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
- Don't choose easy to guess/crack passwords

For More Help...

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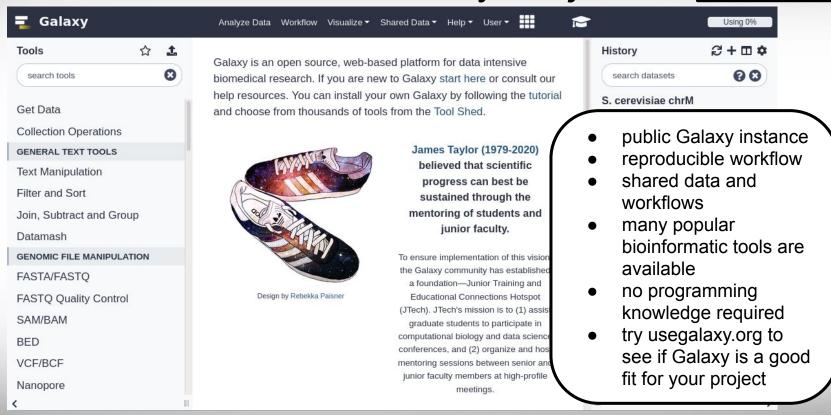Visit us in person: Henderson Hall, Room 114A

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
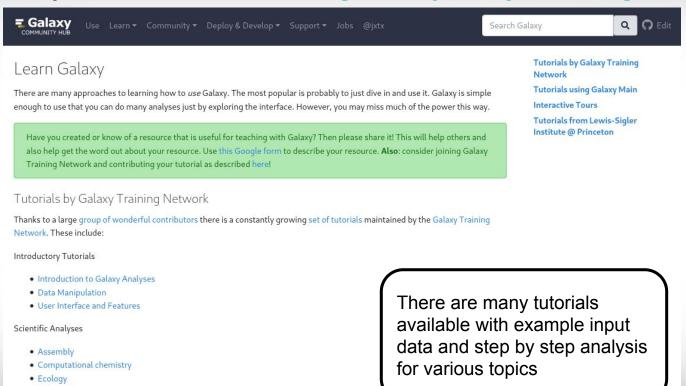- Steps you have taken, so we can reproduce the problem

What is the Galaxy Project?

Galaxy

Analyze Data   Workflow   Visualize ▾   Shared Data ▾   Help ▾   User ▾   Using 0%

Tools ☆ ⬆

search tools

History ⟳ + ▯ ⚙

search datasets ❓ ✕

S. cerevisiae chrM

Get Data
Collection Operations
GENERAL TEXT TOOLS
Text Manipulation
Filter and Sort
Join, Subtract and Group
Datamash
GENOMIC FILE MANIPULATION
FASTA/FASTQ
FASTQ Quality Control
SAM/BAM
BED
VCF/BCF
Nanopore

Galaxy is an open source, web-based platform for data intensive biomedical research. If you are new to Galaxy start here or consult our help resources. You can install your own Galaxy by following the tutorial and choose from thousands of tools from the Tool Shed.

James Taylor (1979-2020)
believed that scientific progress can best be sustained through the mentoring of students and junior faculty.

To ensure implementation of this vision the Galaxy community has established a foundation—Junior Training and Educational Connections Hotspot (JTech). JTech's mission is to (1) assist graduate students to participate in computational biology and data science conferences, and (2) organize and host mentoring sessions between senior and junior faculty members at high-profile meetings.

Design by Rebekka Paisner

- 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 Maroon Galaxy

HPRC Maroon Galaxy

HPRC Maroon Galaxy

Analyze Data   Workflow   Visualize ▾   Shared Data ▾   Admin   Help ▾   User ▾

Using 0%

Tools ☆

search tools

⬆ Upload Data

HPRC
Get Data
Send Data
Collection Operations
Lift-Over
Text Manipulation
Convert Formats
Filter and Sort
Join, Subtract and Group
Fetch Alignments/Sequences
Operate on Genomic Intervals

Best Practices for Maroon Galaxy

- Contact the HPRC helpdesk with an email to request a new or updated tool, indexed genome or to report an error. (Ma...
- All users begin with a file quota of 1TB. Request an increase if you need more disk space but permanently delete nones... requesting.
- FTP uploads are removed from ftp directory 48 hours after uploading so import your ftp files into Galaxy the same day as y...
- The default job resource parameters for all tools is 1 core with 9GB memory for 24 hours (24 SUs).
  - Configuring a job to use 360GB memory for 1 hour requires 48 SUs. (360GB memory for 168 hours = 8,064 SUs).
- Only tools that support multi-core processing have the Job Resources Parameters option which allow you to select cores, memory and time.
  - Configuring a job to use 48 cores and 360GB memory for 1 hour requires 48 SUs. (48 cores for 168 hours = 8,064 SUs).
  - Configuring a job to use 80 cores and 2.93TB memory for 1 hour requires 80 SUs. (80 cores for 168 hours = 13,440 SUs).

ⓘ COVID-19 related research on Galaxy: training, tutorials, documents

⚠ **Current known issues:**
Some tools cannot be removed from favorites. Choose your favorites wisely.

Uploading data from your computer
from Galaxy Project

Local file upload

00:32

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

32: _____
31: Summary Statistics on data 26
30: Summary Statistics on data 26
29: test_sam
28: Map with minimap2 on data 4 and data 5 (mapped reads in BAM format)
26: test.bed
23: Map with minimap2 on data 4 and data 5 (mapped reads in BAM format)
22: Map with minimap2 on data 4 and data 5 (mapped reads in BAM format)

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

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

Texas A&M University   **High Performance Research Computing**   **https://hprc.tamu.edu**

HPRC Galaxy Notes

Galaxy

Maroon Galaxy Accounts

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

It is currently in the testing mode and it will be production mode on June 1st so jobs may be stopped without notice in order to apply updates.

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🔗

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 the browser when you are finished using Galaxy.
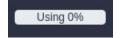
hprc.tamu.edu/wiki/SW:Galaxy

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

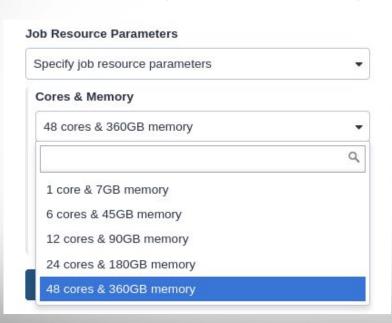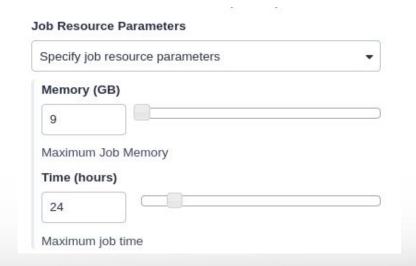4. You can add tools to your favorites to easily locate later

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

Job Resource Parameters

Specify job resource parameters ▾

Cores & Memory

48 cores & 360GB memory ▾

🔍

1 core & 7GB memory

6 cores & 45GB memory

12 cores & 90GB memory

24 cores & 180GB memory

48 cores & 360GB memory

Job Resource Parameters

Specify job resource parameters ▾

Memory (GB)

9

Maximum Job Memory

Time (hours)

24

Maximum job time

New Galaxy Features cont.

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

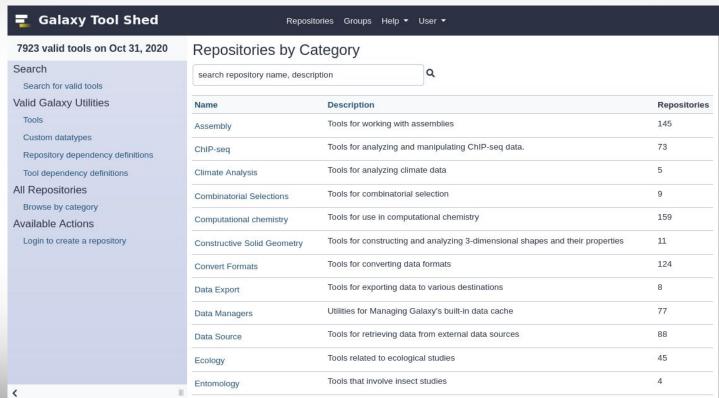   tools supporting **multi-core** and requiring big memory processing

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

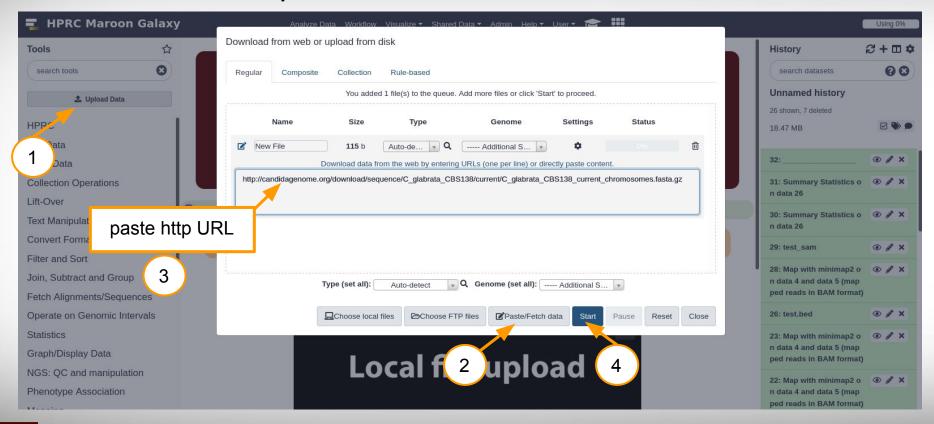# Uploading files to your Galaxy History

Get Data into Your Galaxy History

HPRC Maroon Galaxy

Tools ☆

search tools ✕

⬆ Upload Data

upload button

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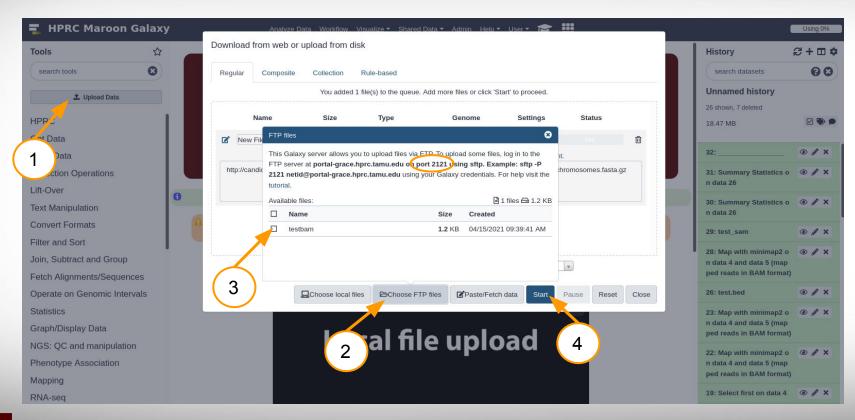
HTTP URL Upload File < 2GB in size or Direct Paste

Uploading a 2GB+ size file to Maroon 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 Grace $SCRATCH directory to ftp directory using sftp on Grace
  c. Use sftp on BitVise on your Windows desktop to copy from your desktop to 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

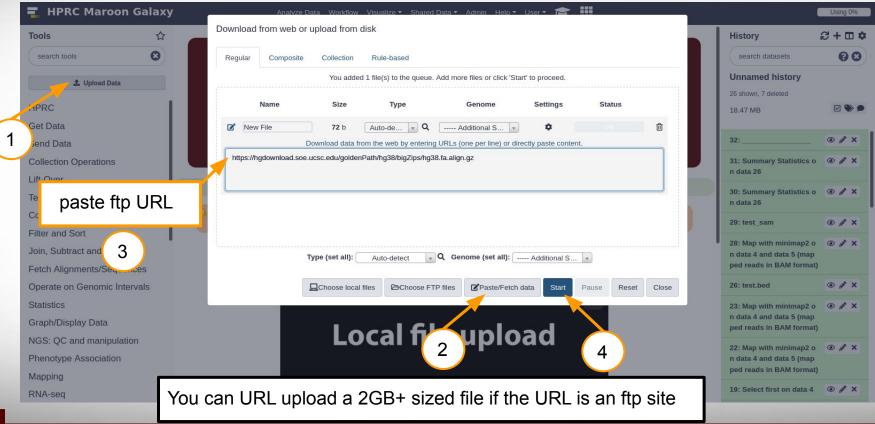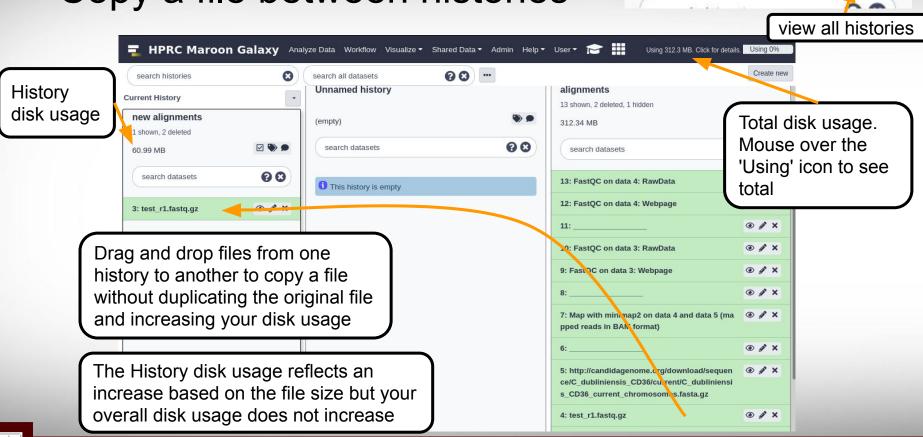# Add Your FTP Uploaded 2GB+ size File to Your History

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

paste ftp URL

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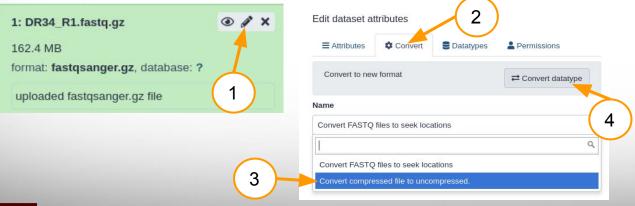
Copy a file between histories

File and History Management
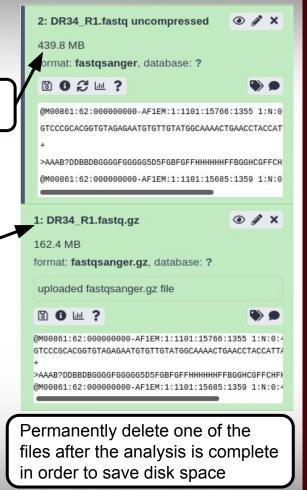
Converting and Using gzipped files

- 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

2: DR34_R1.fastq uncompressed

439.8 MB

format: **fastqsanger**, database: **?**

@M00861:62:000000000-AF1EM:1:1101:15766:1355 1:N:0
GTCCCGCACGGTGTAGAGAATGTGTTGTATGGCAAAACTGAACCTACCAT
+
>AAAB?DDBBDBGGGGFGGGGG5D5FGBFGFFHHHHHHHFFBGGHCGFFCH
@M00861:62:000000000-AF1EM:1:1101:15685:1359 1:N:0

1: DR34_R1.fastq.gz

162.4 MB

format: **fastqsanger.gz**, database: **?**

uploaded fastqsanger.gz file

@M00861:62:000000000-AF1EM:1:1101:15766:1355 1:N:0:4
GTCCCGCACGGTGTAGAGAATGTGTTGTATGGCAAAACTGAACCTACCATTA
+
>AAAB?DDBBDBGGGGFGGGGG5D5FGBFGFFHHHHHHHFFBGGHCGFFCHFH
@M00861:62:000000000-AF1EM:1:1101:15685:1359 1:N:0:4

1: DR34_R1.fastq.gz

162.4 MB

format: **fastqsanger.gz**, database: **?**

uploaded fastqsanger.gz file

①

Edit dataset attributes

≡ Attributes   ✿ Convert   ≋ Datatypes   ⚹ Permissions

②

Convert to new format   ⇄ Convert datatype

④

Name

Convert FASTQ files to seek locations

Convert FASTQ files to seek locations

Convert compressed file to uncompressed.

③

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

History

search datasets

candida variant calling

64 shown, 14 deleted, 5 hidden

4.33 GB

39: http://candidagenome.
org/download/sequence/C
_dubliniensis_CD36/current/C_dublini
ensis_CD36_current_chromosomes.fa
sta.gz

Delete

History

search datasets

candida variant calling

63 shown, 15 deleted, 5 hidden

4.33 GB

40: https://raw.githubuser
content.com/bgruening/ga
laxytools/adf077b912ddebd97b07b947
b855cdd2862ed8ef/tools/augustus/test
-data/human_augustus.fa

History

search datasets

candida variant calling

78 shown, hide deleted, 5 hidden

4.33 GB

⚠ This dataset has been deleted
Undelete it
Permanently remove it from disk

39: http://candidagenome.
org/download/sequence/C
_dubliniensis_CD36/current/C_dublini
ensis_CD36_current_chromosomes.fa

History

search datasets

candida variant calling

78 shown, hide deleted, 5 hidden

4.32 GB

⚠ This dataset has been deleted and
removed from disk

39: http://candidagenome.
org/download/sequence/C
_dubliniensis_CD36/current/C_dublini
ensis_CD36_current_chromosomes.fa
sta.gz

Notice how History disk usage went down only after permanently deleting file
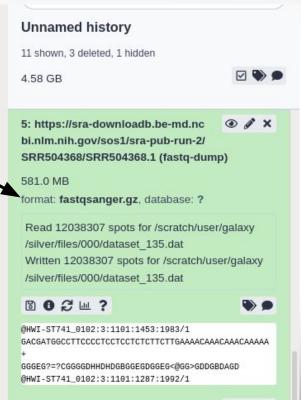
Automatic detection of fastqsanger format for fastq Uploads

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

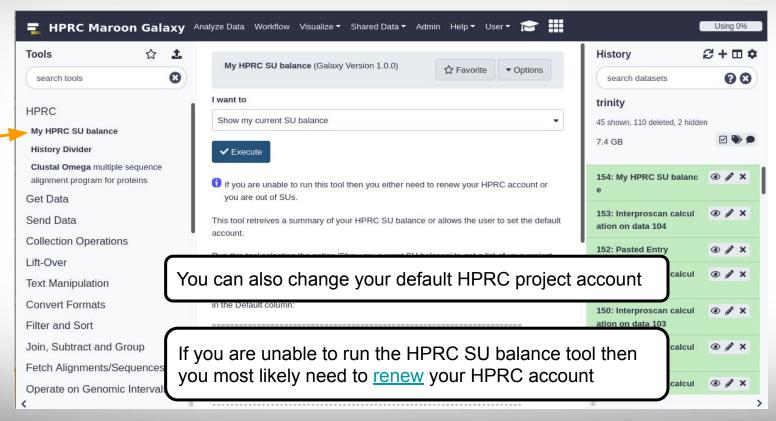There is no need to run the Fastq Groomer tool on this file.

Unnamed history

11 shown, 3 deleted, 1 hidden

4.58 GB

5: https://sra-downloadb.be-md.ncbi.nlm.nih.gov/sos1/sra-pub-run-2/SRR504368/SRR504368.1 (fastq-dump)

581.0 MB

format: **fastqsanger.gz**, database: **?**

Read 12038307 spots for /scratch/user/galaxy/silver/files/000/dataset_135.dat
Written 12038307 spots for /scratch/user/galaxy/silver/files/000/dataset_135.dat

@HWI-ST741_0102:3:1101:1453:1983/1
GACGATGGCCTTCCCCTCCTCCTCTCTTCTTGAAAACAAACAAACAAAAA
+
GGGEG?=?CGGGGDHHDHDGBGGEGDGGEG<@GG>GDDGBDAGD
@HWI-ST741_0102:3:1101:1287:1992/1

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
CCCFFFFFHHHHHJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJIJJJHIJHHGFFFFFFFEEEEEEDDDDDDDDDCDDDD
```

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
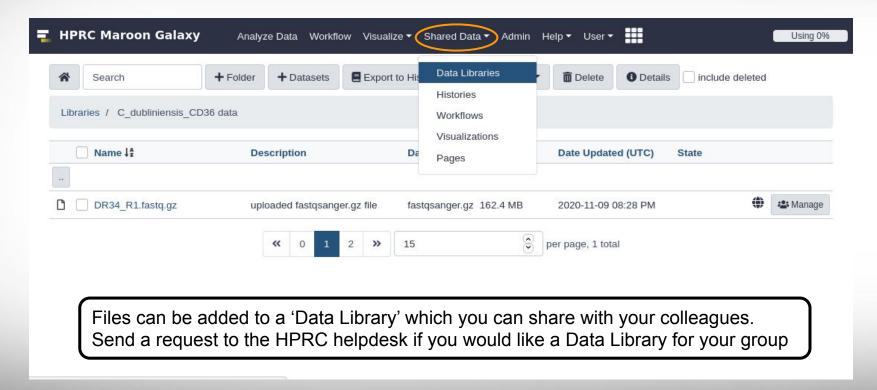
History Divider

Shared Data Libraries

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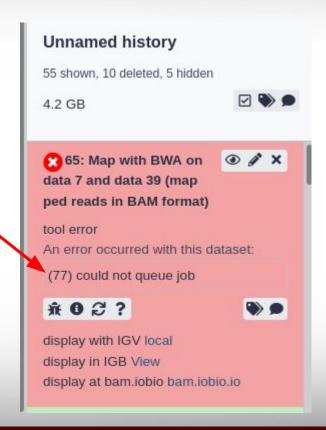Send a request to the HPRC helpdesk if you would like a Data Library for your group
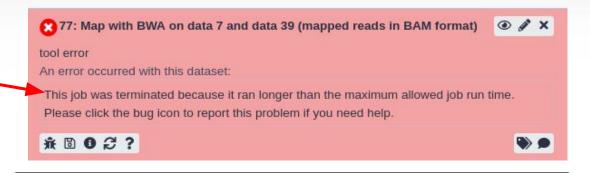
Debugging Failed Jobs

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
- 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

Unnamed history

55 shown, 10 deleted, 5 hidden

4.2 GB

☒ 🏷 💬

❌ 65: Map with BWA on data 7 and data 39 (mapped reads in BAM format) 👁 ✏ ✖

tool error
An error occurred with this dataset:

(77) could not queue job

⚕ ⓘ ⟳ ❓   🏷 💬

display with IGV local
display in IGB View
display at bam.iobio bam.iobio.io

Failed Job: walltime limit reached

❌ 77: Map with BWA on data 7 and data 39 (mapped reads in BAM format) 👁 ✏ ✖

tool error
An error occurred with this dataset:

This job was terminated because it ran longer than the maximum allowed job run time.
Please click the bug icon to report this problem if you need help.

- Configure the job to use more Time

Time (hours)

24
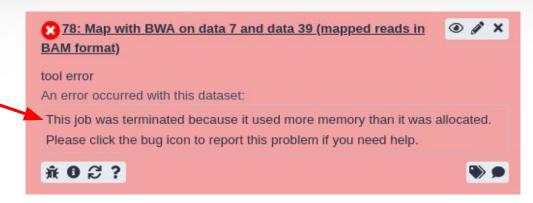
Maximum job time.

Failed Job: memory limit reached

78: Map with BWA on data 7 and data 39 (mapped reads in BAM format)

tool error
An error occurred with this dataset:

This job was terminated because it used more memory than it was allocated.
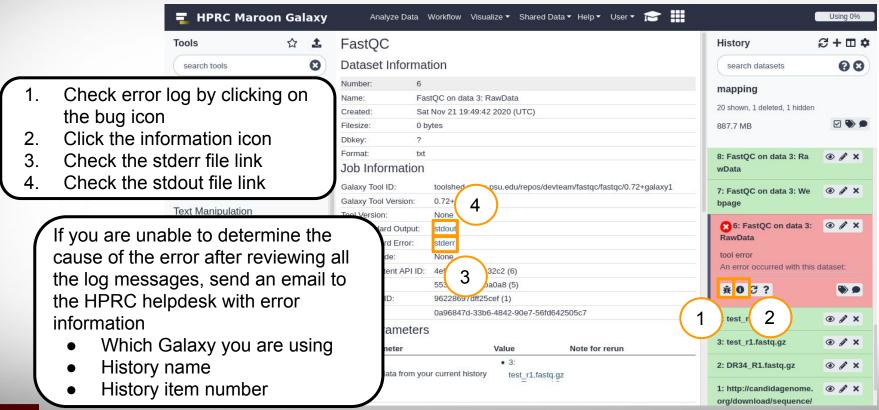Please click the bug icon to report this problem if you need help.

- Configure the job to use more Memory
- Contact the HPRC helpdesk if you configured the job to use the max allowed memory

Cores & Memory

28 cores & 54GB memory

Number of processing cores & max total job memory

or

Memory (GB)

5

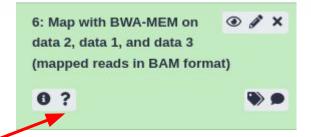Maximum job memory.

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

Unable to Download Files

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

6: Map with BWA-MEM on data 2, data 1, and data 3 (mapped reads in BAM format)

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?