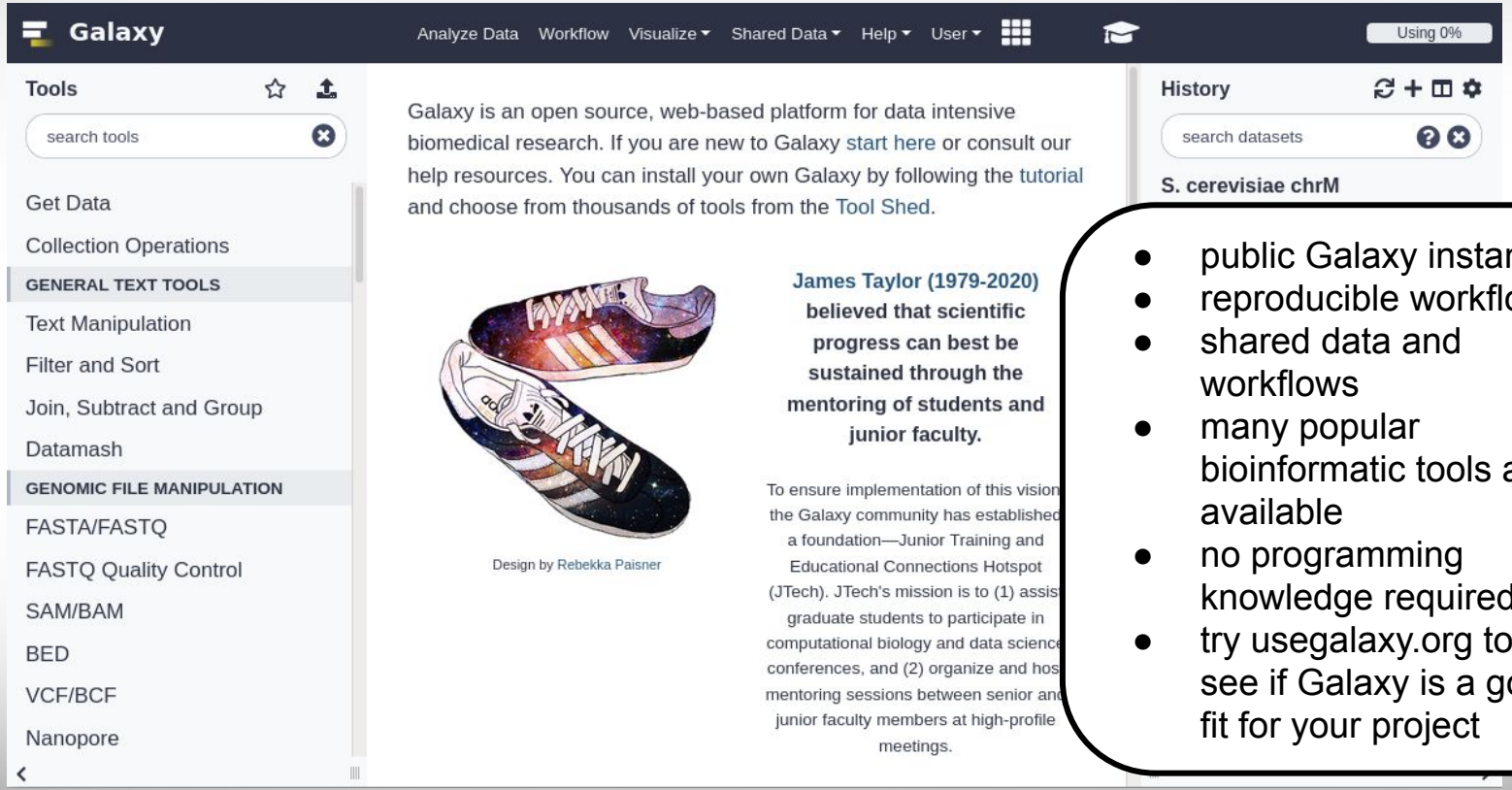
HPRC Maroon Galaxy

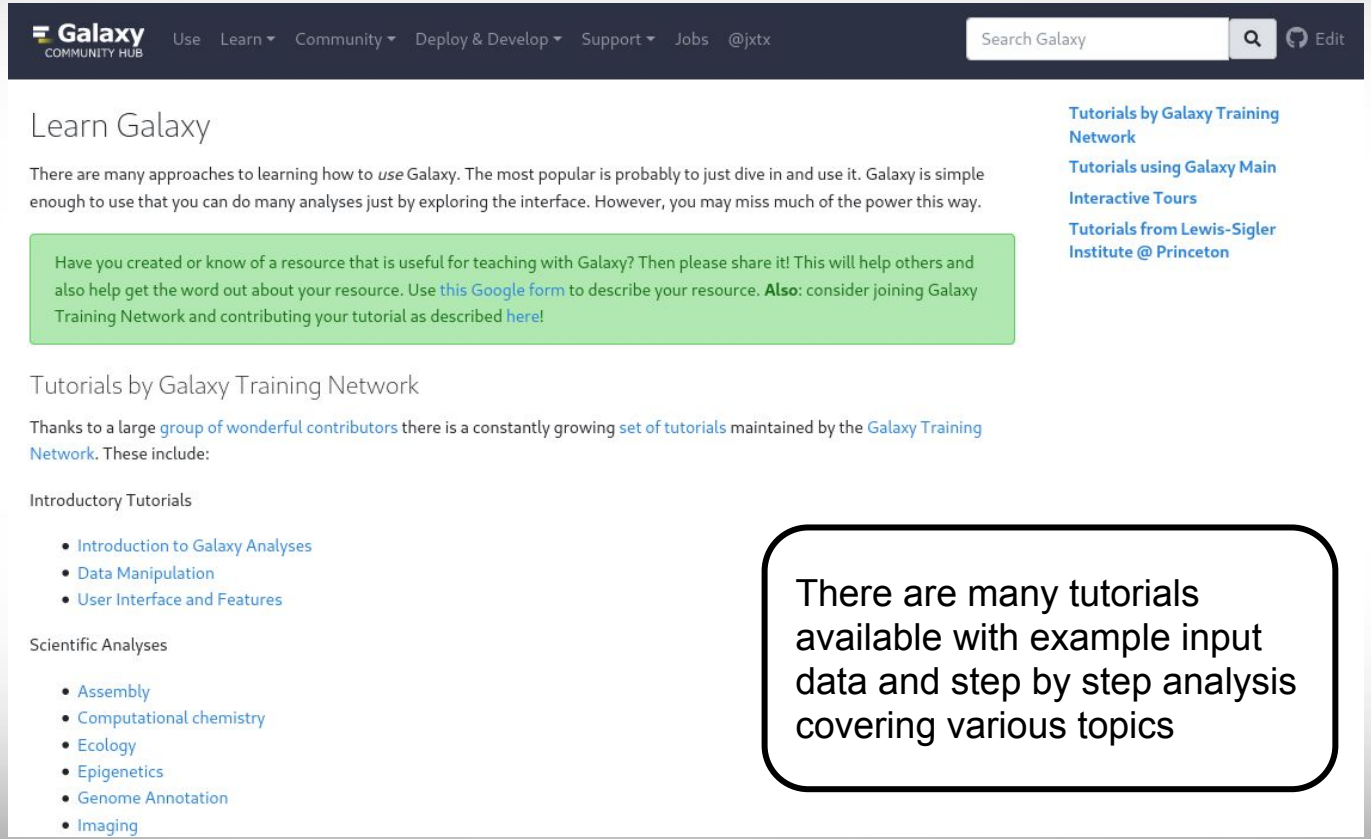
What is the Galaxy Project?

usegalaxy.org



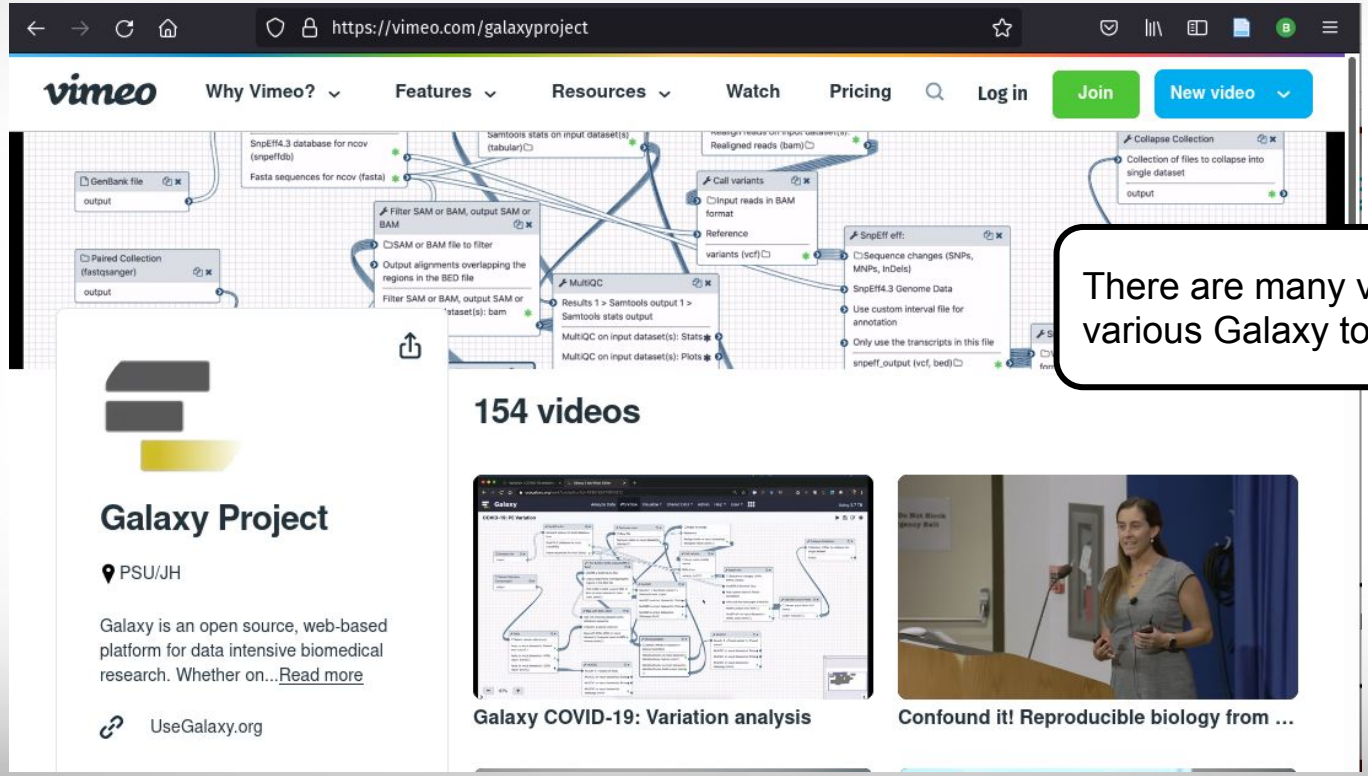
- public Galaxy instance
- reproducible workflow
- shared data and workflows
- many popular bioinformatic tools are available
- no programming knowledge required
- try usegalaxy.org to see if Galaxy is a good fit for your project

Galaxy 101 https://galaxyproject.org/learn



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

Galaxy Videos https://vimeo.com/galaxyproject

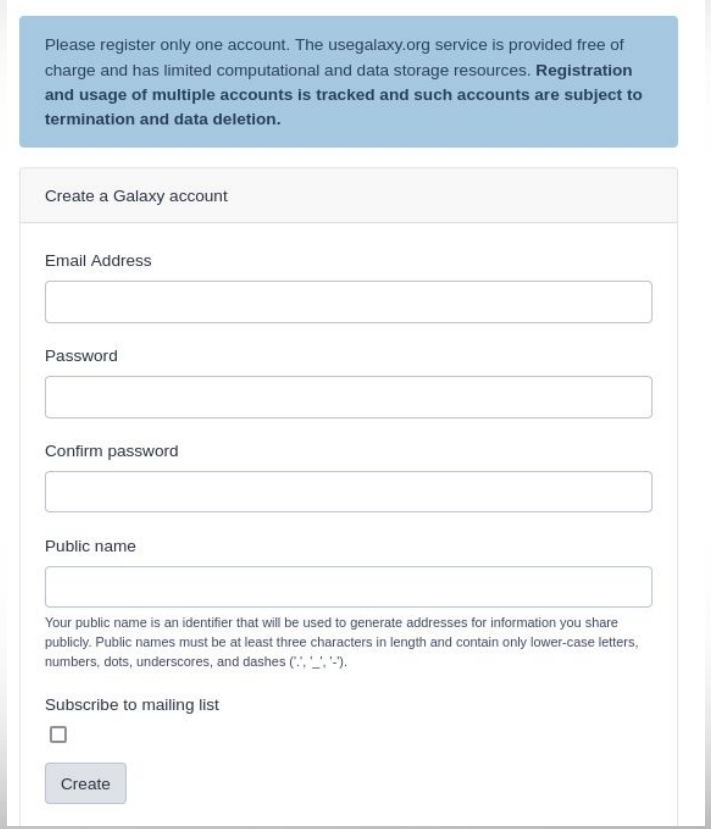


There are many videos available on various Galaxy topics

Apply for a usegalaxy.org account

usegalaxy.org/login

click 'Register here' then fill out the form, submit and check your email to validate your account



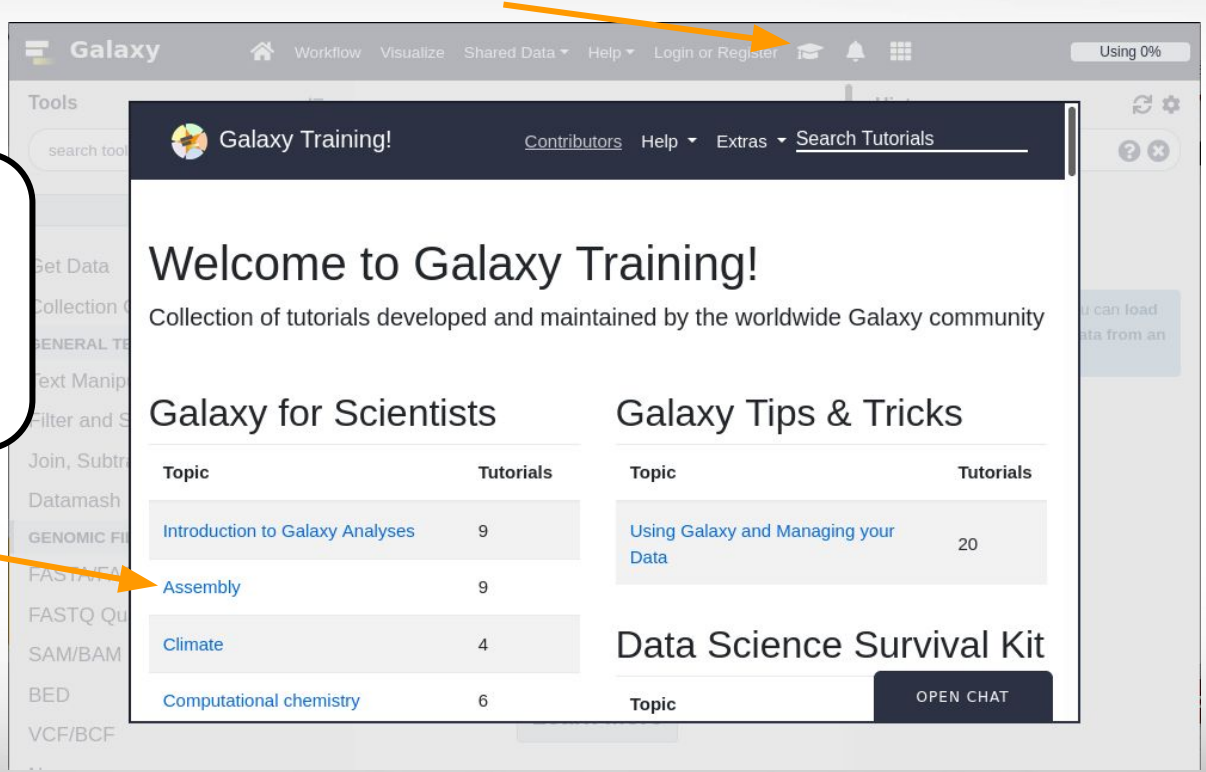
Galaxy Training Objectives

- use Galaxy Training tutorials
- upload and download data
- search and use tools
  - switch tool version
- view data
  - resize windows
- view/change file attributes
- debug jobs
- create new histories
- copy datasets between histories
- create workflow
- permanently delete files

Galaxy Training!

usegalaxy.org/login

Login to your account and click the graduation hat then click Assembly

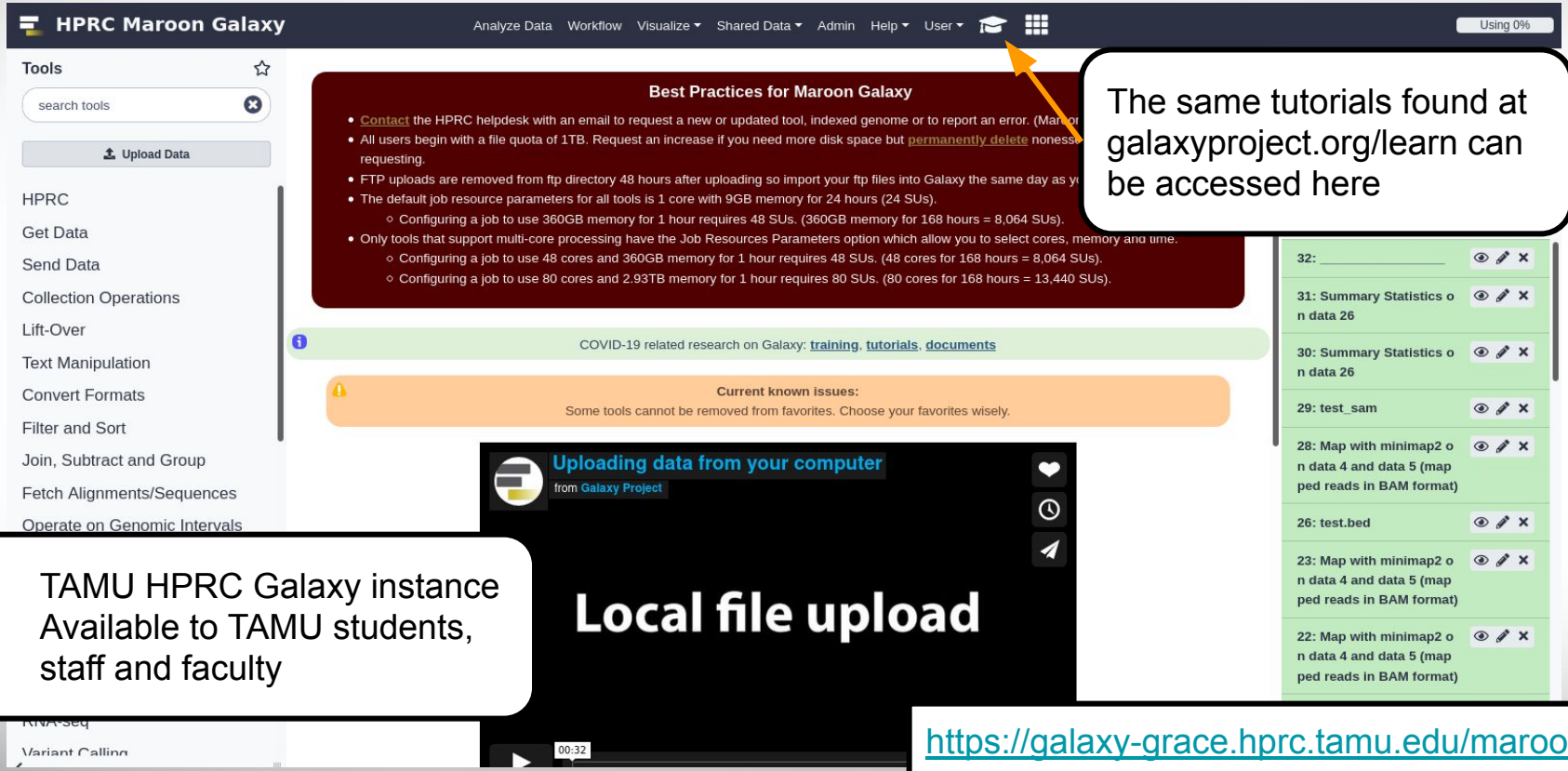


HPRC Maroon Galaxy

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

HPRC Maroon 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-grace.hprc.tamu.edu/maroon

HPRC Galaxy Notes

Galaxy

Contents [hide]

1 Galaxy
- 1.1 Maroon Galaxy Accounts
- 1.2 Silver Galaxy
- 1.3 Set Your Default Account
- 1.4 Account Security
- 1.5 Permanently Delete unwanted files
- 1.6 Uploading Files > 2GB via FTP to Maroon Galaxy
  - 1.6.1 From a Unix Computer (Mac or Linux)
  - 1.6.2 From one of your directories on Ada
  - 1.6.3 Using Bitvise
  - 1.6.4 Using MobaXterm
  - 1.6.5 Using Filezilla
  - 1.6.6 Using WinSCP
- 1.7 Downloading Files > 2GB via FTP from Maroon Galaxy to your desktop
  - 1.7.1 Stage files to copy
    - 1.7.1.1 Download using a terminal
    - 1.7.1.2 Download using a GUI client
  - 1.7.2 Delete files after copying
- 1.8 Requesting New Galaxy Tools
  - 1.8.1 When a Galaxy tool is Available
  - 1.8.2 When a tool has no Galaxy interface
- 1.9 FAQ
- 1.10 Tool specific notes
  - 1.10.1 Fastq groomer
  - 1.10.2 Trinity
    - 1.10.2.1 Before you run a Trinity job
    - 1.10.2.2 If your Trinity job Fails
  - 1.10.3 RSEM
  - 1.10.4 BLAST
  - 1.10.5 bwa, bowtie, bowtie2 hisat2
  - 1.10.6 HISAT2
- 1.11 Share your History
  - 1.11.1 with user(s) on the same Galaxy instance
- 1.12 Manage your HPRC Accounts

Maroon Galaxy Accounts

The new Maroon Galaxy (v21.01) on Grace is available to students, faculty and staff for research use.

See the Maroon Galaxy usage slides

Before you request an account on Maroon 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.
- Request a Grace Maroon Galaxy account only if you have data to analyze, otherwise use usegalaxy.org Galaxy for training and practice.
- If you decide that Galaxy is a good choice for your research project then do the following
  - Establish an HPRC account by sending a request. See the NewUser page for details on how to request an account.
  - After you have your HPRC account approved, send an email to help@hprc.tamu.edu requesting an account on Maroon 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 Maroon Galaxy.

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

https://galaxy-grace.hprc.tamu.edu/maroon

Set Your Default Account

Use the HPRC *My HPRC SU Balance* tool. If you are unable to use the tool, you are either out of SUs or need to renew your HPRC account.

You can also check your Grace SU balance at the Grace portal

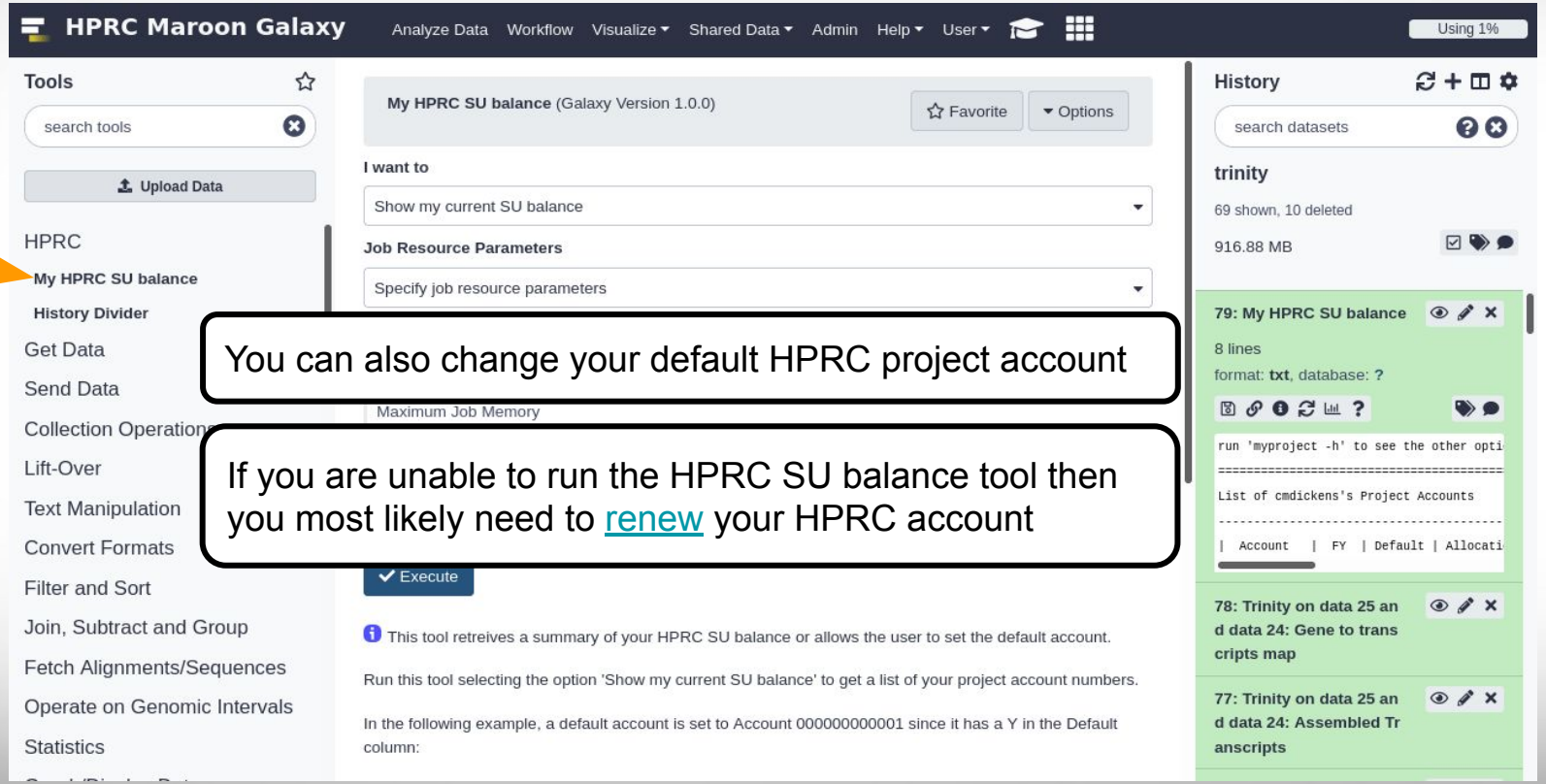
hprc.tamu.edu/wiki/SW:Galaxy

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 using Galaxy.

Check Your HPRC SUs Balance



You can also change your default HPRC project account

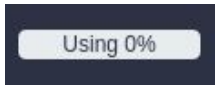
If you are unable to run the HPRC SU balance tool then you most likely need to renew your HPRC account

New Galaxy Features

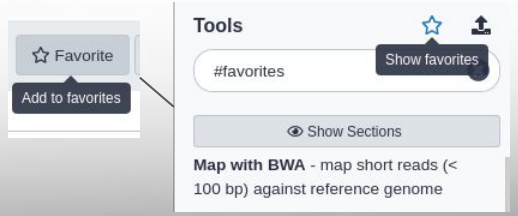
1. You can easily switch to an older software version

2. Tutorials are easy to follow along without leaving Maroon Galaxy

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 for uploads but not for all tools



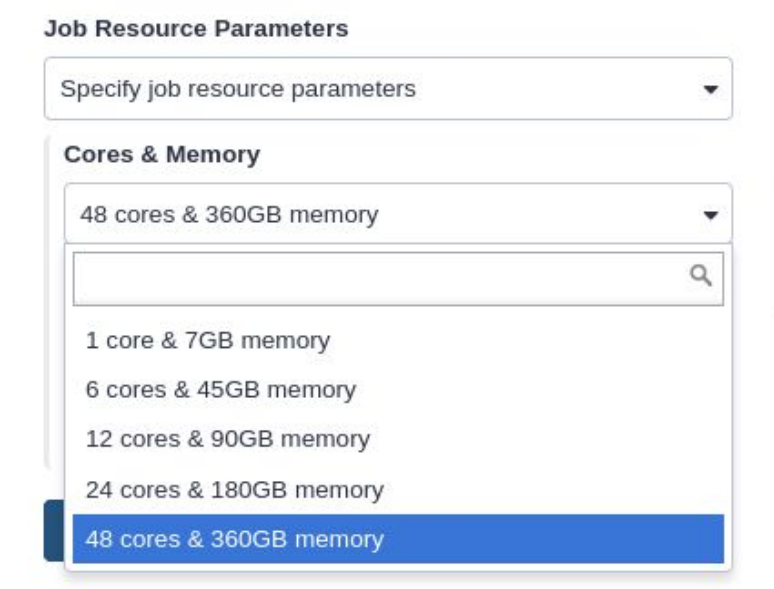
4. You can add tools to your favorites to easily locate later



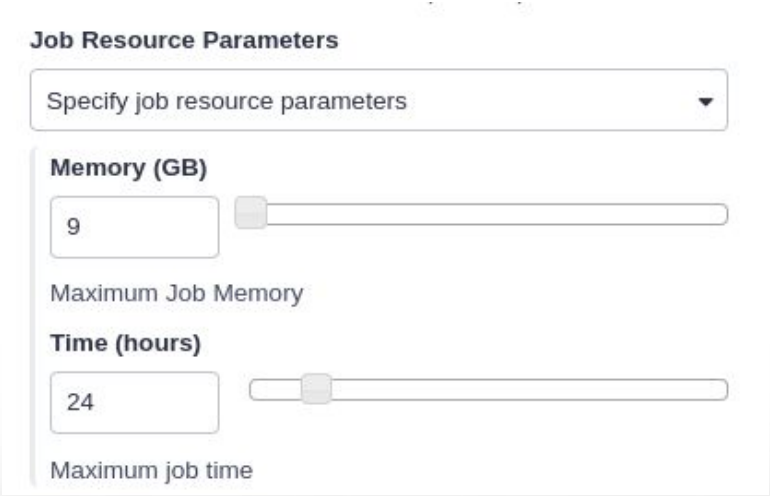
New Galaxy Features cont.

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

tools supporting **multi-core** processing



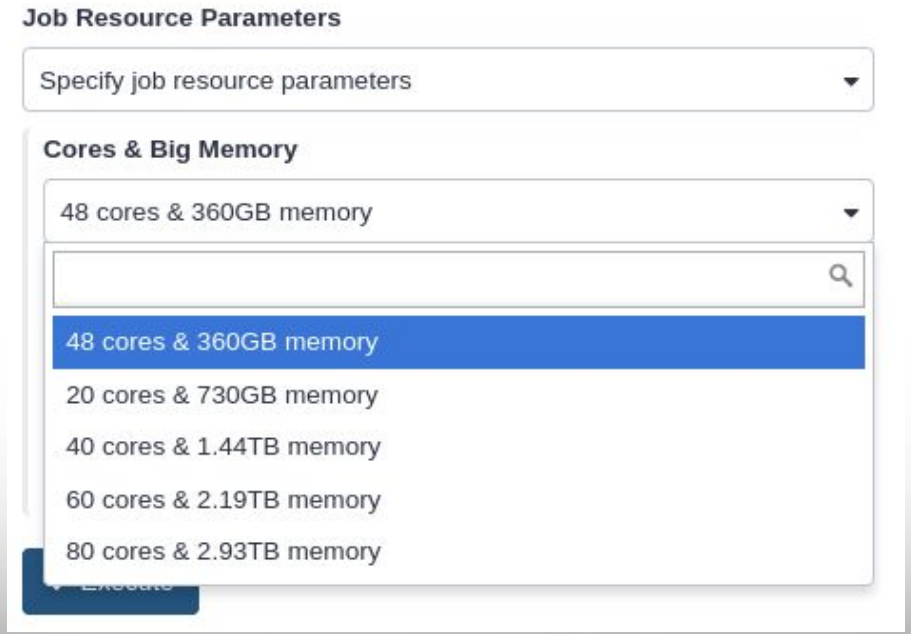
tools supporting **single-core** processing



New Galaxy Features cont.

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

tools supporting **multi-core** and requiring big memory processing

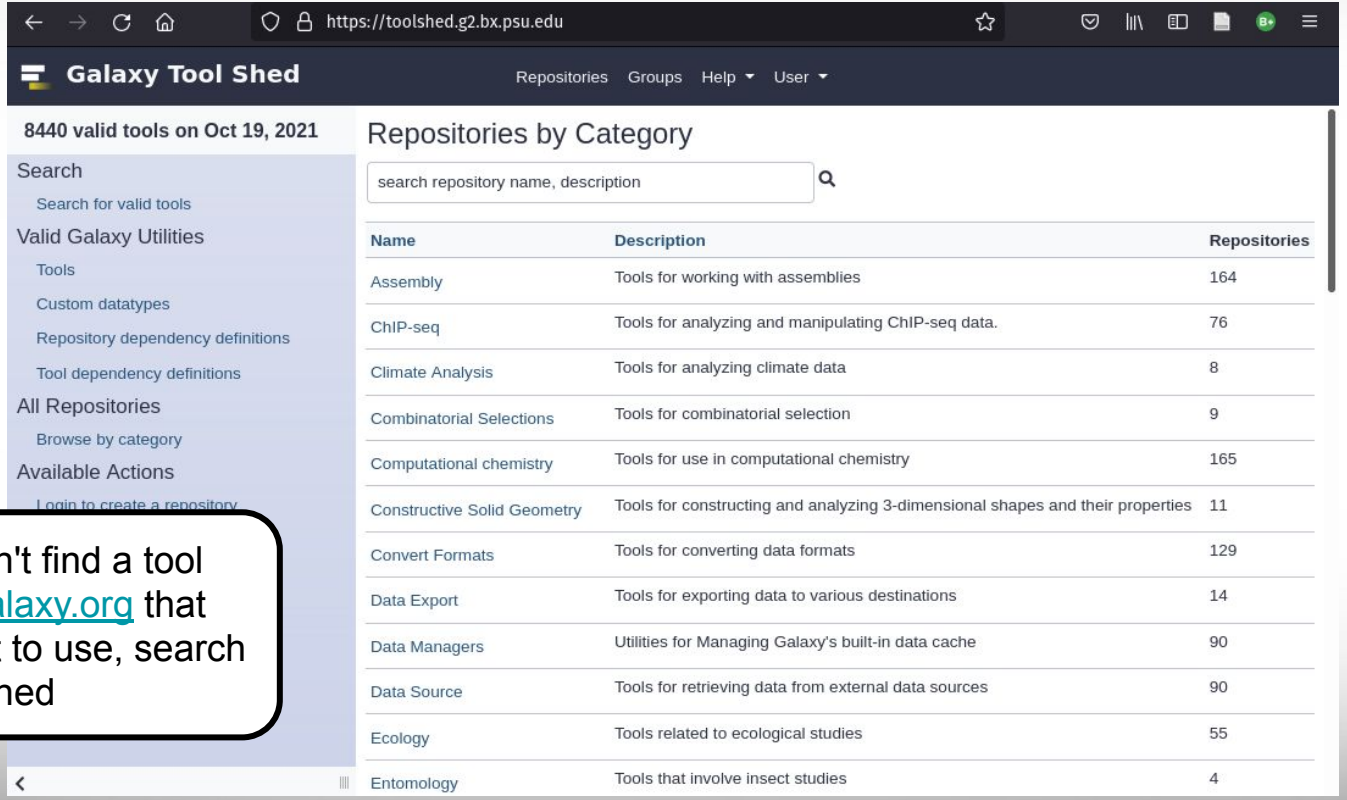


Maroon Galaxy

- Try Galaxy at usegalaxy.org to see if it appropriate for your project
- Getting Access to HPRC Maroon 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 Maroon Galaxy account
    - help@hprc.tamu.edu
  - Need to use VPN when connecting to Galaxy from off campus
  - Login to Maroon 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

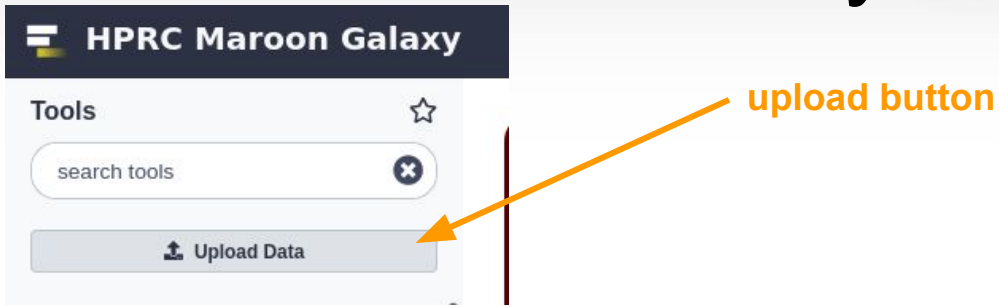
https://toolshed.g2.bx.psu.edu



If you can't find a tool on usegalaxy.org that you want to use, search the toolshed

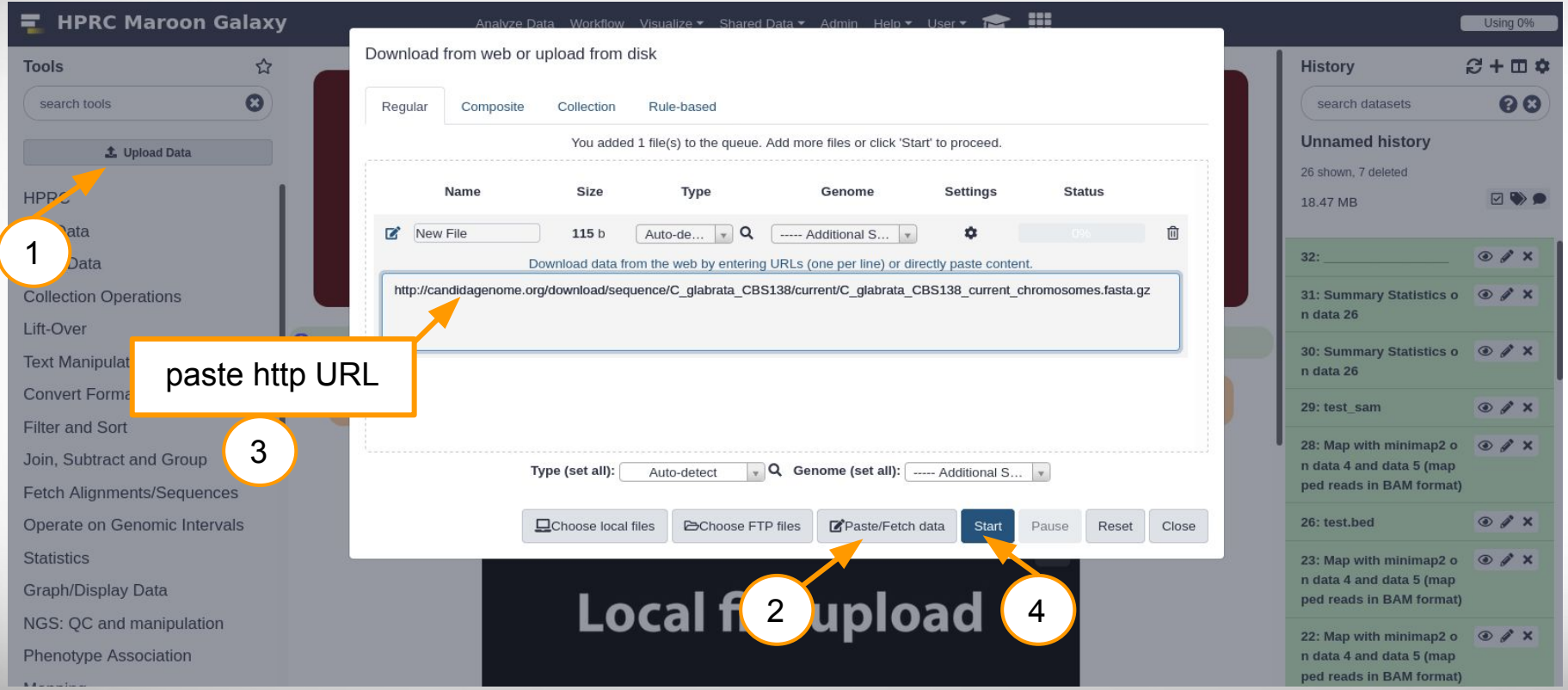
Uploading files to your Galaxy History

Get Data into Your Galaxy History



- Upload files < 2GB in size using Galaxy web interface
  - select local file on your computer to upload
  - or 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 Maroon Galaxy

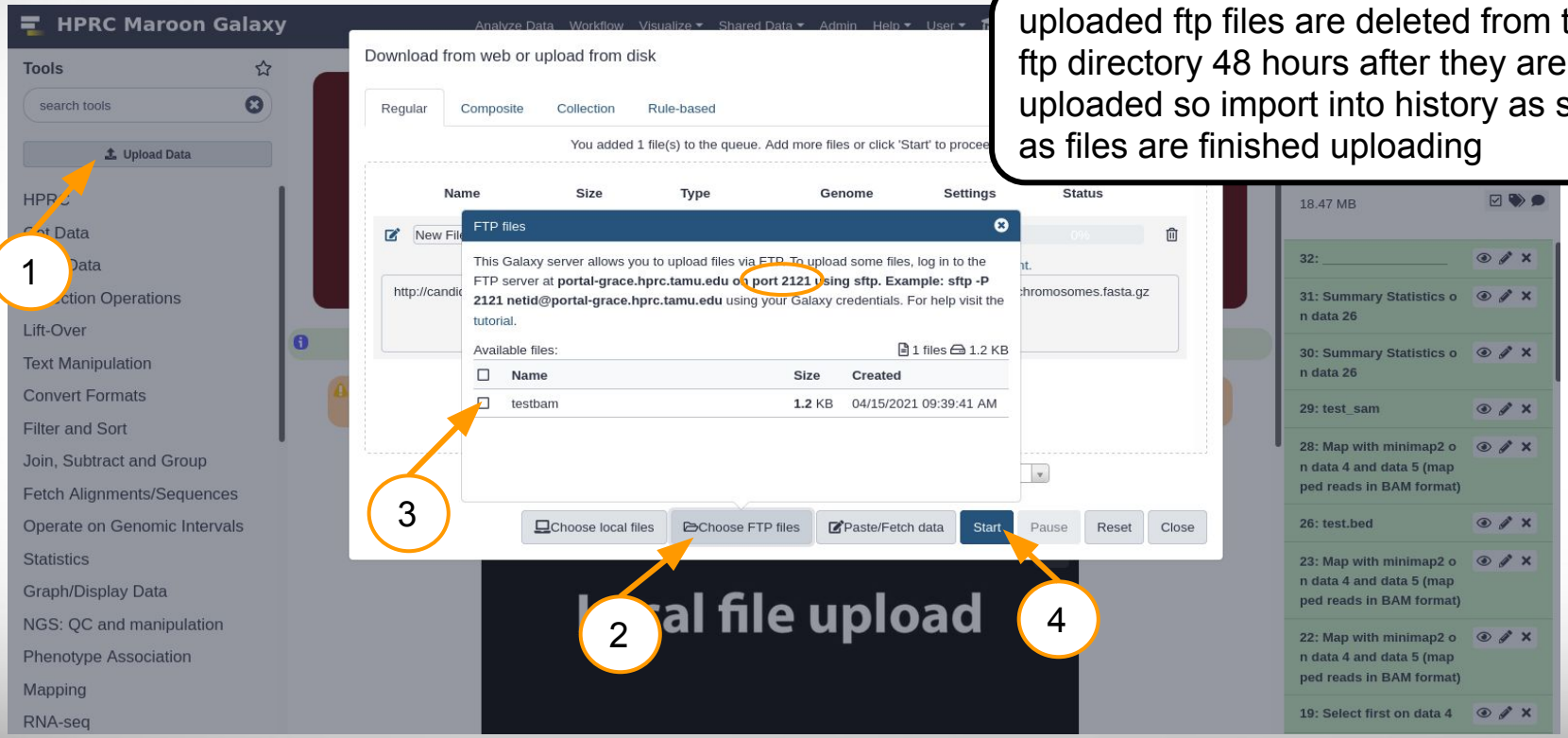
- Files larger than 2GB should be copied using ftp instead of the upload button which uses the http protocol which has limitations on file sizes for file transfers
- There are three options for uploading files via ftp
  a. Use the sftp command in a Unix terminal on your Mac or Linux desktop
  b. Use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

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

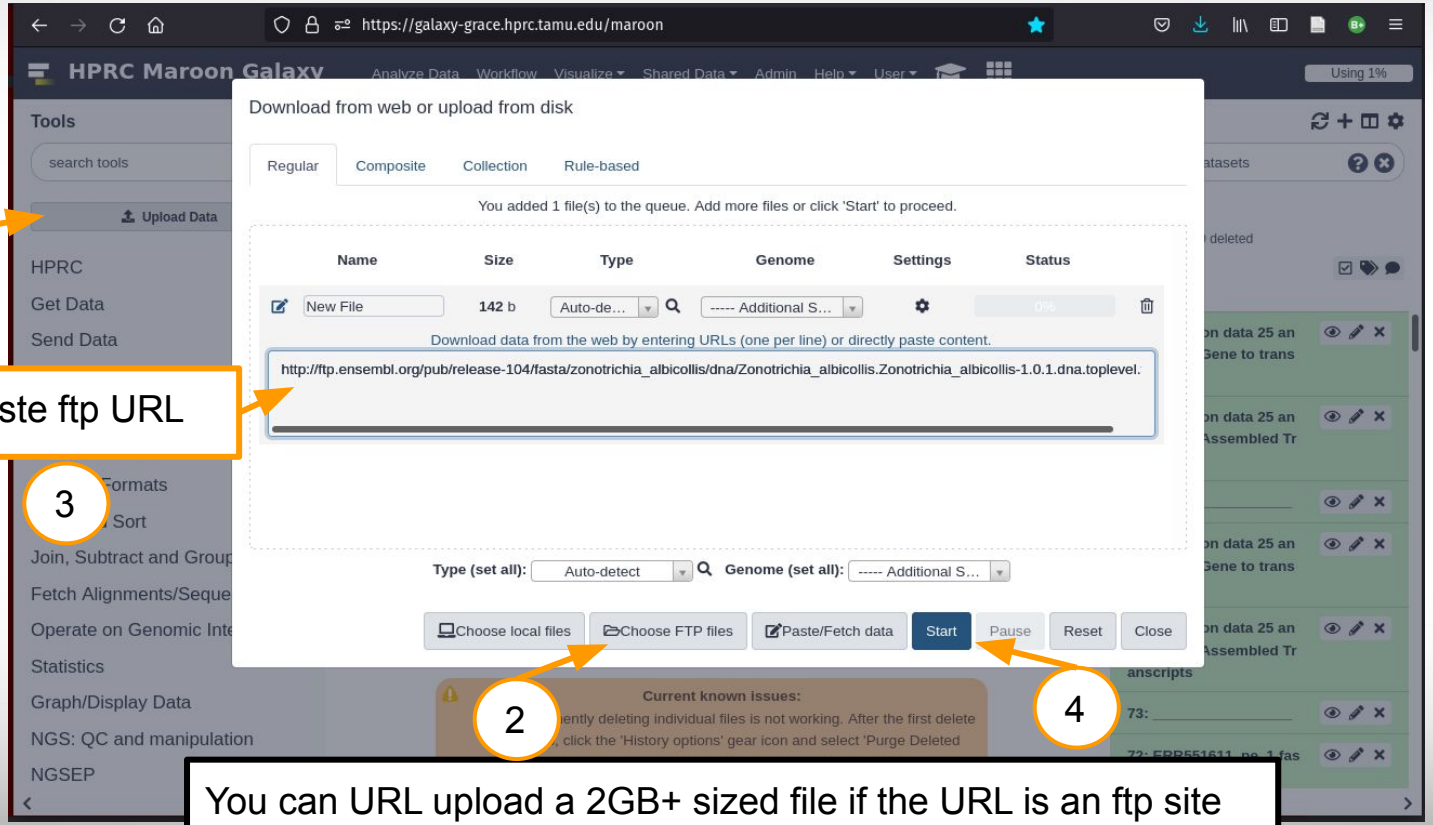
https://hprc.tamu.edu/wiki/SW:Galaxy#Uploading_Files_.3E_2GB_via_FTP_to_Maroon_Galaxy

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

uploaded ftp files are deleted from the ftp directory 48 hours after they are uploaded so import into history as soon as files are finished uploading

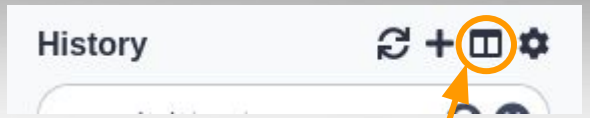


FTP Upload File 2GB+ size directly to History via URL



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

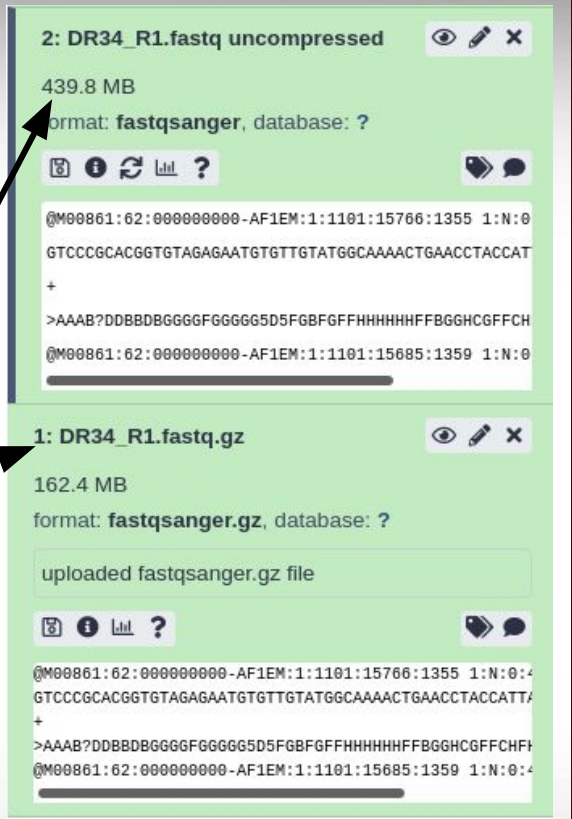
File and History Management

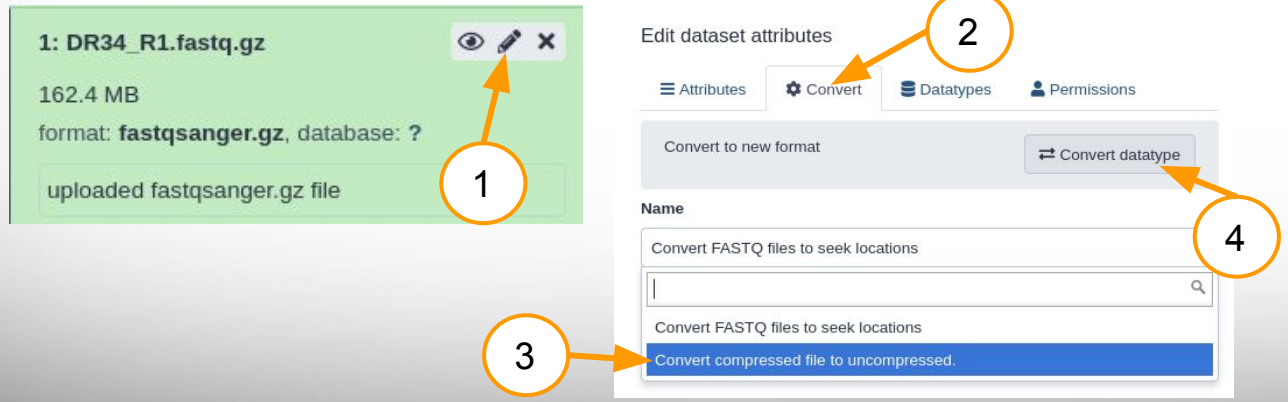
Converting and Using gzipped files

- Although Galaxy now supports uploading files in gzipped format, some Galaxy tools do not support gzipped format for input files.
- Files can be converted to uncompressed format but the original gzipped file should be ***permanently*** deleted or the resulting unzipped file should be ***permanently*** deleted after it is used in an analysis in order not to have duplicated data that takes up disk space

uncompressed file

original file compressed





Permanently delete one of the files after the analysis is complete in order to save disk space

Permanently Delete Nonessential Files

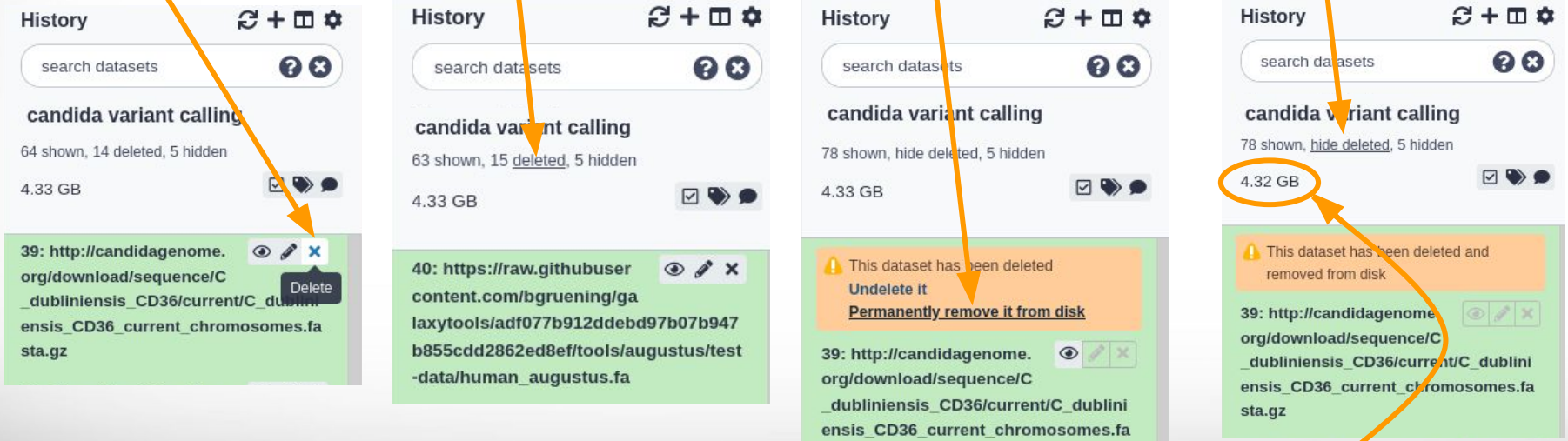
deletes files from disk and reduces your Galaxy disk usage

delete file

show deleted files

permanently delete file

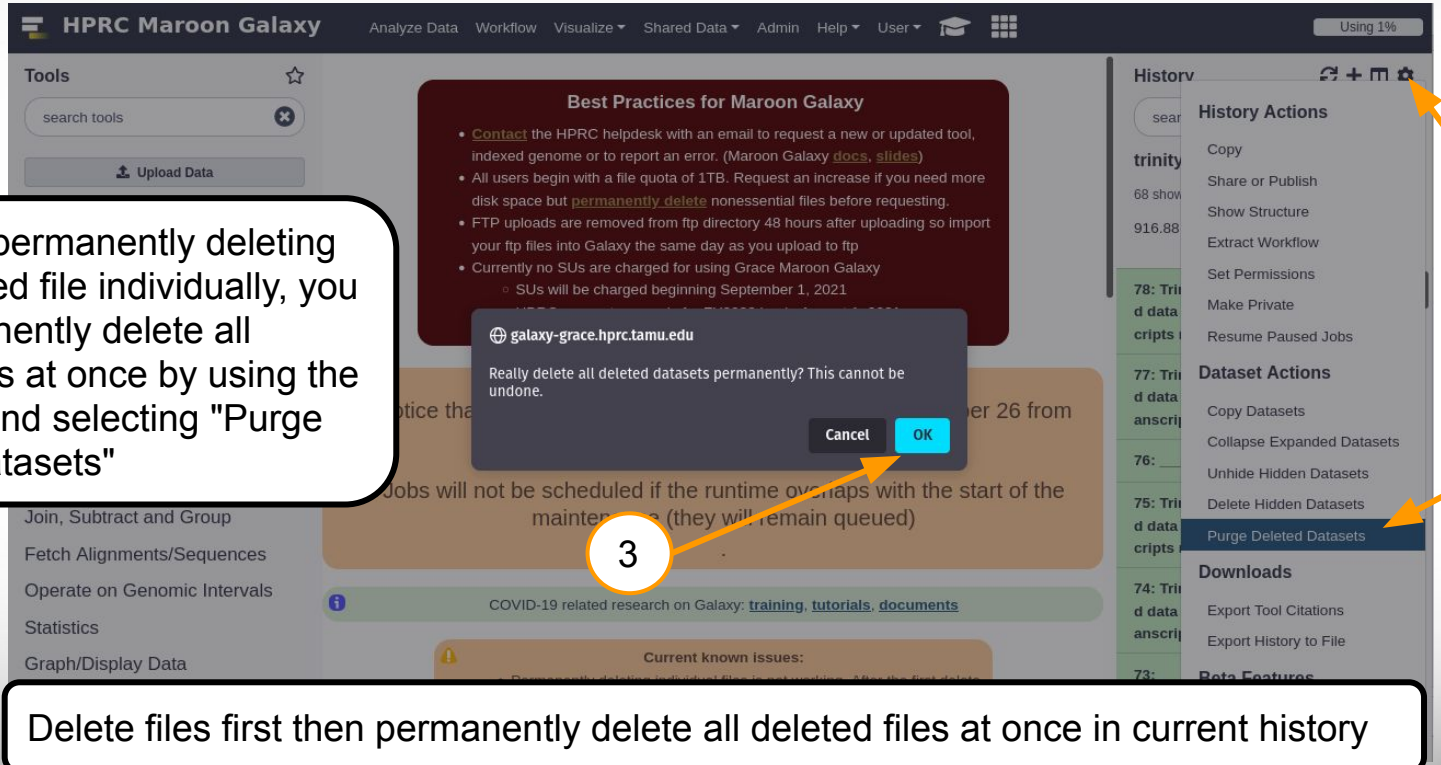
hide deleted files



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

Permanently Delete Nonessential Files

deletes files from disk and reduces your Galaxy disk usage



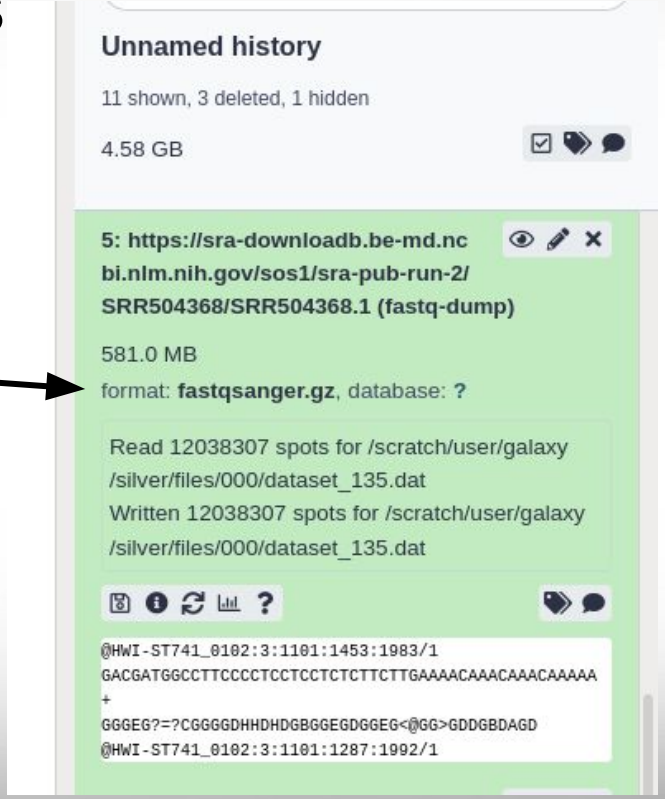
Instead of permanently deleting each deleted file individually, you can permanently delete all deleted files at once by using the gear icon and selecting "Purge Deleted Datasets"

Delete files first then permanently delete all deleted files at once in current history

Automatic detection of fastqsanger format for fastq Uploads

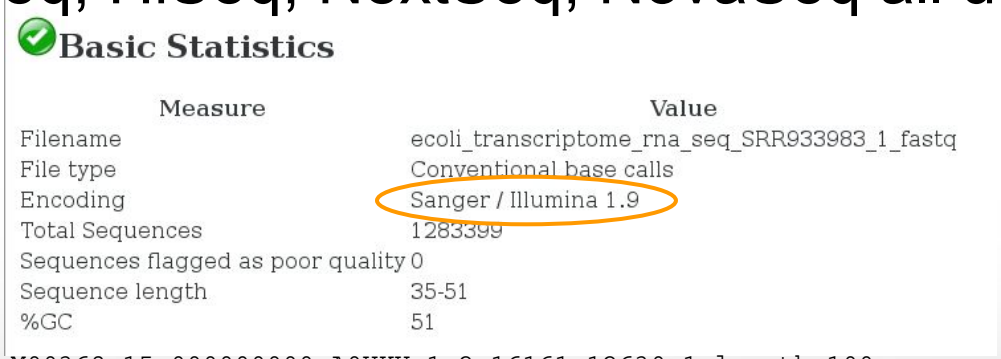
Galaxy will automatically detect fastqsanger format and set format attribute accordingly.

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 for uploaded files
- FastQC tool will check fastq format of a fastq file
  - MiSeq, HiSeq, NextSeq, NovaSeq all use 1.8+

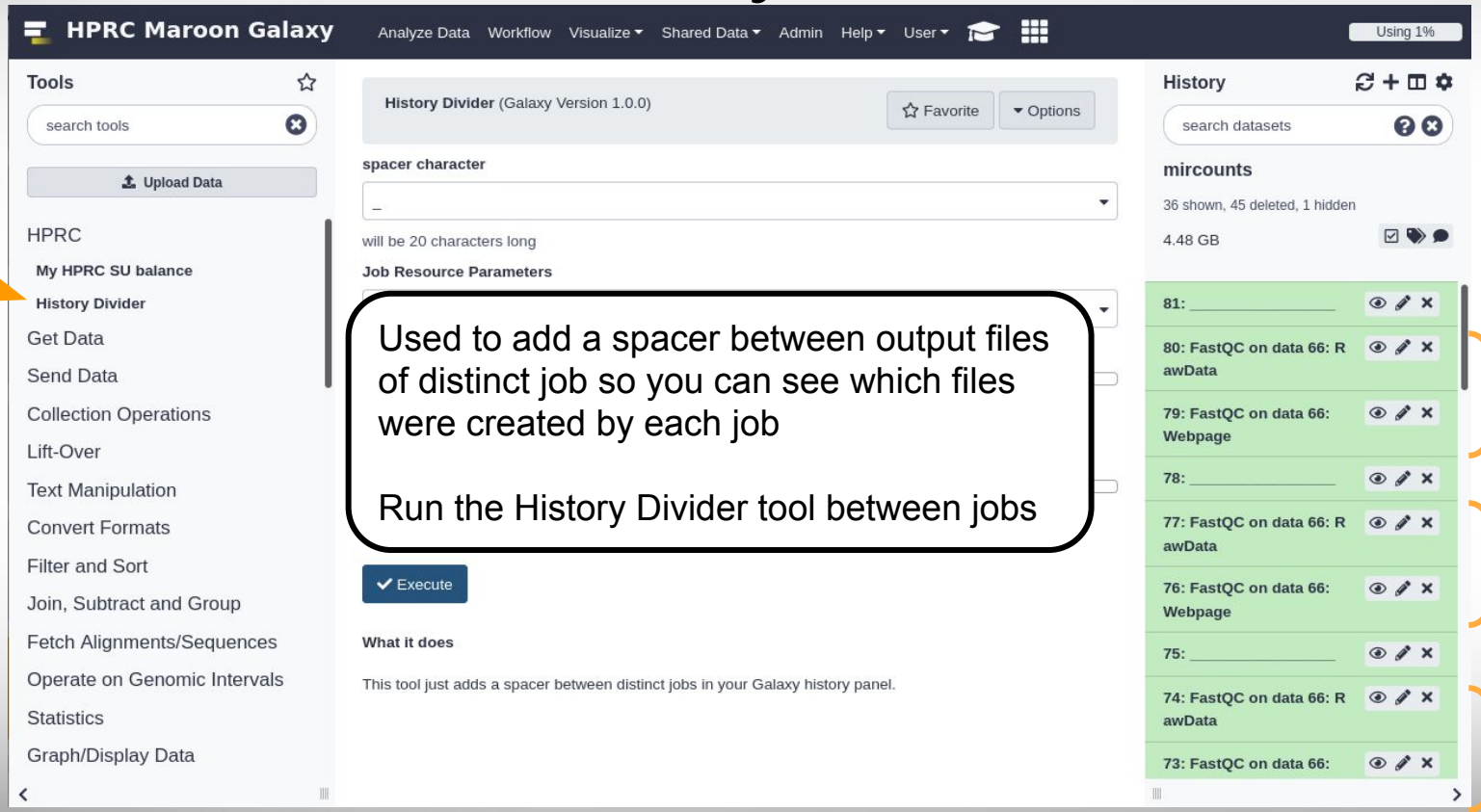


Basic Statistics

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
```

History Divider



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

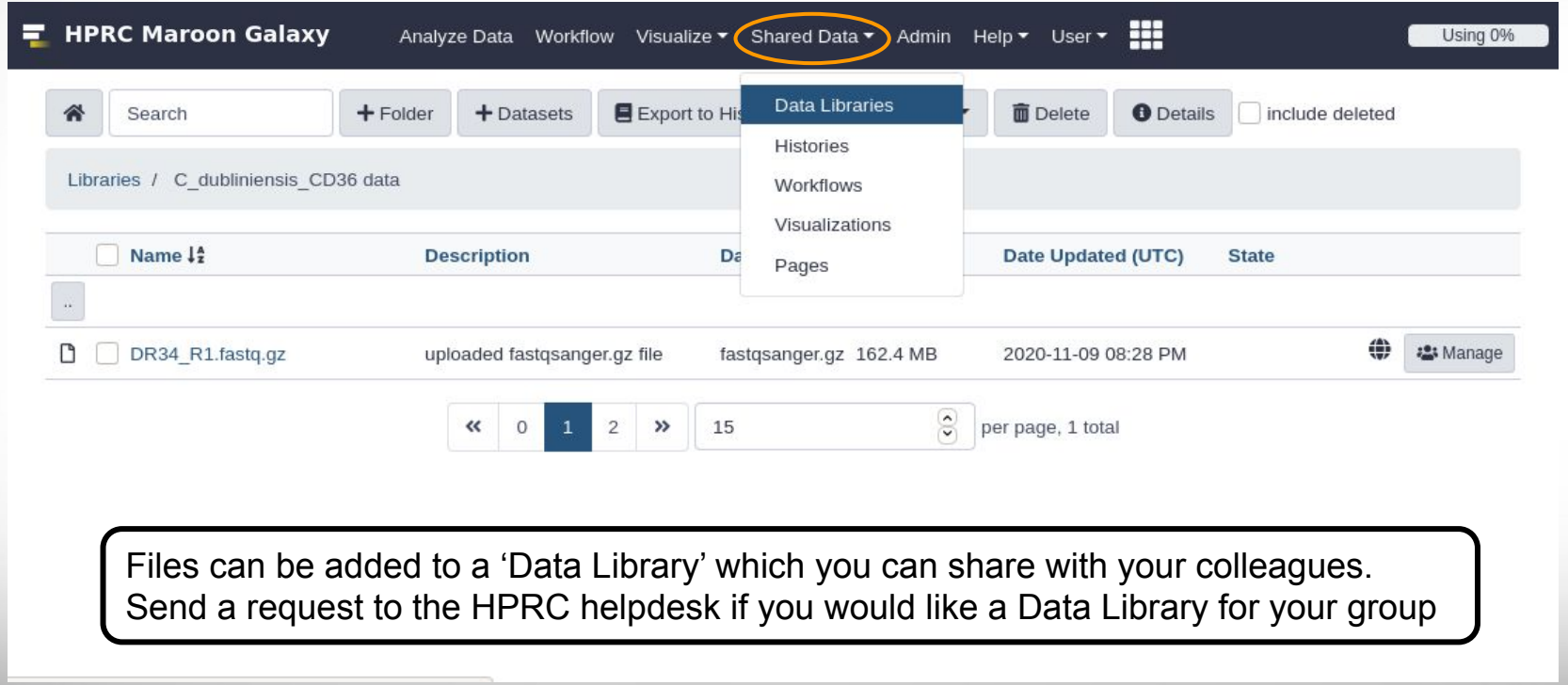
Run the History Divider tool between jobs

job with 2 output files

job with 2 output files

job with 2 output files

Shared Data Libraries

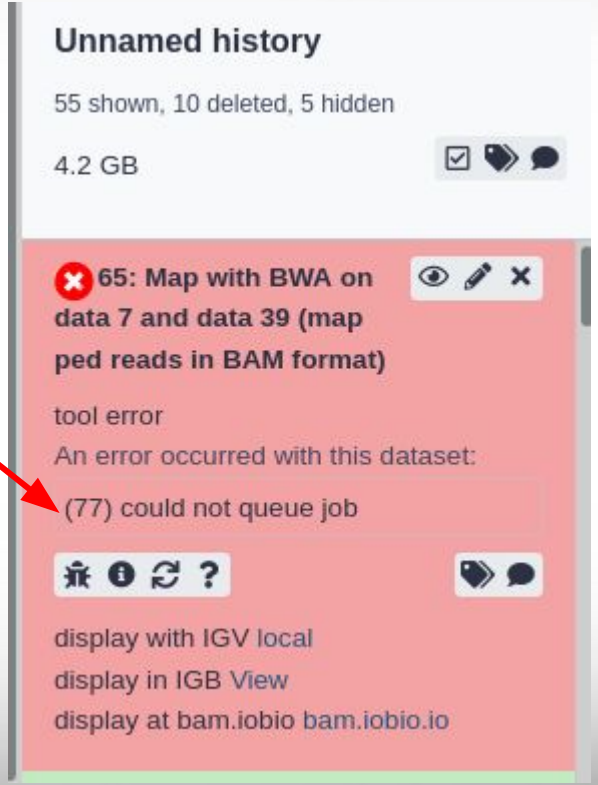


Files can be added to a ‘Data Library’ which you can share with your colleagues. Send a request to the HPRC helpdesk if you would like a Data Library for your group

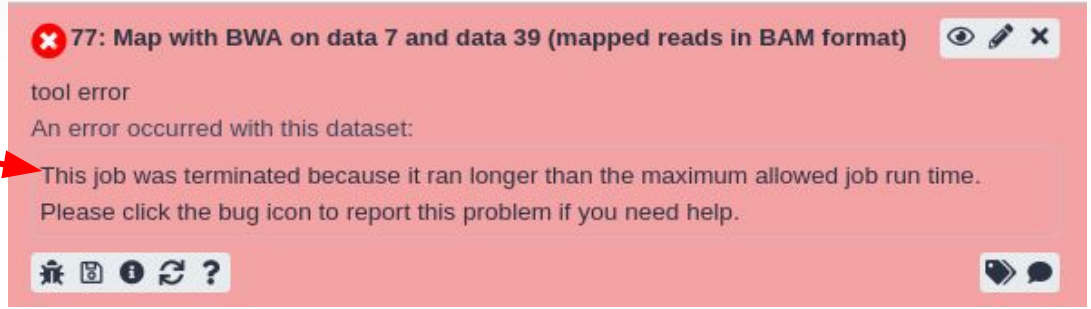
Debugging Failed Jobs

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 the current fiscal year which starts September 1 each year
- 77 is the Galaxy job ID and not a fail code
- Check with the HPRC helpdesk if you have enough SUs and your job does not queue

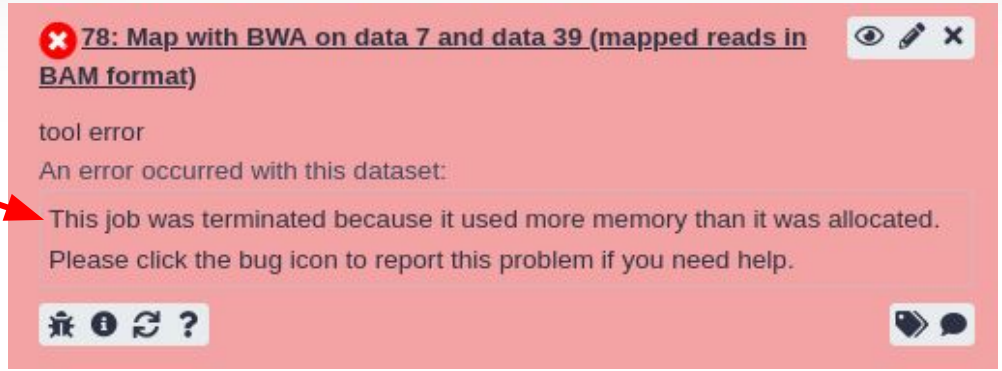


Failed Job: walltime limit reached

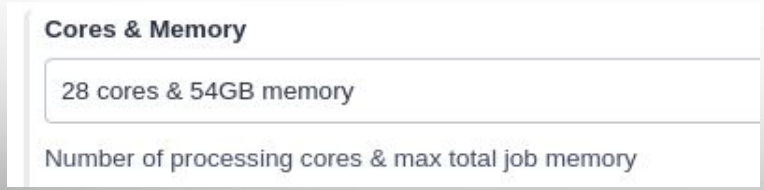


- Configure the job to use more Time

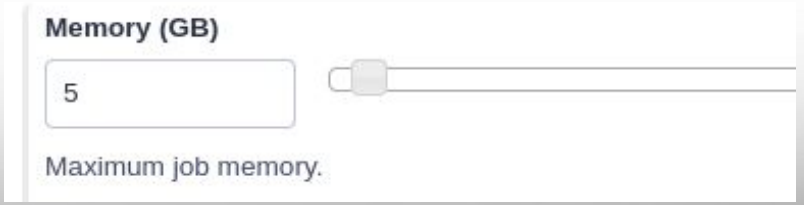
Failed Job: memory limit reached



- Configure the job to use more Memory
- Contact the HPRC helpdesk if you configured the job to use the max allowed memory and it still ran out of memory



or

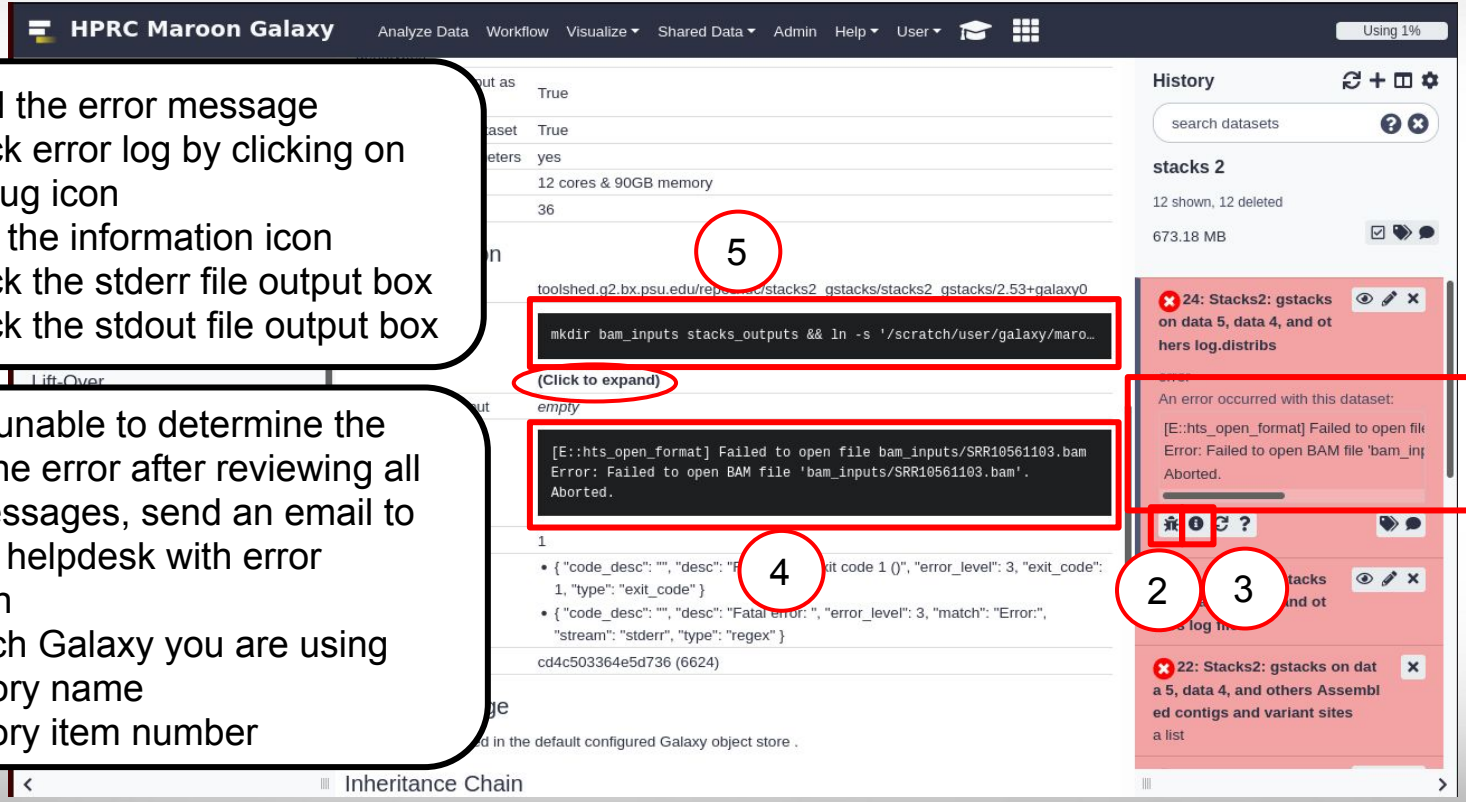


See Logs of Failed Jobs

1. Read the error message
2. Check error log by clicking on the bug icon
3. Click the information icon
4. Check the stderr file output box
5. Check the stdout file output box

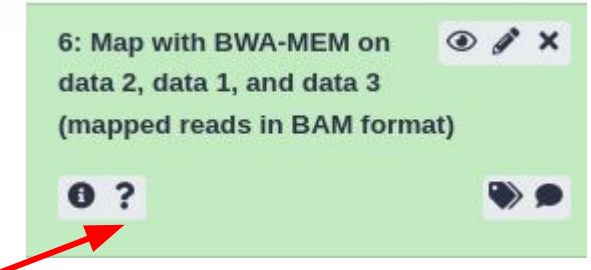
If you are unable to determine the cause of the error after reviewing all the log messages, send an email to the HPRC helpdesk with error information

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



Unable to Download Files

- Problem
  - The disk image download button is missing
- Solution
  - Refresh browser



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

Let us know when the issue has been resolved so we can close the helpdesk ticket