HPRC Galaxy
Maroon, Silver, Kaiser

Your HPRC Galaxy Username and Password

- Your HPRC Galaxy login is the same as your TAMU account NetID and password
- Both state of Texas law and TAMU regulations prohibit the sharing and/or illegal use of computer passwords and accounts
- Don’t write down passwords
- Don’t choose easy to guess/crack passwords

For More Help...

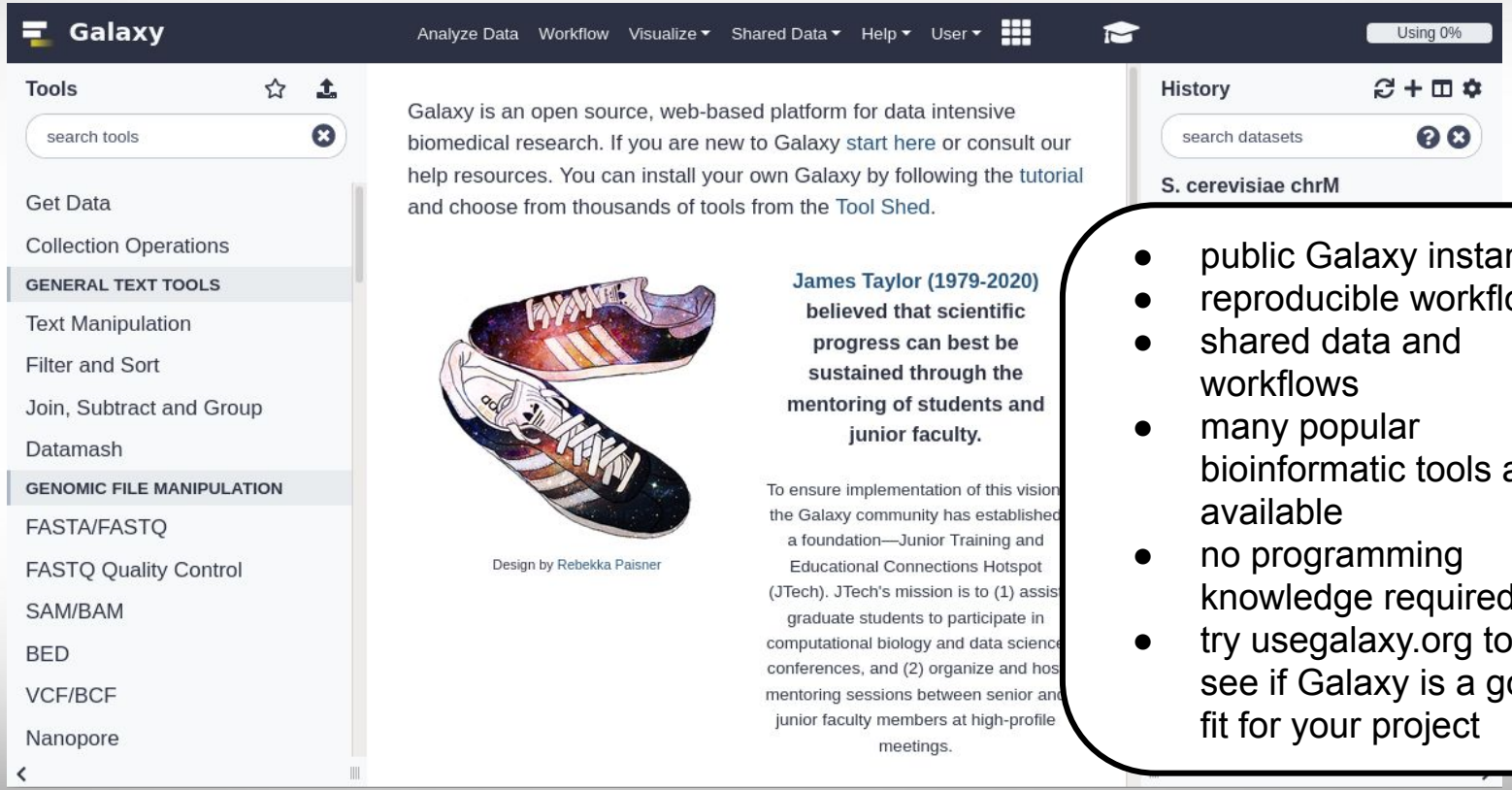
Website:	hprc.tamu.edu
Email:	help@hprc.tamu.edu
Telephone:	(979) 845-0219
Visit us in person:	Henderson Hall, Room 114A

Help us, help you -- we need more info

- Which Cluster
- UserID/NetID
- Job id(s) if any
- Location of your jobfile, input/output files
- Application used if any
- Module(s) loaded if any
- Error messages
- Steps you have taken, so we can reproduce the problem

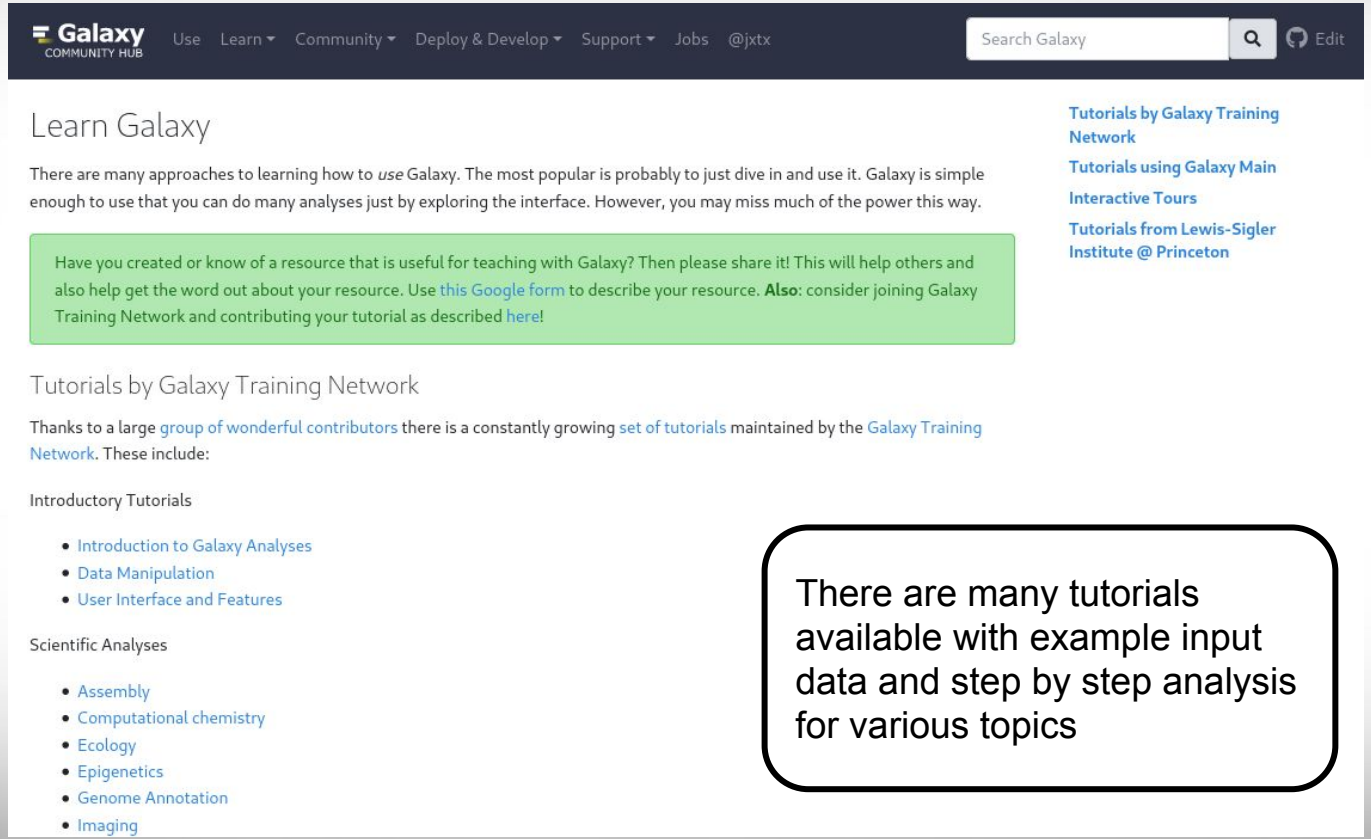
What is the Galaxy Project?

usegalaxy.org



- public Galaxy instance
- reproducible workflow
- shared data and workflows
- many popular bioinformatic tools are available
- no programming knowledge required
- try usegalaxy.org to see if Galaxy is a good fit for your project

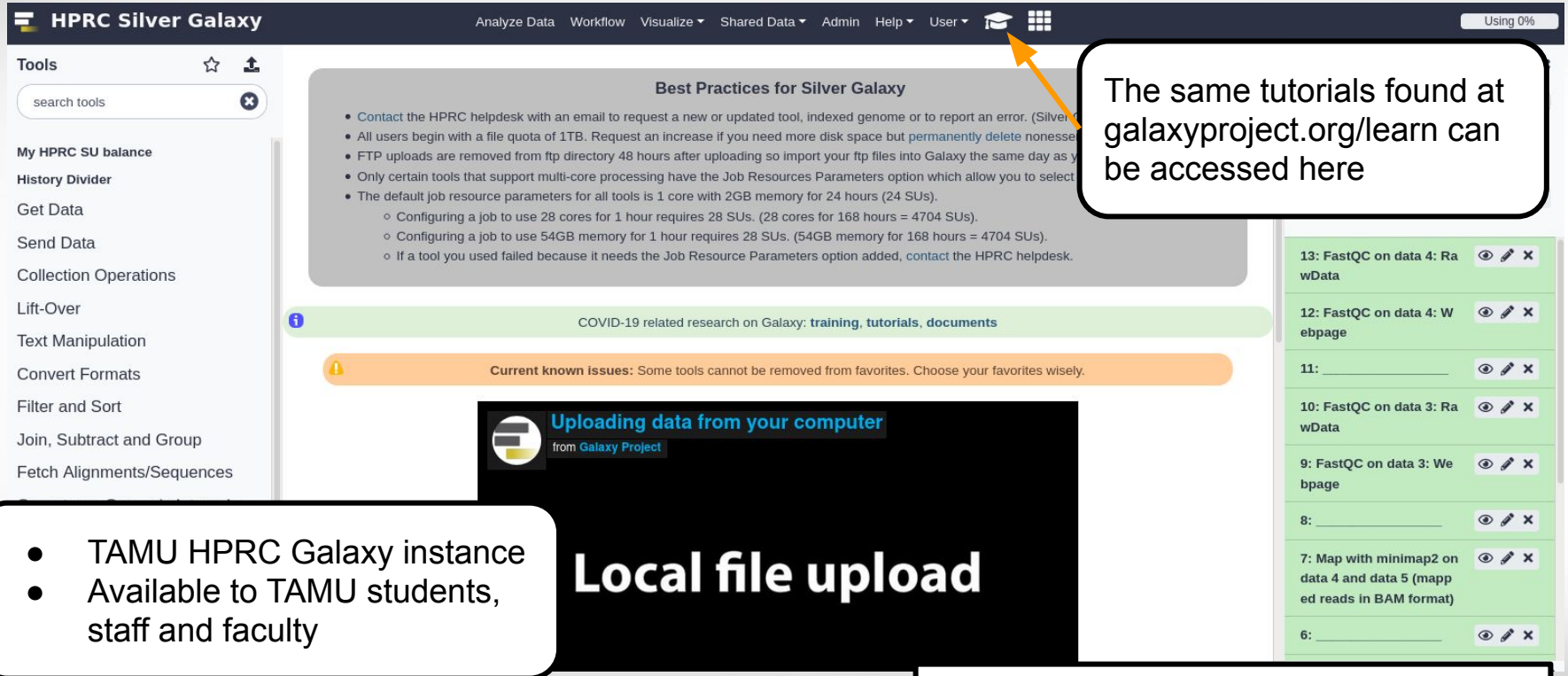
Galaxy 101 https://galaxyproject.org/learn



There are many tutorials available with example input data and step by step analysis for various topics

HPRC Galaxy
Maroon, Silver, Kaiser

HPRC Silver Galaxy



The same tutorials found at galaxyproject.org/learn can be accessed here

- TAMU HPRC Galaxy instance
- Available to TAMU students, staff and faculty

https://galaxy-terra.hprc.tamu.edu/silver

HPRC Galaxy Notes

Galaxy

Contents [hide]

HPRC Galaxy Tutorial

Silver Galaxy slides

Maroon Galaxy slides

Silver Galaxy

Before you request an account on Silver Galaxy, you must do the following:

- Go to usegalaxy.org and get familiar with Galaxy. You can start with a free account and learn about Galaxy tools.
- If you decide that Galaxy is a good choice for your research project then do the following
  - Apply for an HPRC account. See the NewUser page for details on how to request an account.
  - After your HPRC account is approved, send an email to help@hprc.tamu.edu requesting an account on Silver Galaxy
  - Send us information on what type of data you will be analyzing and which tools you expect to use for your research project.

If you are off campus then you will have to install and run the TAMU VPN to connect to Silver Galaxy.

Silver Galaxy can be accessed using your favorite web browser such as Firefox, Chrome or IE.

Account Security

Do not share your Galaxy account with anyone. Galaxy uses the TAMU Central Authentication Service which is linked to your TAMU account.

Make sure you always logout of Galaxy by selecting User -> Logout and then click the Logout button on the next screen and then close your browser when you are finished u

Permanently Delete unwanted files

In order to free disk space, you should permanently delete files that you have already deleted from the history.

Please make this part of your Galaxy work routine in order to free up disk space.

This is a two step process.

1. Click the X on the history item you want deleted.

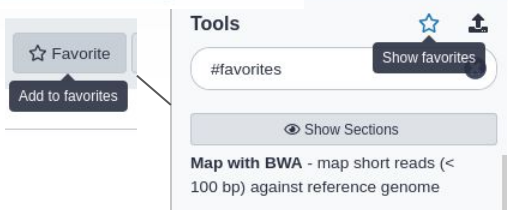
2. At the top of the history panel, you will see a link to deleted files. Click that link and th ant remo

hprc.tamu.edu/wiki/SW:Galaxy

New Galaxy Features

1. You can easily switch to an older software version

2. You can add tools to your favorites to easily locate later

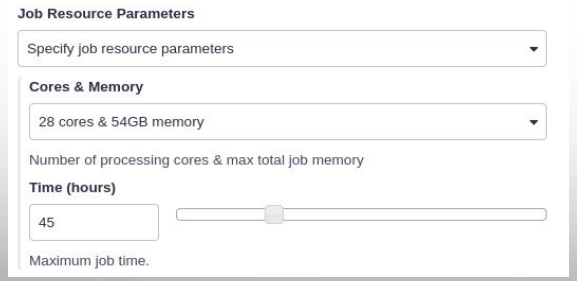


3. Disk quota enforced for all users
   a. all users begin with 1TB disk quota

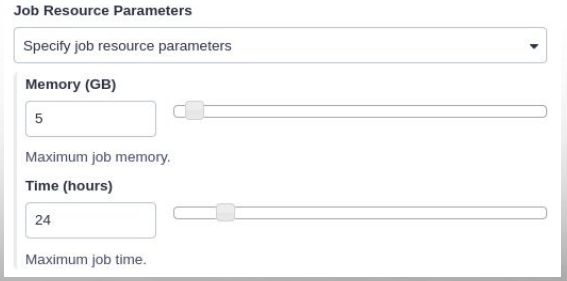
   b. request an increase if needed but permanently deleted nonessential files first
   c. compressed (gzipped) file format is supported
4. Tutorials are easy to follow along without leaving Silver Galaxy

5. Job Resource Parameters available for cores, memory and job time

tools supporting multi-core processing



tools supporting single-core processing

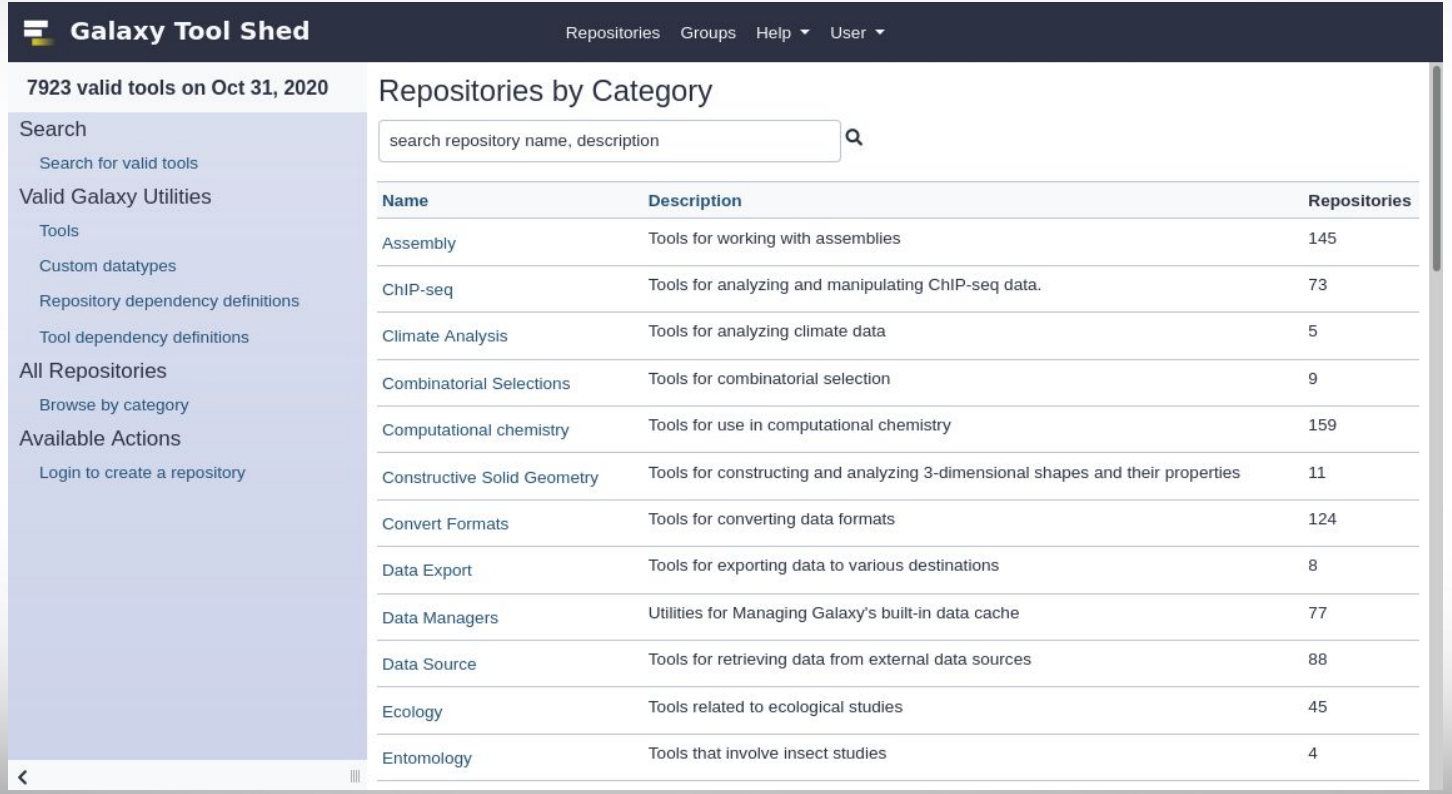


Silver Galaxy

- Try Galaxy at usegalaxy.org to see if it appropriate for your project
- Getting Access to HPRC Silver Galaxy
  - Available to Texas A&M students, staff and faculty with a NetID and an HPRC account
  - Apply for an HPRC account first
    - https://hprc.tamu.edu/apply
  - Then send an email request for a Silver Galaxy account
    - help@hprc.tamu.edu
  - Need to use VPN when connecting to Galaxy from off campus
  - Login to Silver Galaxy using your TAMU NetID and password
- Read the Galaxy Usage Notes
  - https://hprc.tamu.edu/wiki/SW:Galaxy
- There are no backups of users' Galaxy files

Look for Additionally Available Tools

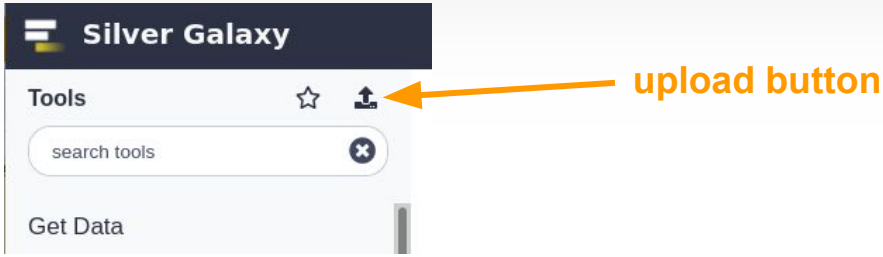
https://toolshed.g2.bx.psu.edu



Name	Description	Repositories
Assembly	Tools for working with assemblies	145
ChIP-seq	Tools for analyzing and manipulating ChIP-seq data.	73
Climate Analysis	Tools for analyzing climate data	5
Combinatorial Selections	Tools for combinatorial selection	9
Computational chemistry	Tools for use in computational chemistry	159
Constructive Solid Geometry	Tools for constructing and analyzing 3-dimensional shapes and their properties	11
Convert Formats	Tools for converting data formats	124
Data Export	Tools for exporting data to various destinations	8
Data Managers	Utilities for Managing Galaxy's built-in data cache	77
Data Source	Tools for retrieving data from external data sources	88
Ecology	Tools related to ecological studies	45
Entomology	Tools that involve insect studies	4

Uploading files to your Galaxy History

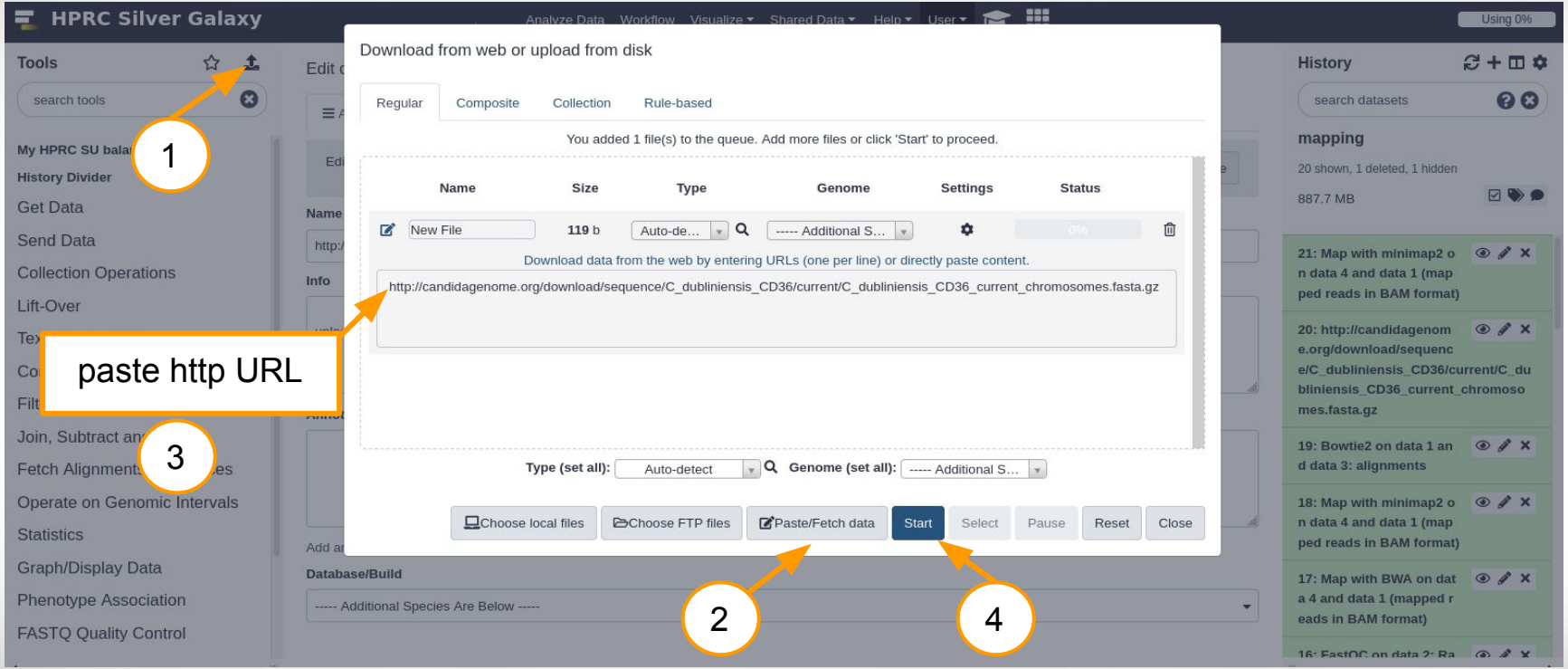
Get Data into Your Galaxy History



- Upload files < 2GB in size using Galaxy web interface
  - select local file on your computer to upload
  - paste URL address of any size file

- Can retrieve data from external websites directly into your Galaxy history with ‘Get Data’ tools
  - UCSC, BioMart, Ratmine, ...

HTTP URL Upload File < 2GB in size or Direct Paste



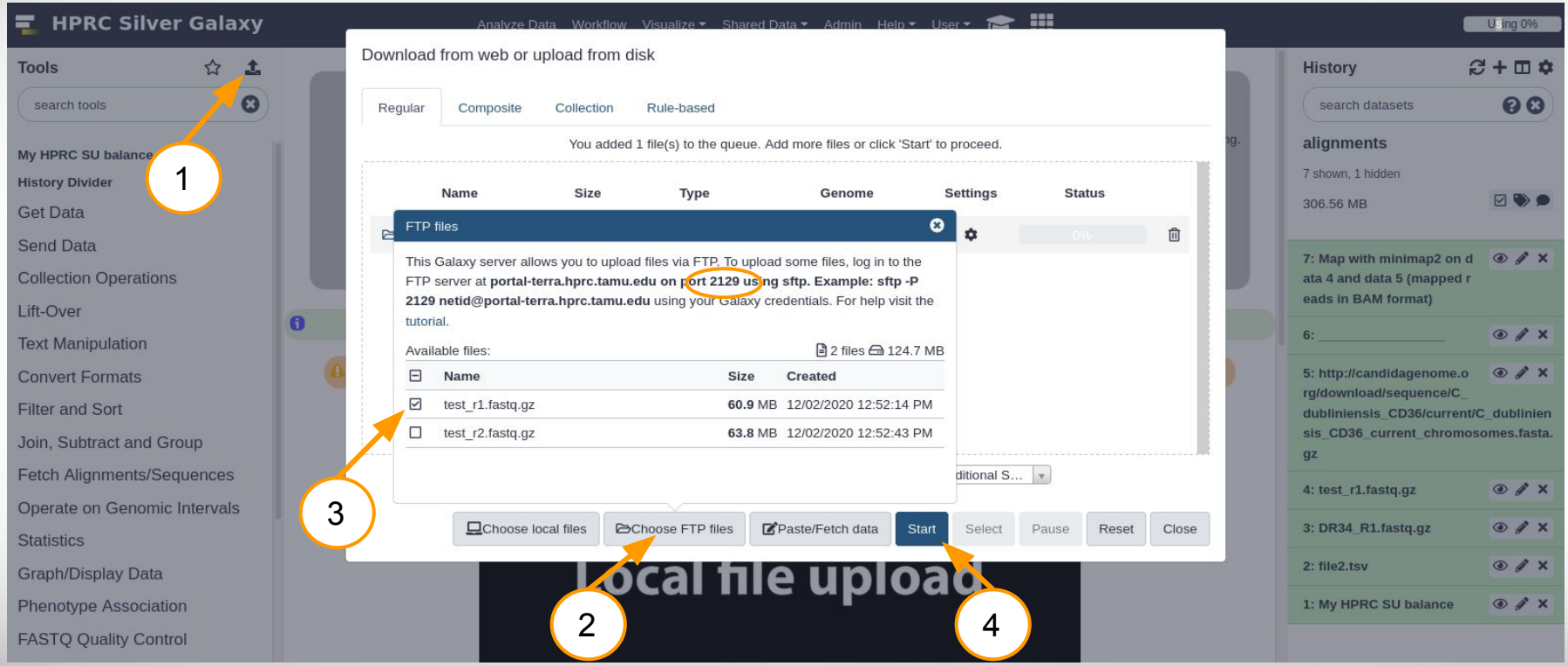
Uploading a 2GB+ size file to Silver Galaxy

- Files larger than 2GB should be copied using ftp instead of the upload button
- There are three options for uploading files via ftp
  a. Use the sftp command in a Unix terminal on your Mac or Linux desktop
  b. Copy files from Terra $SCRATCH directory to ftp directory using sftp on Terra
  c. Use sftp on BitVise on your Windows desktop to copy from your desktop to Terra

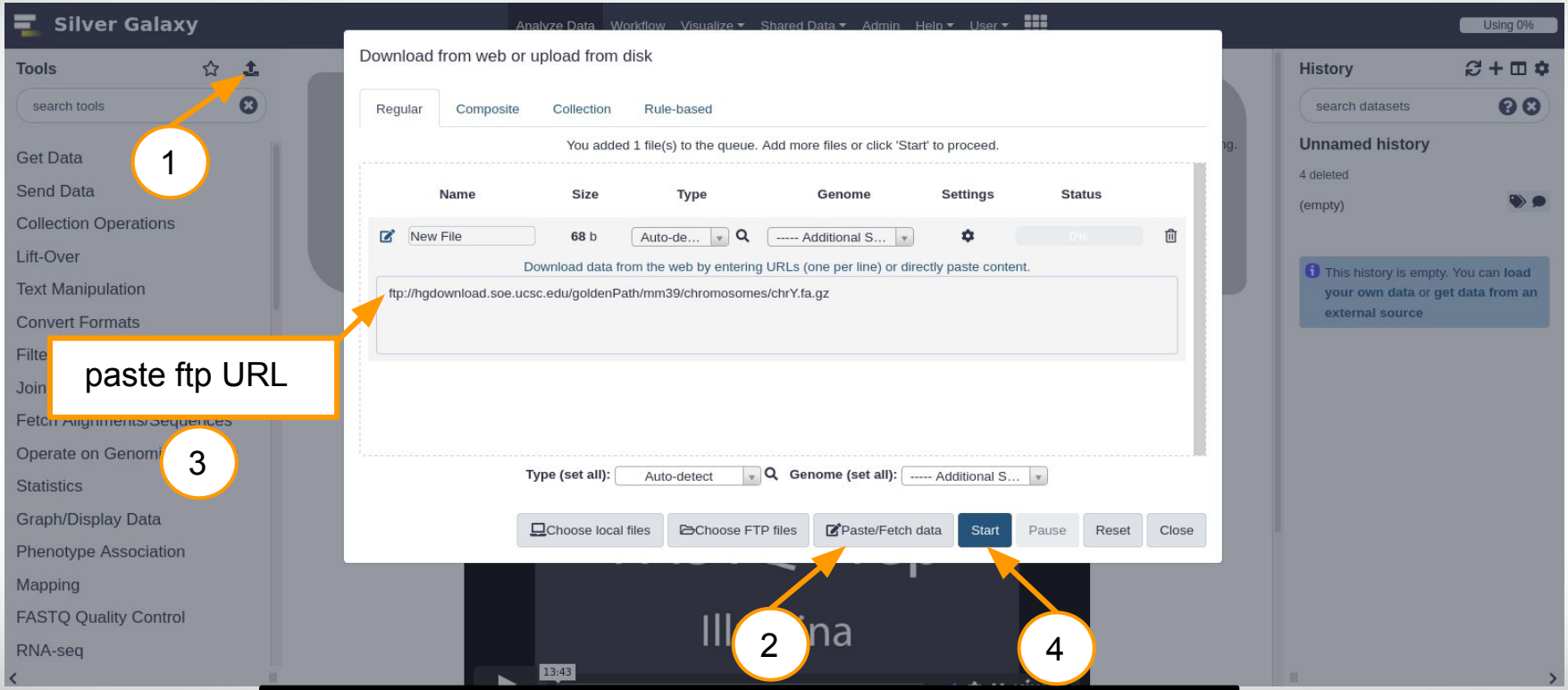
- After copying file to the Galaxy ftp directory, go to Galaxy 'upload file' interface in Galaxy to see your ftp transferred file (next slide)

https://hprc.tamu.edu/wiki/SW:Galaxy#Uploading_Files_.3E_2GB_via_FTP_to_Maroon_Galaxy

Add Your FTP Uploaded 2GB+ size File to Your History

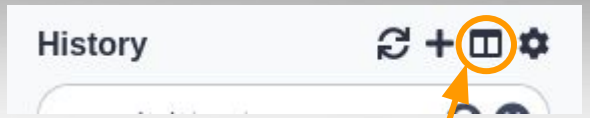


FTP Upload File 2GB+ size directly to History via URL



You can URL upload a 2GB+ sized file if the URL is an ftp site

Copy a file between histories



view all histories

History disk usage

Total disk usage. Mouse over the 'Using' icon to see total

Drag and drop files from one history to another to copy a file without duplicating the original file and increasing your disk usage

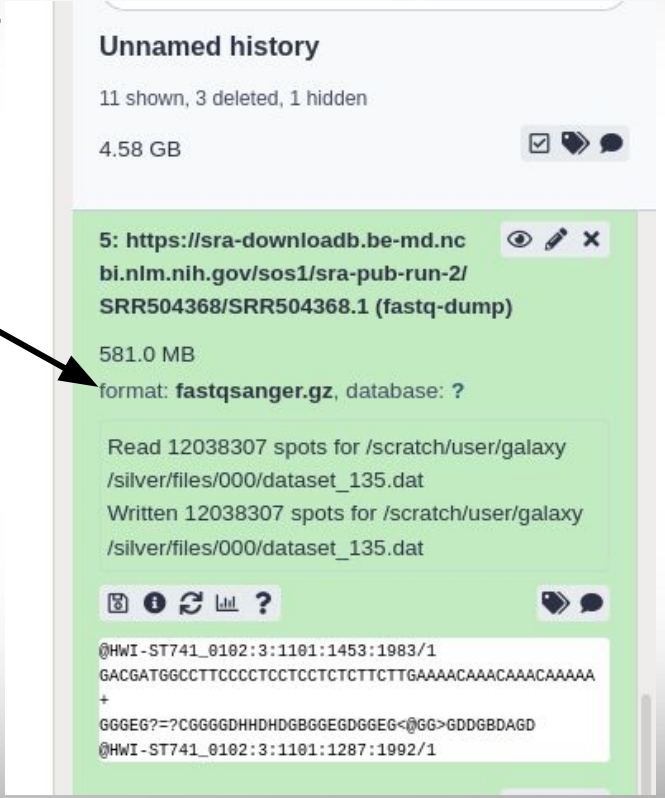
The History disk usage reflects an increase based on the file size but your overall disk usage does not increase

File and History Management

Automatic detection of fastqsanger format for fastq Uploads

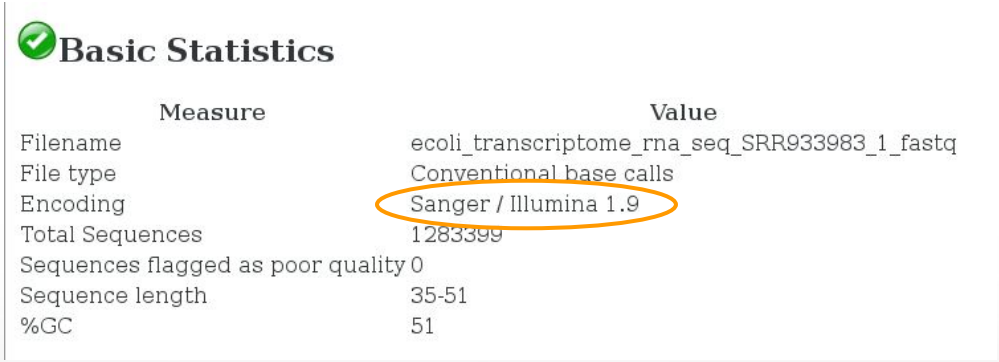
Galaxy will automatically detect fastqsanger format and set format attribute accordingly.

There is no need to run the Fastq Groomer tool on this file.



Galaxy File Formats

- Many Galaxy tools require fastqsanger not just fastq
- Galaxy now **auto detects** and **assigns** fastqsanger format
- FastQC tool will check fastq format of a fastq file
  - MiSeq, HiSeq, NextSeq, NovaSeq all use 1.8+



Basic Statistics

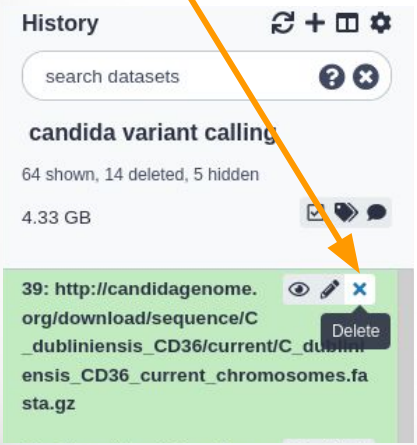
Measure	Value
Filename	ecoli_transcriptome_rna_seq_SRR933983_1_fastq
File type	Conventional base calls
Encoding	Sanger / Illumina 1.9
Total Sequences	1283399
Sequences flagged as poor quality	0
Sequence length	35-51
%GC	51

```
@ERR504787.5.1 M00368:15:000000000-A0HKH:1:2:16161:12630-1 length=100
TATTTTAAGTGACCAAGGAATGACTCCCCAATCATGGCTGTATCAACTCCAAAATTTTCTGCAACAGTCGCTGAAATATCTGCAAAATGCCCTTGTGGAA
+ERR504787.5.1 M00368:15:000000000-A0HKH:1:2:16161:12630-1 length=100
CCCFFFFFHHHHHJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJIJJJHIJHHGFFFFFFFEEEEEEDDDDDDDDDCDDDD
```

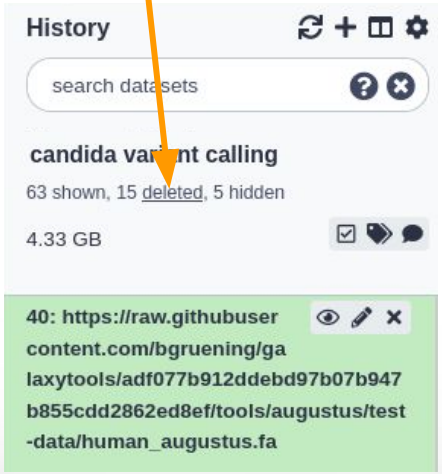
Permanently Delete Nonessential Files

deletes files from disk and reduces your Galaxy disk usage

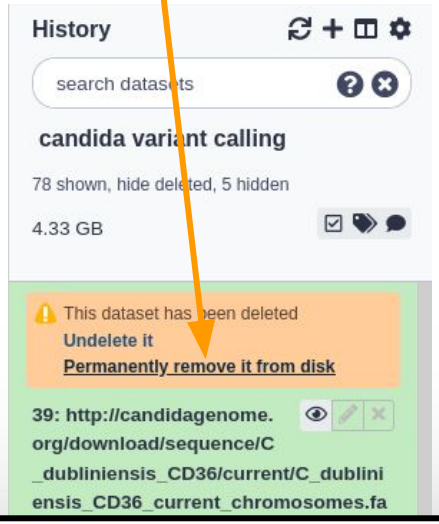
delete file



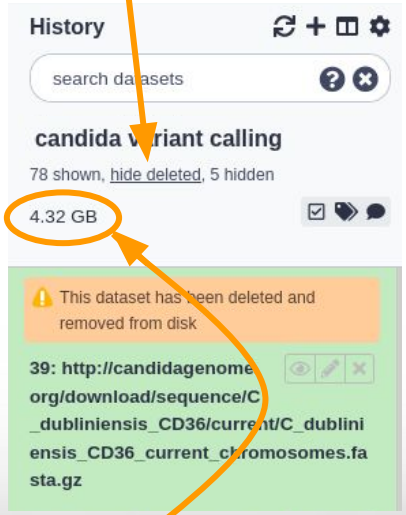
show deleted files



permanently delete file

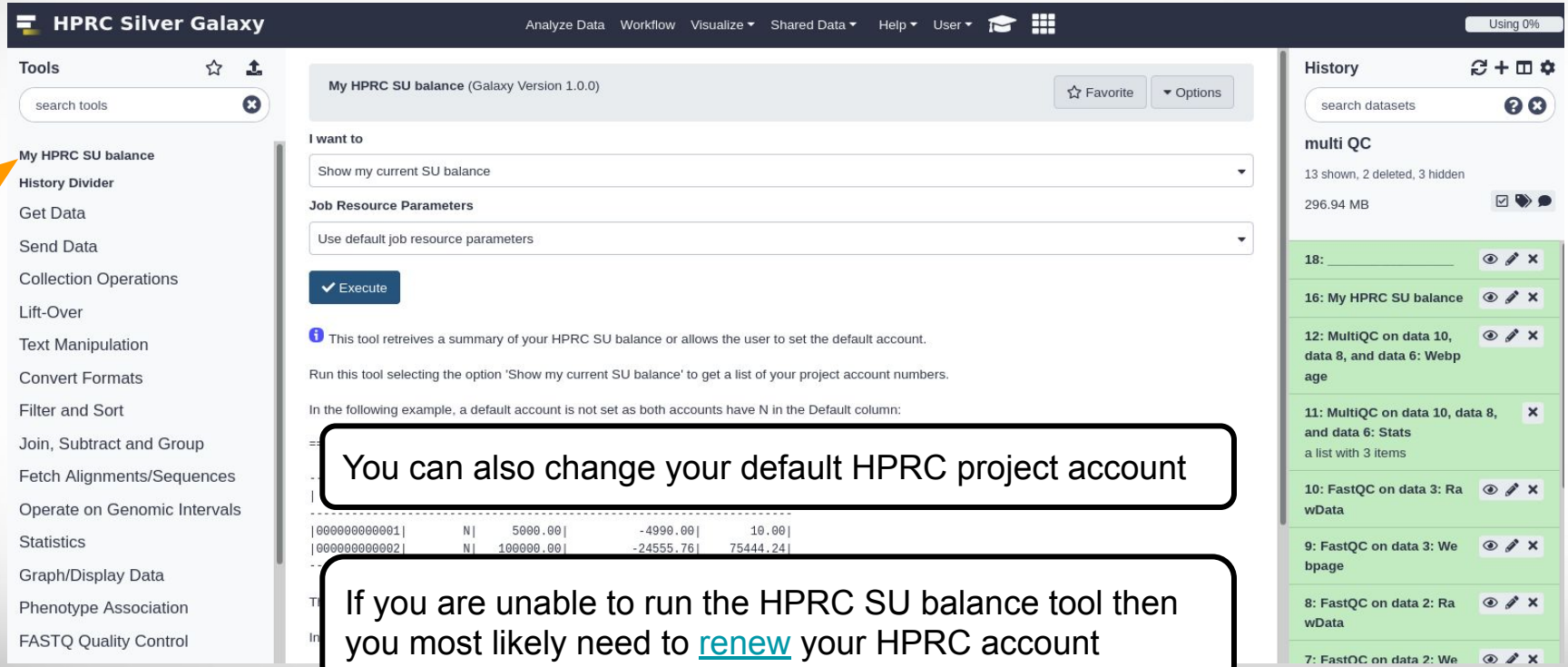


hide deleted files



Notice how History disk usage went down only after permanently deleting file

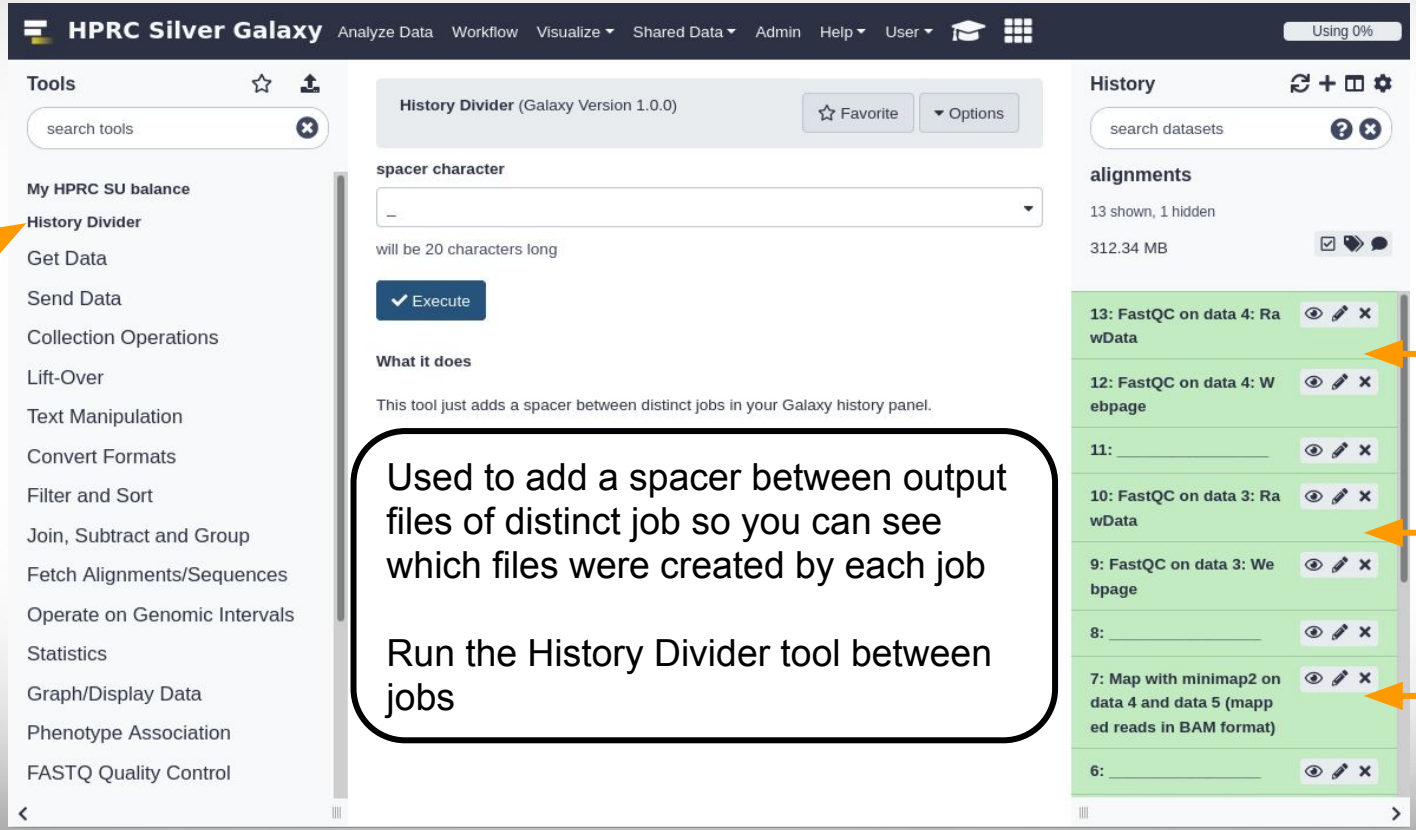
Check Your HPRC SUs Balance



You can also change your default HPRC project account

If you are unable to run the HPRC SU balance tool then you most likely need to renew your HPRC account

History Divider



Used to add a spacer between output files of distinct job so you can see which files were created by each job

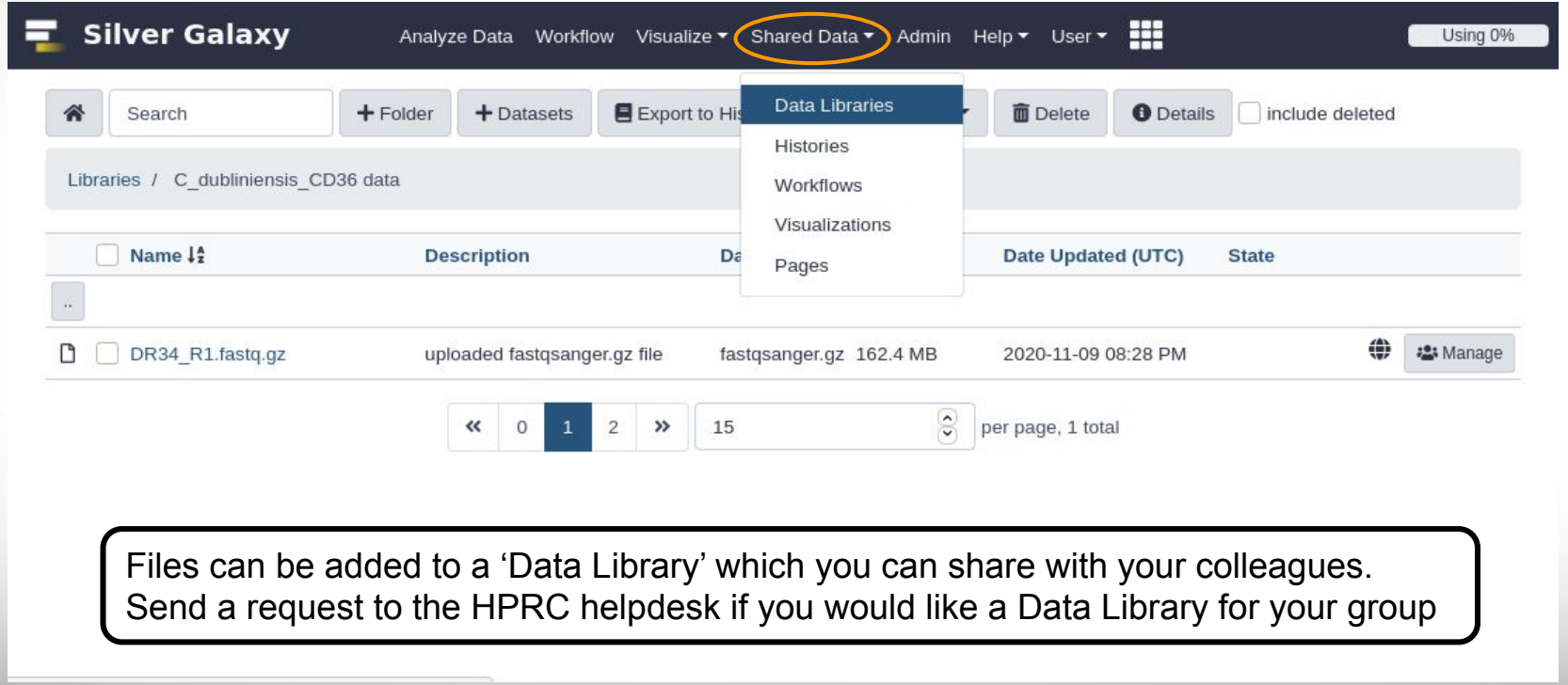
Run the History Divider tool between jobs

job with 2 output files

job with 2 output files

job with 1 output file

Shared Data Libraries

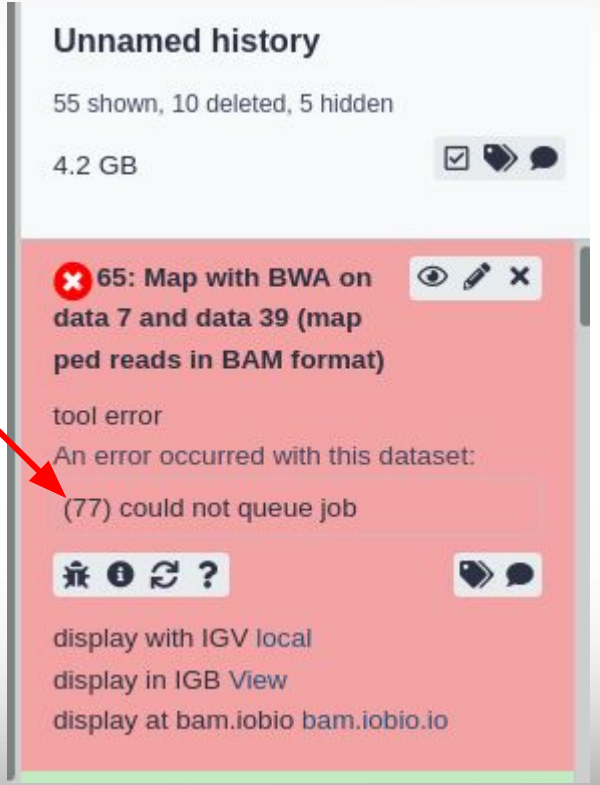


Files can be added to a ‘Data Library’ which you can share with your colleagues. Send a request to the HPRC helpdesk if you would like a Data Library for your group

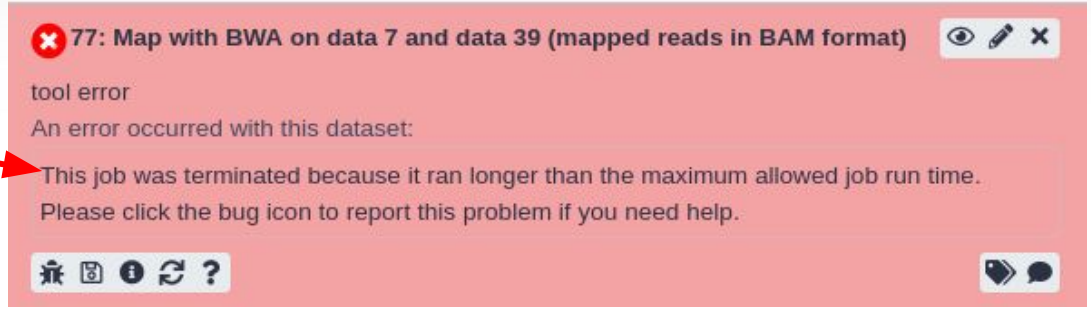
Debugging Failed Jobs

Failed to Queue Job

- Make sure you have enough SUs to run the job
  - My HPRC SU Balance
- Make sure your account has been renewed for current fiscal year

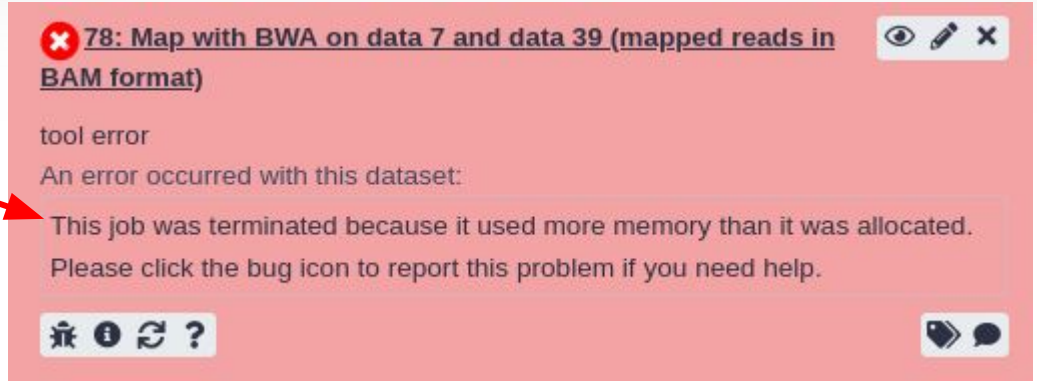


Failed Job: walltime limit reached

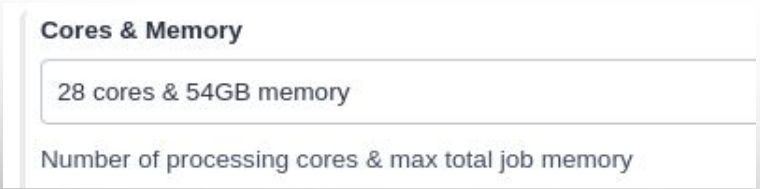


- Configure the job to use more Time

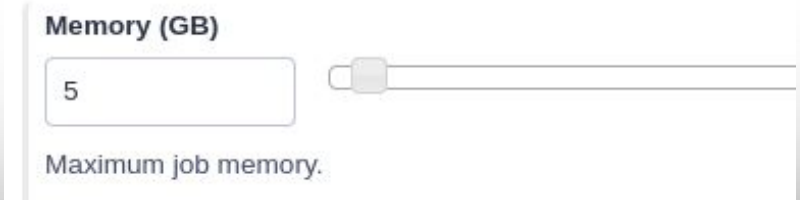
Failed Job: memory limit reached



- Configure the job to use more Memory



or

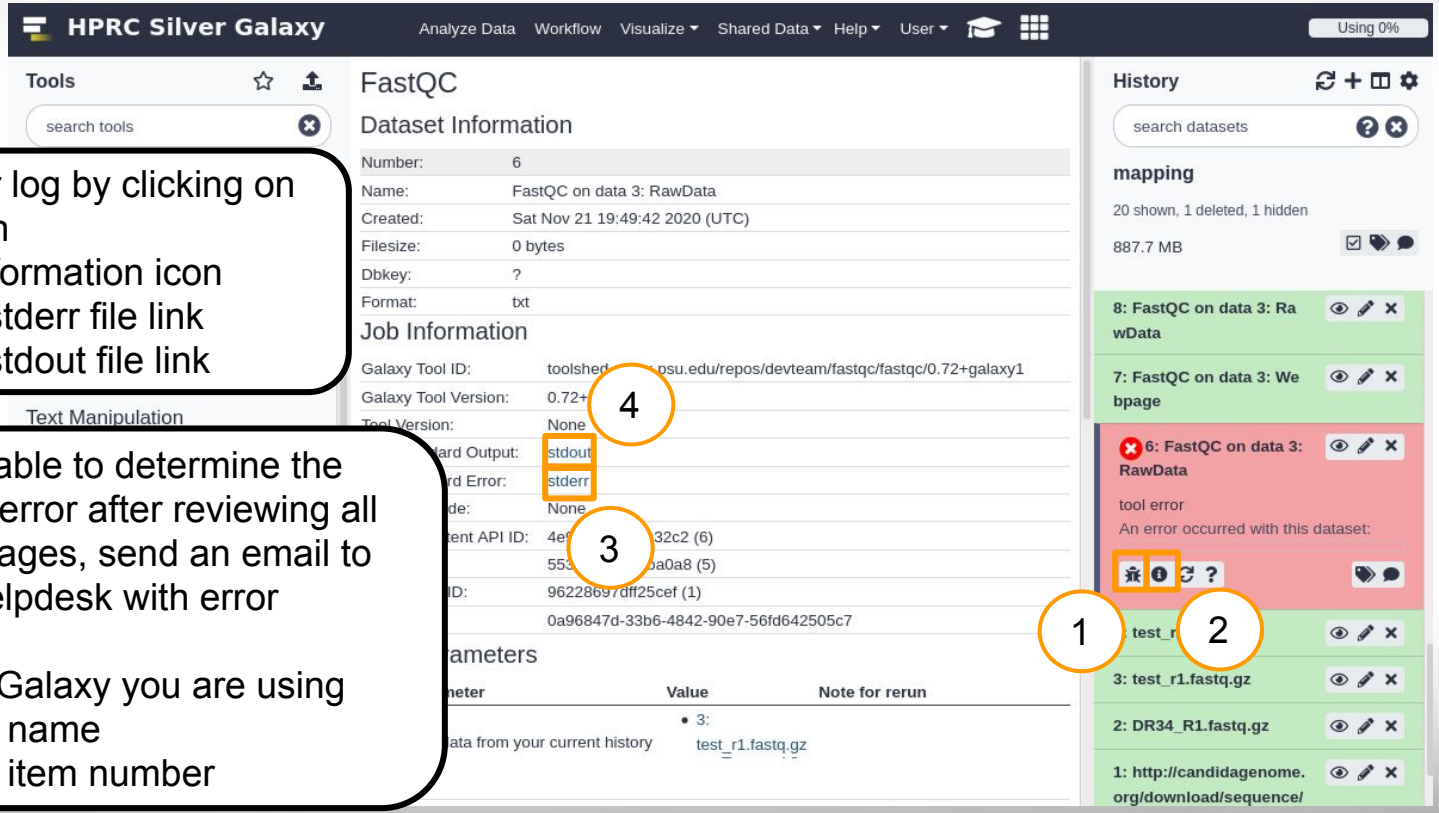


See Logs of Failed Jobs

1. Check error log by clicking on the bug icon
2. Click the information icon
3. Check the stderr file link
4. Check the stdout file link

If you are unable to determine the cause of the error after reviewing all the log messages, send an email to the HPRC helpdesk with error information

- Which Galaxy you are using
- History name
- History item number



usegalaxy.org Exercise

1. Add fasta file for S. cerevisiae chromosome I from genome.ucsc.edu to your current history and set the “Database/Build” attribute to the proper genome assembly
2. Run the “Compute Sequence Length” tool to get length of chromosome I
3. Add fasta file for chromosome M from genome.ucsc.edu to your history
4. Concatenate the two files into one file
5. Run the “Compute Sequence Length” tool on the concatenated file
6. Permanently delete the file you created in step 2.

Thank you.

Any questions?