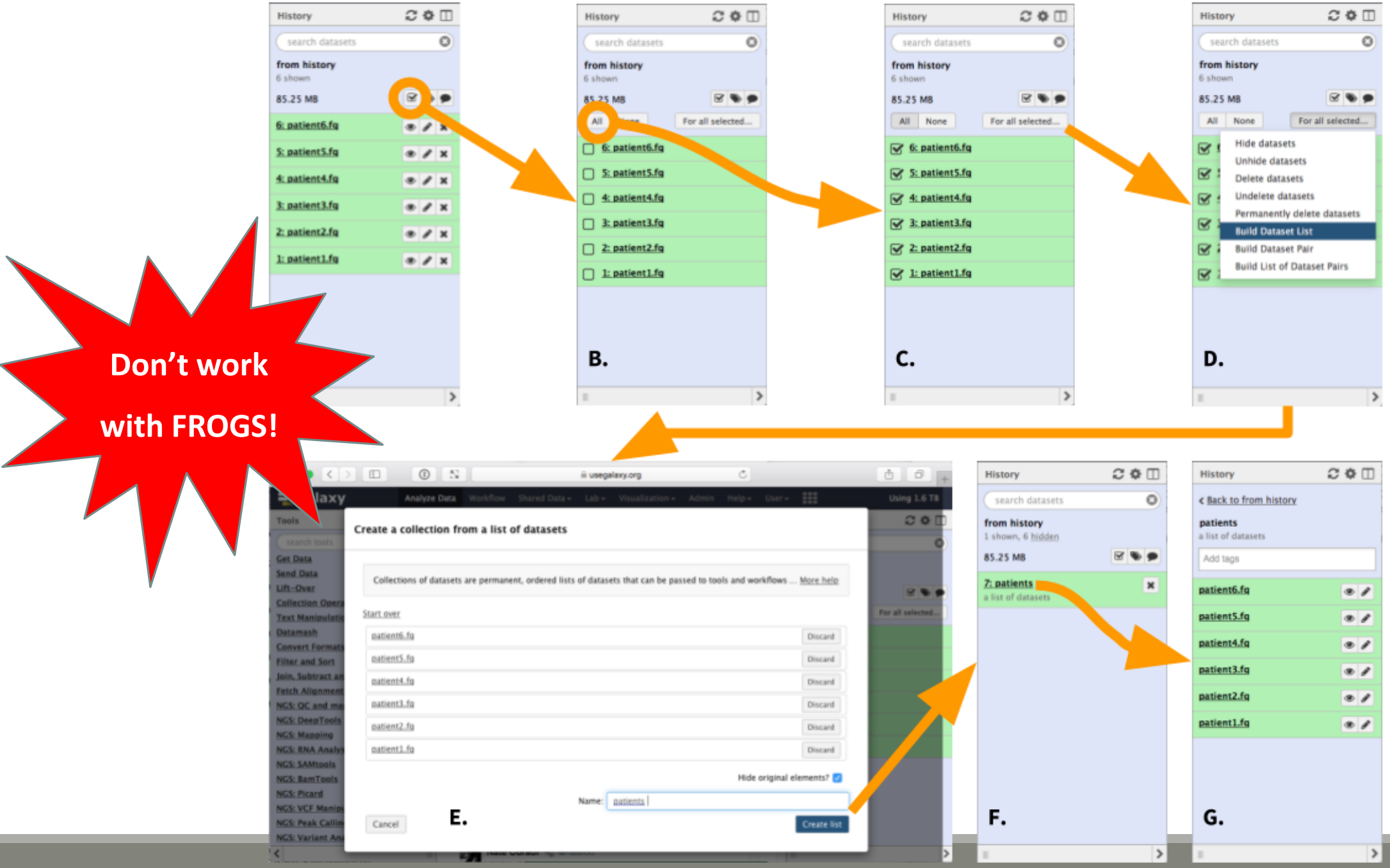
- Stay in the temp history.
- Click on the wheel.
- Click on delete.
- A deleted item on Galaxy is recoverable.
- To definitively delete it:
  click on « Delete Permanently ».




Current History
Switch to
Switch to
Switch to
multiplex
2 shown, 1 deleted
2.12 MB
search datasets
Drag datasets here to copy them to the current history
2: multiplex.fastq
1: barcode_forward.tabular
merged
1 shown
4.55 MB
search datasets
1: 100_10000seq_sample.tar.gz
454
1 shown
26.13 MB
search datasets
1: 454.fastq
Historique R1R2
2 shown, 1 deleted
126.19 MB
search datasets
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!
History
search datasets
from history
6 shown
85.25 MB
6: patient6.fq
5: patient5.fq
4: patient4.fq
3: patient3.fq
2: patient2.fq
1: patient1.fq
B.
All
None
For all selected...
C.
D.
Hide datasets
Unhide datasets
Delete datasets
Undelete datasets
Permanently delete datasets
Build Dataset List
Build Dataset Pair
Build List of Dataset Pairs
usegalaxy.org
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
Start over
Discard
Hide original elements?
Name:
patients
Cancel
Create list
E.
1 shown, 6 hidden
7: patients
a list of datasets
F.
Back to from history
patients
Add tags
G.

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.