

**High Performance Research Computing**

*A Resource for Research and Discovery*



**TEXAS A&M**  
UNIVERSITY.

# HPRC Maroon Galaxy



# What is the Galaxy Project?

usegalaxy.org

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

History

search datasets

S. cerevisiae chrM

- 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>

 Use Learn ▾ Community ▾ Deploy & Develop ▾ Support ▾ Jobs @jtxx    Edit

## Learn Galaxy

There are many approaches to learning how to use Galaxy. The most popular is probably to just dive in and use it. Galaxy is simple enough to use that you can do many analyses just by exploring the interface. However, you may miss much of the power this way.

Have you created or know of a resource that is useful for teaching with Galaxy? Then please share it! This will help others and also help get the word out about your resource. Use [this Google form](#) to describe your resource. **Also:** consider joining Galaxy Training Network and contributing your tutorial as described [here!](#)

[Tutorials by Galaxy Training Network](#)

[Tutorials using Galaxy Main](#)

[Interactive Tours](#)

[Tutorials from Lewis-Sigler Institute @ Princeton](#)

## Tutorials by Galaxy Training Network

Thanks to a large [group of wonderful contributors](#) there is a constantly growing [set of tutorials](#) maintained by the [Galaxy Training Network](#). These include:

### Introductory Tutorials

- [Introduction to Galaxy Analyses](#)
- [Data Manipulation](#)
- [User Interface and Features](#)

### Scientific Analyses

- [Assembly](#)
- [Computational chemistry](#)
- [Ecology](#)
- [Epigenetics](#)
- [Genome Annotation](#)
- [Imaging](#)

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

154 videos

**Galaxy Project**  
PSU/JH  
Galaxy is an open source, web-based platform for data intensive biomedical research. Whether on...[Read more](#)  
UseGalaxy.org

**Galaxy COVID-19: Variation analysis**

**Confound it! Reproducible biology from ...**



# Apply for a usegalaxy.org account

[usegalaxy.org/login](https://usegalaxy.org/login)

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

Please register only one account. The usegalaxy.org service is provided free of charge and has limited computational and data storage resources. **Registration and usage of multiple accounts is tracked and such accounts are subject to termination and data deletion.**

Create a Galaxy account

Email Address

Password

Confirm password

Public name

Your public name is an identifier that will be used to generate addresses for information you share publicly. Public names must be at least three characters in length and contain only lower-case letters, numbers, dots, underscores, and dashes ('.', '\_', '-').

Subscribe to mailing list

Create



# 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](https://usegalaxy.org/login)

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

The screenshot shows the Galaxy Training website. The top navigation bar includes 'Galaxy', 'Workflow', 'Visualize', 'Shared Data', 'Help', 'Login or Register', and a graduation hat icon. The main content area is titled 'Galaxy Training!' and features a search bar and navigation links for 'Contributors', 'Help', 'Extras', and 'Search Tutorials'. Below the header, there is a 'Welcome to Galaxy Training!' message and a description: 'Collection of tutorials developed and maintained by the worldwide Galaxy community'. Two tables are displayed: 'Galaxy for Scientists' and 'Galaxy Tips & Tricks'. The 'Galaxy for Scientists' table lists topics and the number of tutorials: Introduction to Galaxy Analyses (9), Assembly (9), Climate (4), and Computational chemistry (6). The 'Galaxy Tips & Tricks' table lists 'Using Galaxy and Managing your Data' (20). A 'Data Science Survival Kit' section is also visible with an 'OPEN CHAT' button.

Topic	Tutorials
<a href="#">Introduction to Galaxy Analyses</a>	9
<a href="#">Assembly</a>	9
<a href="#">Climate</a>	4
<a href="#">Computational chemistry</a>	6

Topic	Tutorials
<a href="#">Using Galaxy and Managing your Data</a>	20

# 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 screenshot shows the HPRC Maroon Galaxy web interface. At the top, there is a navigation bar with links for 'Analyze Data', 'Workflow', 'Visualize', 'Shared Data', 'Admin', 'Help', 'User', and a user profile icon. A search bar is on the left, and a 'Using 0%' indicator is on the right. The main content area features a dark red box titled 'Best Practices for Maroon Galaxy' with a list of guidelines. Below this is a green banner for COVID-19 research and an orange banner for current known issues. On the right, a list of data items is visible, including '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)', and '22: Map with minimap2 on data 4 and data 5 (mapped reads in BAM format)'. A video player at the bottom center shows a video titled 'Uploading data from your computer' from Galaxy Project, with the text 'Local file upload' overlaid. A callout box on the right points to the user profile icon in the navigation bar.

**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. (Maroon Galaxy)
- All users begin with a file quota of 1TB. Request an increase if you need more disk space but permanently delete nonessential data.
- FTP uploads are removed from ftp directory 48 hours after uploading so import your ftp files into Galaxy the same day as you upload.
- 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.

**Local file upload**

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

The same tutorials found at [galaxyproject.org/learn](https://galaxyproject.org/learn) can be accessed here

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



# 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](http://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](http://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](mailto: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](https://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

The screenshot shows the HPRC Maroon Galaxy web interface. The top navigation bar includes 'Analyze Data', 'Workflow', 'Visualize', 'Shared Data', 'Admin', 'Help', 'User', and a 'Using 1%' indicator. The left sidebar lists various tools, with 'HPRC' and 'My HPRC SU balance' highlighted by an orange arrow. The main content area displays the configuration for the 'My HPRC SU balance' tool (Galaxy Version 1.0.0). It includes a 'Favorite' button and an 'Options' dropdown. The 'I want to' section has a dropdown menu set to 'Show my current SU balance'. The 'Job Resource Parameters' section has a dropdown set to 'Specify job resource parameters'. Below these are fields for 'Maximum Job Memory' and an 'Execute' button. A blue information icon is present above the execution instructions.

**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**

This tool retrieves a summary of your HPRC SU balance or allows the user to set the default account. Run this tool selecting the option 'Show my current SU balance' to get a list of your project account numbers. In the following example, a default account is set to Account 000000000001 since it has a Y in the Default column:

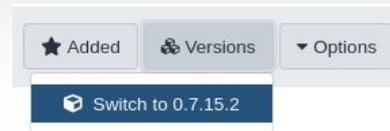
The History panel shows a search for 'trinity' with 69 shown and 10 deleted. It lists several runs, including '79: My HPRC SU balance' which is expanded to show 8 lines of output. The output includes a table of project accounts.

Account	FY	Default	Allocati

Other runs listed include '78: Trinity on data 25 an d data 24: Gene to trans cripts map' and '77: Trinity on data 25 an d data 24: Assembled Tr anscripts'.

# 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

**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**

tools supporting **single-core** processing

**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

**Job Resource Parameters**

Specify job resource parameters ▼

**Cores & Big Memory**

48 cores & 360GB memory ▼

- 48 cores & 360GB memory
- 20 cores & 730GB memory
- 40 cores & 1.44TB memory
- 60 cores & 2.19TB memory
- 80 cores & 2.93TB memory

Execute

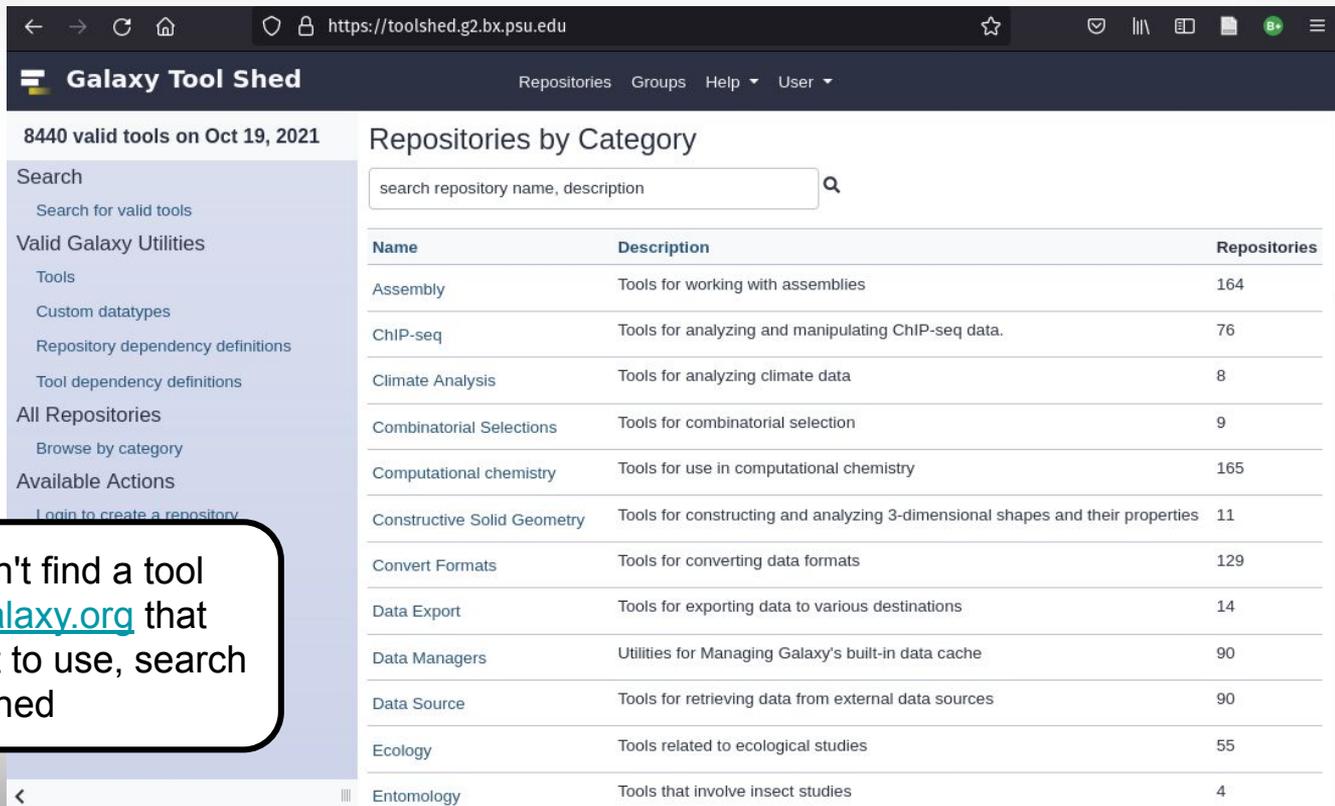
# Maroon Galaxy

- 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](mailto: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>



The screenshot shows the Galaxy Tool Shed interface. At the top, the browser address bar displays <https://toolshed.g2.bx.psu.edu>. The main header includes the Galaxy Tool Shed logo and navigation links for Repositories, Groups, Help, and User. A notification states "8440 valid tools on Oct 19, 2021". On the left, a search sidebar offers options like "Search for valid tools", "Valid Galaxy Utilities", and "All Repositories". The main content area, titled "Repositories by Category", features a search input and a table listing various tool categories and their repository counts.

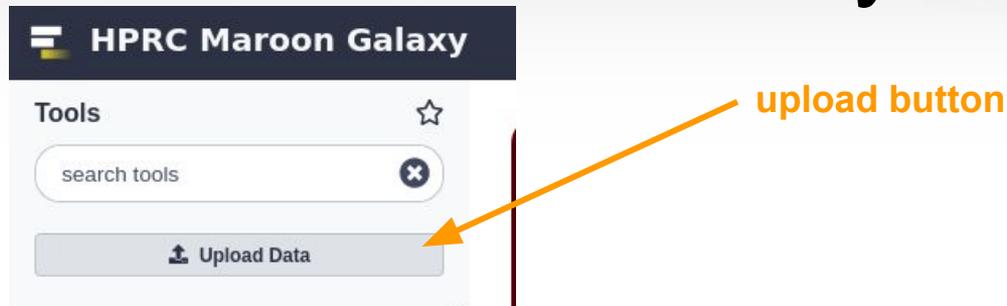
Name	Description	Repositories
Assembly	Tools for working with assemblies	164
ChIP-seq	Tools for analyzing and manipulating ChIP-seq data.	76
Climate Analysis	Tools for analyzing climate data	8
Combinatorial Selections	Tools for combinatorial selection	9
Computational chemistry	Tools for use in computational chemistry	165
Constructive Solid Geometry	Tools for constructing and analyzing 3-dimensional shapes and their properties	11
Convert Formats	Tools for converting data formats	129
Data Export	Tools for exporting data to various destinations	14
Data Managers	Utilities for Managing Galaxy's built-in data cache	90
Data Source	Tools for retrieving data from external data sources	90
Ecology	Tools related to ecological studies	55
Entomology	Tools that involve insect studies	4

If you can't find a tool on [usegalaxy.org](https://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

Download from web or upload from disk

Regular Composite Collection Rule-based

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

Name	Size	Type	Genome	Settings	Status
New File	115 b	Auto-de...	----- Additional S...	⚙️	0%

Download data from the web by entering URLs (one per line) or directly paste content.

`http://candidagenome.org/download/sequence/C_glabrata_CBS138/current/C_glabrata_CBS138_current_chromosomes.fasta.gz`

Type (set all): Auto-detect Genome (set all): ----- Additional S...

Choose local files Choose FTP files **Paste/Fetch data** **Start** Pause Reset Close

1

2

3

4

paste http URL

Local file upload

# 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](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

1

2

3

4

FTP files

This Galaxy server allows you to upload files via FTP. To upload some files, log in to the FTP server at `portal-grace.hprc.tamu.edu` on port 2121 using sftp. Example: `sftp -P 2121 netid@portal-grace.hprc.tamu.edu` using your Galaxy credentials. For help visit the tutorial.

Available files: 1 files 1.2 KB

<input type="checkbox"/>	Name	Size	Created
<input checked="" type="checkbox"/>	testbam	1.2 KB	04/15/2021 09:39:41 AM

Choose local files Choose FTP files Paste/Fetch data Start Pause Reset Close

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

Download from web or upload from disk

Regular Composite Collection Rule-based

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

Name	Size	Type	Genome	Settings	Status
<input checked="" type="checkbox"/> New File	142 b	Auto-de...	----- Additional S...		0%

Download data from the web by entering URLs (one per line) or directly paste content.

`http://ftp.ensembl.org/pub/release-104/fasta/zonotrichia_albicollis/dna/Zonotrichia_albicollis.Zonotrichia_albicollis-1.0.1.dna.toplevel.`

Type (set all):  Genome (set all):

Current known issues:  
...ently deleting individual files is not working. After the first delete  
...click the 'History options' gear icon and select 'Purge Deleted'

**1** paste ftp URL

**2**

**3**

**4**

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

# Copy a file between histories

The screenshot displays the HPRC Maroon Galaxy web interface. At the top, a navigation bar includes 'Analyze Data', 'Workflow', 'Visualize', 'Shared Data', 'Admin', 'Help', 'User', and a 'Using 312.3 MB' status indicator. Below the navigation bar, there are search bars for 'search histories' and 'search all datasets'. The main content area is divided into three panels: 'Current History' (containing 'new alignments' with 1 shown and 2 deleted, and a file '3: test\_r1.fastq.gz'), 'Unnamed history' (empty), and 'alignments' (a list of 13 items, including 'FastQC on data 4: RawData' and 'test\_r1.fastq.gz'). A callout box 'view all histories' points to a grid icon in the top right. Another callout 'History disk usage' points to the 'new alignments' panel. A third callout 'Total disk usage. Mouse over the 'Using' icon to see total' points to the 'Using 312.3 MB' status. A fourth callout 'Drag and drop files from one history to another to copy a file without duplicating the original file and increasing your disk usage' points to a drag-and-drop icon in the 'Current History' panel. A fifth callout 'The History disk usage reflects an increase based on the file size but your overall disk usage does not increase' points to the 'alignments' panel.

History disk usage

view all histories

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

1: DR34\_R1.fastq.gz  
162.4 MB  
format: fastqsanger.gz, database: ?  
uploaded fastqsanger.gz file

Edit dataset attributes

Attributes **Convert** Datatypes Permissions

Convert to new format Convert datatype

Name

Convert FASTQ files to seek locations

Convert FASTQ files to seek locations

Convert compressed file to uncompressed.

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  
GTCCCGCACGGTGTAGAGAATGTGTTGTATGGCAAACCTAACCAT  
+  
>AAAB?DDBBDBGGGGFGGGG5D5FGBFGFFHHHHHHFFBGGHCGFFCH  
@M00861:62:000000000-AF1EM:1:1101:15685:1359 1:N:0

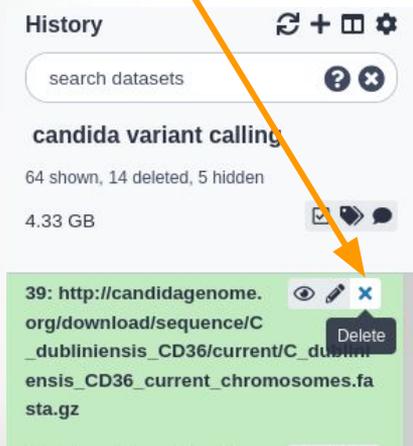
1: DR34\_R1.fastq.gz  
162.4 MB  
format: fastqsanger.gz, database: ?  
uploaded fastqsanger.gz file  
@M00861:62:000000000-AF1EM:1:1101:15766:1355 1:N:0:  
GTCCCGCACGGTGTAGAGAATGTGTTGTATGGCAAACCTAACCAT/  
+  
>AAAB?DDBBDBGGGGFGGGG5D5FGBFGFFHHHHHHFFBGGHCGFFCH/  
@M00861:62:000000000-AF1EM:1:1101:15685:1359 1:N:0:0

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



History

search datasets

candida variant calling

64 shown, 14 deleted, 5 hidden

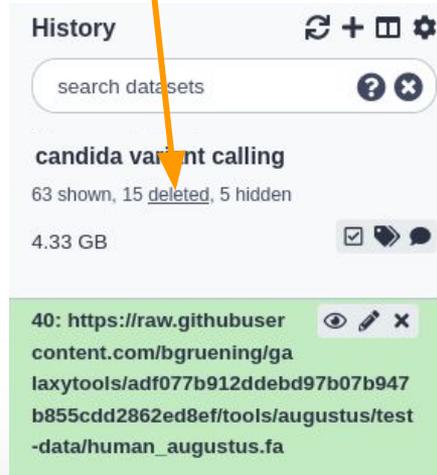
4.33 GB

39: [http://candidagenome.org/download/sequence/C\\_dublينيensis\\_CD36/current/C\\_dublينيensis\\_CD36\\_current\\_chromosomes.fasta.gz](http://candidagenome.org/download/sequence/C_dublينيensis_CD36/current/C_dublينيensis_CD36_current_chromosomes.fasta.gz)

Delete

An orange arrow points from the 'delete file' text to the 'Delete' button in the file entry.

show deleted files



History

search datasets

candida variant calling

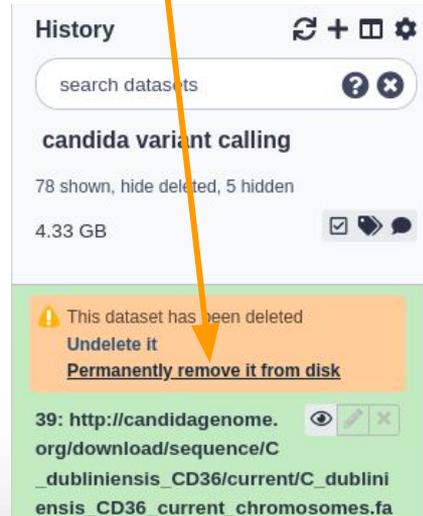
63 shown, 15 deleted, 5 hidden

4.33 GB

40: [https://raw.githubusercontent.com/bgruening/galaxytools/adf077b912ddebd97b07b947b855cdd2862ed8ef/tools/augustus/test-data/human\\_augustus.fa](https://raw.githubusercontent.com/bgruening/galaxytools/adf077b912ddebd97b07b947b855cdd2862ed8ef/tools/augustus/test-data/human_augustus.fa)

An orange arrow points from the 'show deleted files' text to the 'deleted' link in the file entry.

permanently delete file



History

search datasets

candida variant calling

78 shown, hide deleted, 5 hidden

4.33 GB

This dataset has been deleted

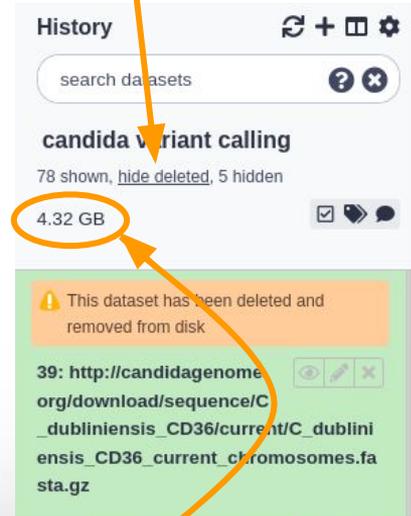
Undo it

Permanently remove it from disk

39: [http://candidagenome.org/download/sequence/C\\_dublينيensis\\_CD36/current/C\\_dublينيensis\\_CD36\\_current\\_chromosomes.fasta.gz](http://candidagenome.org/download/sequence/C_dublينيensis_CD36/current/C_dublينيensis_CD36_current_chromosomes.fasta.gz)

An orange arrow points from the 'permanently delete file' text to the 'Permanently remove it from disk' link.

hide deleted files



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\\_dublينيensis\\_CD36/current/C\\_dublينيensis\\_CD36\\_current\\_chromosomes.fasta.gz](http://candidagenome.org/download/sequence/C_dublينيensis_CD36/current/C_dublينيensis_CD36_current_chromosomes.fasta.gz)

An orange arrow points from the 'hide deleted files' text to the 'hide deleted' link. Another orange arrow points from the '4.32 GB' disk usage to a callout box.

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"

The screenshot shows the HPRC Maroon Galaxy interface. The 'History' panel on the right is open, showing a list of datasets. The 'History Actions' menu is expanded, and 'Purge Deleted Datasets' is highlighted. A confirmation dialog box is overlaid on the screen, asking 'Really delete all deleted datasets permanently? This cannot be undone.' with 'Cancel' and 'OK' buttons. Three orange arrows point to the gear icon (1), the 'Purge Deleted Datasets' option (2), and the 'OK' button (3).

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. (Maroon Galaxy [docs](#), [slides](#))
- All users begin with a file quota of 1TB. Request an increase if you need more disk space but **permanently delete** nonessential files before requesting.
- FTP uploads are removed from ftp directory 48 hours after uploading so import your ftp files into Galaxy the same day as you upload to ftp
- Currently no SUs are charged for using Grace Maroon Galaxy
  - SUs will be charged beginning September 1, 2021

Really delete all deleted datasets permanently? This cannot be undone.

Cancel OK

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.



The screenshot shows a Galaxy history entry for a file upload. The entry is titled "5: https://sra-downloadb.be-md.ncbi.nlm.nih.gov/sos1/sra-pub-run-2/SRR504368/SRR504368.1 (fastq-dump)" and has a size of 581.0 MB. The format is automatically detected as "fastqsanger.gz" and the database is set to "?". Below the entry, the Galaxy log shows that 1,203,830 spots were read and written for the dataset. A preview of the fastq data is shown at the bottom, including headers like "@HWI-ST741\_0102:3:1101:1453:1983/1" and sequence lines.

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
GACGATGGCCTTCCCCTCCTCTCTTCTTGA AAAACAAAACAAAAA
+
GGGEG?-?CGGGGDHHDHDBGGEGDGGEG<@GG>GDDGBDAGD
@HWI-ST741_0102:3:1101:1287:1992/1
```



# History Divider

The screenshot shows the HPRC Maroon Galaxy interface. On the left is a 'Tools' sidebar with a search bar and an 'Upload Data' button. Below these are categories: 'HPRC', 'My HPRC SU balance', 'History Divider' (highlighted with an orange arrow), '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', 'Statistics', and 'Graph/Display Data'. The main area displays the 'History Divider (Galaxy Version 1.0.0)' tool configuration. It includes a 'spacer character' dropdown menu (currently empty), a character count 'will be 20 characters long', and 'Job Resource Parameters'. An 'Execute' button is visible. Below the configuration is a 'What it does' section: 'This tool just adds a spacer between distinct jobs in your Galaxy history panel.' On the right, the 'History' panel shows a search bar and a list of datasets under the name 'mircounts'. The list includes jobs 81, 80, 79, 78, 77, 76, 75, 74, and 73. Jobs 80, 79, 76, and 73 are grouped with orange brackets and labeled 'job with 2 output files'. Each job entry shows a job ID, a description (e.g., 'FastQC on data 66: R awData'), and icons for viewing, editing, and deleting.

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

The screenshot shows the HPRC Maroon Galaxy interface. The top navigation bar includes 'Analyze Data', 'Workflow', 'Visualize', 'Shared Data' (circled in orange), 'Admin', 'Help', and 'User'. A dropdown menu for 'Shared Data' is open, showing options: 'Data Libraries', 'Histories', 'Workflows', 'Visualizations', and 'Pages'. Below the navigation, there is a search bar and buttons for '+ Folder', '+ Datasets', 'Export to His...', 'Delete', 'Details', and 'include deleted'. The breadcrumb path is 'Libraries / C\_dubliniensis\_CD36 data'. A table lists data libraries with columns: Name, Description, Date Updated (UTC), and State. One entry is visible: 'DR34\_R1.fastq.gz' with description 'uploaded fastqsanger.gz file', size '162.4 MB', and date '2020-11-09 08:28 PM'. A 'Manage' button is next to it. At the bottom, there is a pagination control showing '1' of 1 page, 15 items per page, and 1 total item.

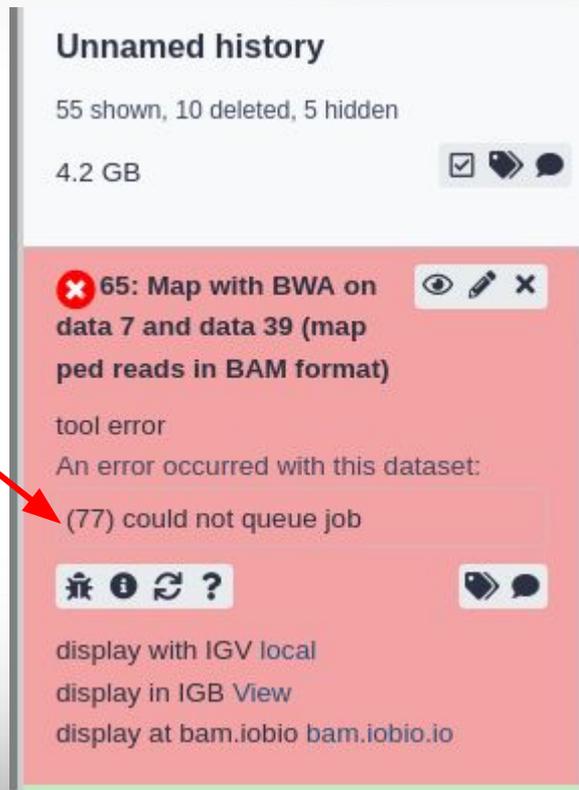
Files can be added to a 'Data Library' which you can share with your colleagues. Send a request to the HPRC helpdesk if you would like a Data Library for your group

# 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



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:

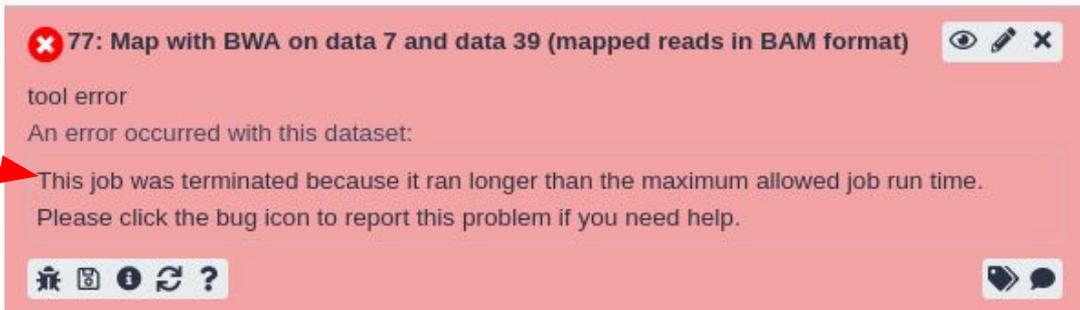
(77) could not queue job

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

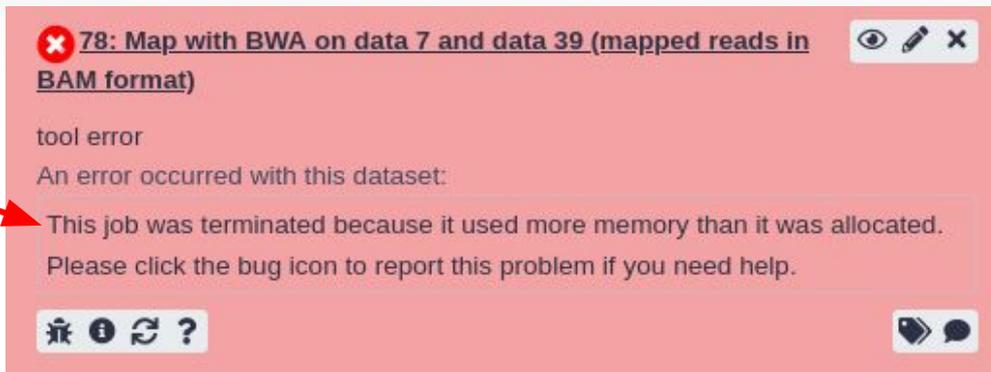
**Time (hours)**

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 and it still ran out of 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. 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

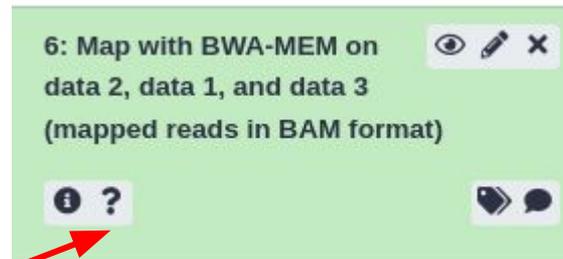
If you are unable to determine the cause of the error after reviewing all the log messages, send an email to the HPRC helpdesk with error information

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

The screenshot shows the HPRC Maroon Galaxy interface. The top navigation bar includes 'Analyze Data', 'Workflow', 'Visualize', 'Shared Data', 'Admin', 'Help', 'User', and 'Using 1%'. The main content area displays job details for 'Stacks2'. A red box labeled '5' highlights the command: `mkdir bam_inputs stacks_outputs && ln -s '/scratch/user/galaxy/maroon...'`. Below it, a red box labeled '4' highlights the error message: `[E::hts_open_format] Failed to open file bam_inputs/SRR10561103.bam Error: Failed to open BAM file 'bam_inputs/SRR10561103.bam'. Aborted.`. On the right, the 'History' panel shows a list of jobs. A red box labeled '1' highlights a job entry: '24: Stacks2: gstacks on data 5, data 4, and others log.distrib'. A red box labeled '2' highlights the bug icon, and a red box labeled '3' highlights the information icon. The 'Inheritance Chain' is visible at the bottom.

# Unable to Download Files

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



# For More Help...

Website: [hprc.tamu.edu](http://hprc.tamu.edu)

Email: [help@hprc.tamu.edu](mailto: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

