

# HIGH PERFORMANCE RESEARCH COMPUTING

## Introduction to HPRC Galaxy

November 21, 2025

Michael Dickens



High Performance  
Research Computing

DIVISION OF RESEARCH



# Galaxy Training Objectives

- usegalaxy.org
- upload and download data
- create new histories
- search and switch tool version
- view data
- view/change file attributes
- permanently delete files
- debug jobs
- copy datasets between histories
- run a variant calling analysis
- create a workflow

# 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.

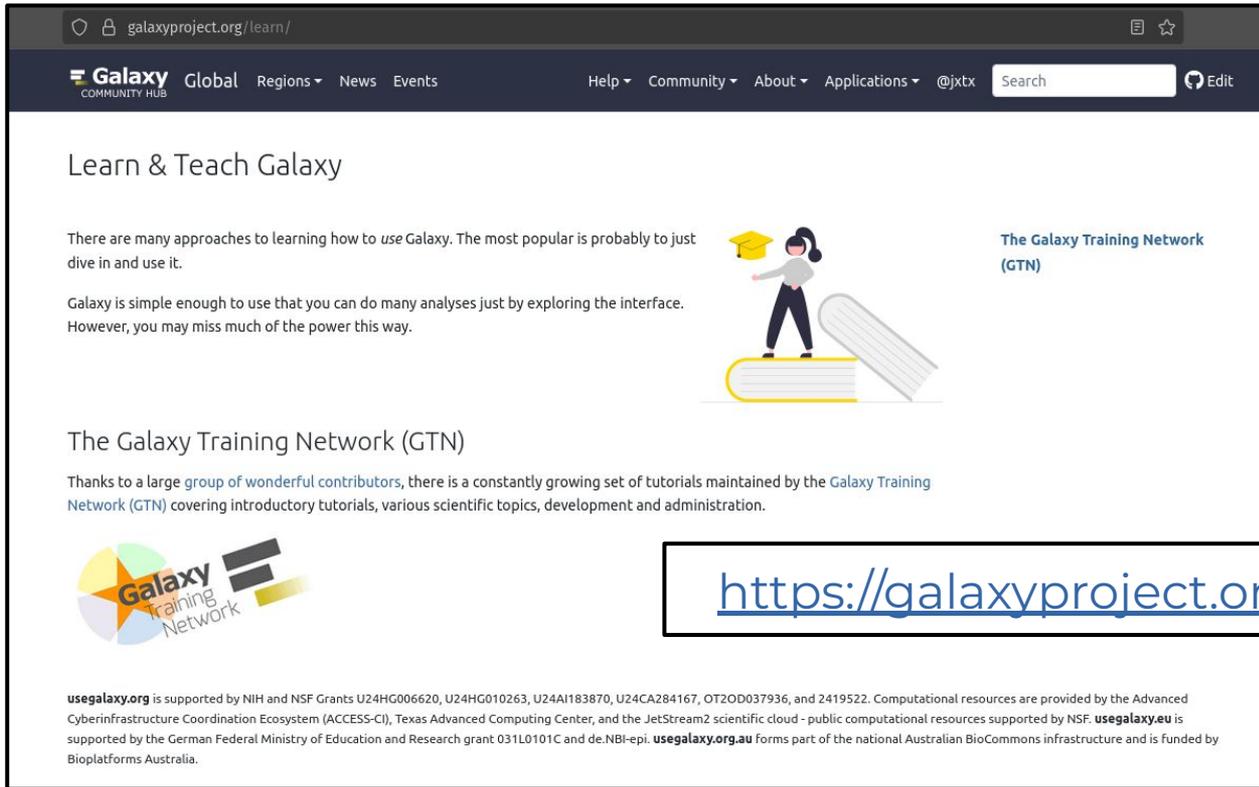
**James P. Taylor**  
Foundation for Open Science.  
“The most important job of senior faculty is to mentor junior faculty and students.” — @jtx

History

- 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

Donate to the James P. Taylor Foundation for Open Science

# Galaxy Videos and Tutorials



The screenshot shows the Galaxy Project Learn page. The browser address bar displays 'galaxyproject.org/learn/'. The navigation bar includes the Galaxy Community Hub logo, 'Global', 'Regions', 'News', 'Events', 'Help', 'Community', 'About', 'Applications', '@jxtb', a search box, and an 'Edit' button. The main content area is titled 'Learn & Teach Galaxy' and contains two paragraphs of text, an illustration of a person with a graduation cap, and a link to 'The Galaxy Training Network (GTN)'. Below this is a section for 'The Galaxy Training Network (GTN)' with a paragraph of text and the GTN logo. A URL box highlights 'https://galaxyproject.org/learn'. At the bottom, a footer paragraph provides funding and support information for usegalaxy.org and usegalaxy.eu.

galaxyproject.org/learn/

Galaxy COMMUNITY HUB Global Regions News Events Help Community About Applications @jxtb Search Edit

## Learn & Teach 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.



[The Galaxy Training Network \(GTN\)](#)

## The Galaxy Training Network (GTN)

Thanks to a large group of wonderful contributors, there is a constantly growing set of tutorials maintained by the Galaxy Training Network (GTN) covering introductory tutorials, various scientific topics, development and administration.



<https://galaxyproject.org/learn>

usegalaxy.org is supported by NIH and NSF Grants U24HG006620, U24HG010263, U24AI183870, U24CA284167, OT2OD037936, and 2419522. Computational resources are provided by the Advanced Cyberinfrastructure Coordination Ecosystem (ACCESS-CI), Texas Advanced Computing Center, and the JetStream2 scientific cloud - public computational resources supported by NSF. usegalaxy.eu is supported by the German Federal Ministry of Education and Research grant 031L0101C and de.NBI-epi. usegalaxy.org.au forms part of the national Australian BioCommons infrastructure and is funded by Bioplatforms Australia.

# Maroon Galaxy

- Try Galaxy at [usegalaxy.org](http://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
    - [hprc.tamu.edu/apply](http://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
  - [hprc.tamu.edu/kb/User-Guides/Galaxy](http://hprc.tamu.edu/kb/User-Guides/Galaxy)
- There are no backups of users' Galaxy files

# HPRC Maroon Galaxy

**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 96 hours after upload. Please do not delete the same day as you upload to ftp

Galaxy data privacy [not](#)  
If you have restricted access data, then you should  
Contact the HPRC helpdesk for an alternative

COVID-19 related research on Galaxy: [training](#)

**Current known issues:**

- Tools will fail when parentheses are in the input file names.
  - Click the pencil icon to access Edit attributes to rename a file and remove the parentheses.
- Some tools cannot be removed from favorites. Choose your favorites wisely.

Take an interactive tour: [Galaxy UI History Scratchbook](#)

platform for supporting data intensive research. Galaxy is developed by [The Galaxy Team](#) with the support of [many](#)

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

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

# HPRC Galaxy Documentation Knowledge Base

**TAMU Texas A&M HPRC** Search

Home User Guides Software Helpful Pages FAQ

**User Guides**

- ACES >
- FASTER >
- Grace >
- Launch >
- Portal >
- Galaxy >
- LMS
- AMS

## Galaxy

### Maroon Galaxy Accounts

Maroon Galaxy (v24.2) 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](https://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](https://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.

Send an email to [help@hprc.tamu.edu](mailto:help@hprc.tamu.edu) requesting

**Table of contents**

- Maroon Galaxy Accounts
- Set Your Default Account
- Data Security
- Account Security
- Permanently Delete unwanted files
- Download all files in a history
- Uploading Files > 2GB via FTP to Maroon Galaxy
  - From a Unix Computer (Mac or Linux)
  - From one of your directories on Grace
  - Using Bitvise
  - Using MobaXterm
  - Using Filezilla
  - Using WinSCP
- Downloading Files > 2GB via FTP from Maroon Galaxy to your desktop
  - Stage files to copy

<https://hprc.tamu.edu/kb/User-Guides/Galaxy>

# Check Your HPRC SUs Balance

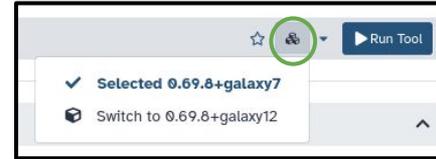
The screenshot shows the Galaxy HPRC Maroon interface. The left sidebar contains navigation options: Upload, Tools, History Divider, My HPRC SU Balance, Get Data, Send Data, Collection Operations, Lift-Over, Text Manipulation, Convert Formats, Filter and Sort, Join, Subtract and Group, Fetch Alignments/Sequences, Operate on Genomic Intervals, Statistics, Graph/Display Data, NGS: QC and manipulation, RNA-seq, and HUMAnN. The main panel displays the 'My HPRC SU Balance' tool parameters. Under 'I want to', the option 'Show my current SU balance' is selected. Below this, there are sections for 'Job Resource Parameters' with input fields for Memory (7 GB) and Time (24 hours). The right sidebar shows the 'History' section with a table of job records.

Project/Account	JobID	JobArrayIndex
132787212185	16469166	0
132787212185	16469694	0

- 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 or you are out of SUs

# Galaxy Features

1. You can easily switch the software version



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

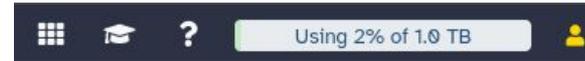


3. Disk quota enforced for all users

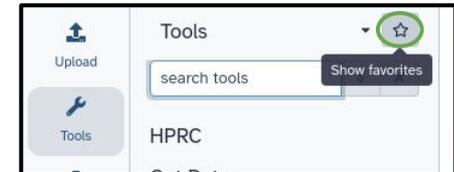
a. all users begin with 1 TB disk quota

b. request an increase if needed but [permanently delete](#) nonessential files first

c. compressed (gzipped) file format is supported for uploads but not as input files for all tools



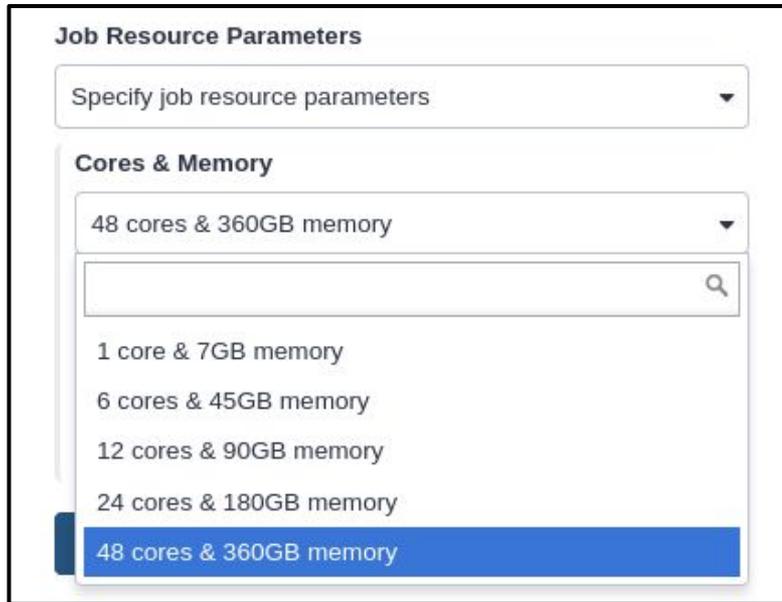
4. You can add tools to your favorites to easily locate



# Galaxy Features cont.

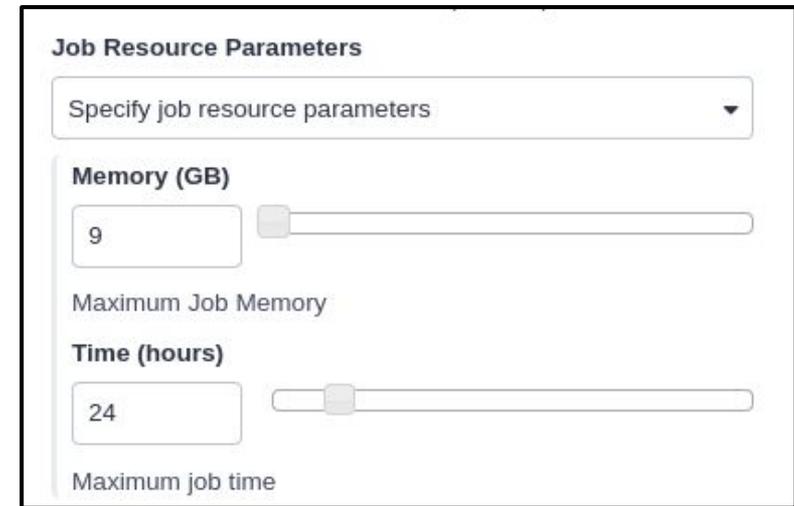
## 5. Job Resource Parameters available for cores, memory and job time

tools supporting multi-core processing



The screenshot shows the 'Job Resource Parameters' section of a Galaxy tool interface. At the top, there is a dropdown menu labeled 'Specify job resource parameters'. Below it, the 'Cores & Memory' section is expanded, showing a dropdown menu with the selected option '48 cores & 360GB memory'. A search bar with a magnifying glass icon is positioned below the dropdown. A list of five options is displayed: '1 core & 7GB memory', '6 cores & 45GB memory', '12 cores & 90GB memory', '24 cores & 180GB memory', and '48 cores & 360GB memory'. The last option is highlighted with a blue background.

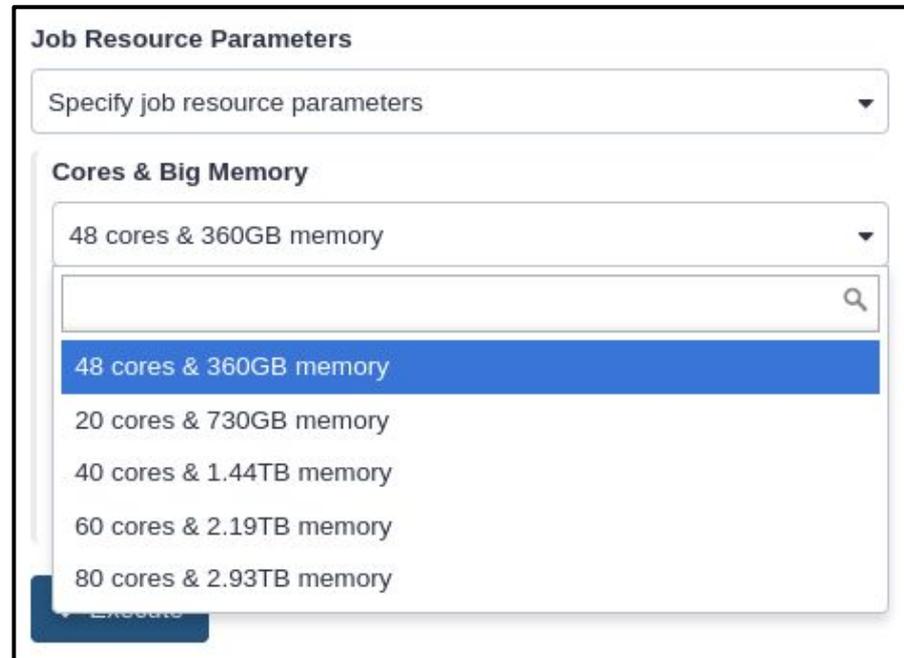
tools supporting single-core processing



The screenshot shows the 'Job Resource Parameters' section of a Galaxy tool interface. At the top, there is a dropdown menu labeled 'Specify job resource parameters'. Below it, the 'Memory (GB)' section is expanded, showing a text input field with the value '9' and a horizontal slider control. Below this, the text 'Maximum Job Memory' is displayed. The 'Time (hours)' section is also expanded, showing a text input field with the value '24' and a horizontal slider control. Below this, the text 'Maximum job time' is displayed.

# Galaxy Features cont.

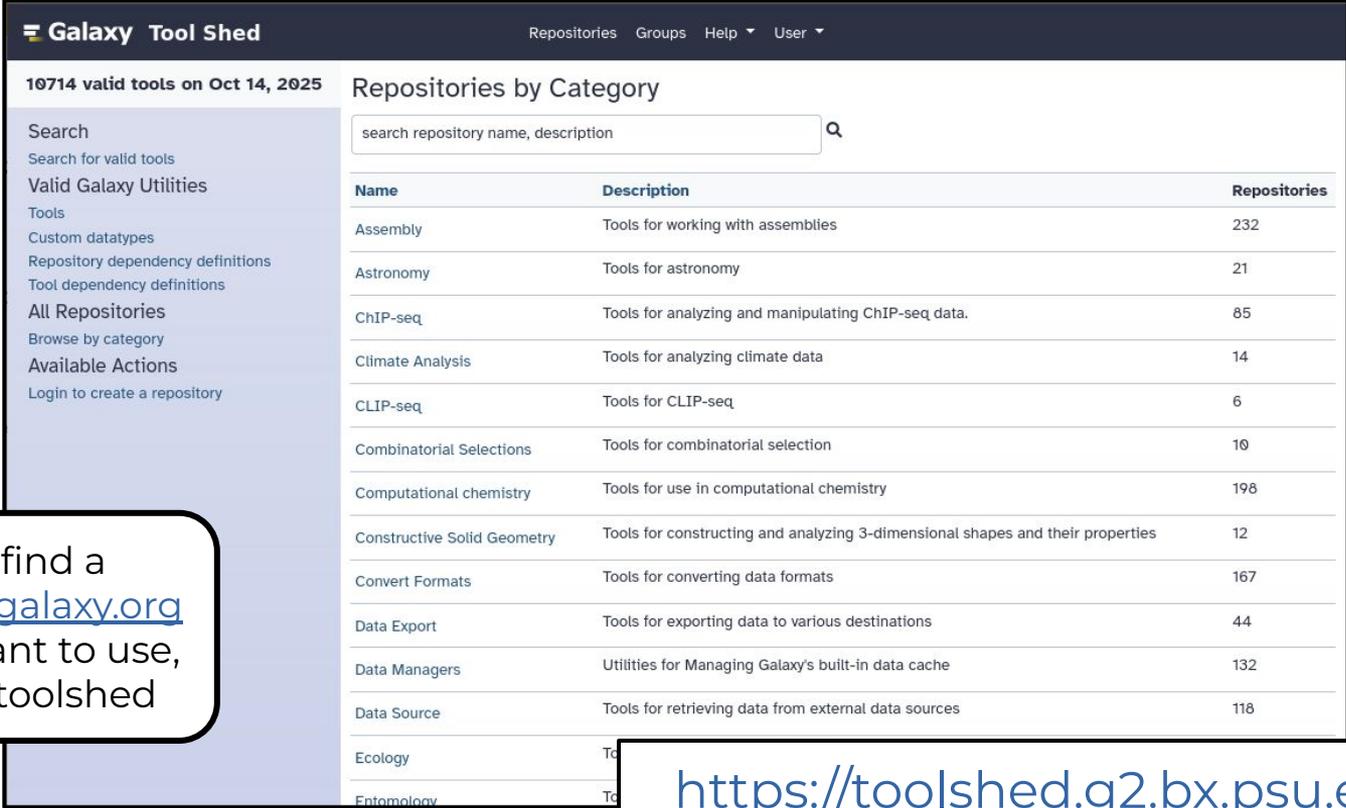
6. Job Resource Parameters available for cores, big memory and job time tools supporting multi-core and requiring big memory processing



The screenshot displays the 'Job Resource Parameters' section of a Galaxy interface. It features a dropdown menu titled 'Specify job resource parameters' which is currently open to show options under the 'Cores & Big Memory' category. The selected option is '48 cores & 360GB memory'. Other visible options include '20 cores & 730GB memory', '40 cores & 1.44TB memory', '60 cores & 2.19TB memory', and '80 cores & 2.93TB memory'. A search icon is present in the dropdown menu, and a blue 'Execute' button is partially visible at the bottom left.

Job Resource Parameters
Specify job resource parameters
<b>Cores &amp; Big Memory</b>
48 cores & 360GB memory
20 cores & 730GB memory
40 cores & 1.44TB memory
60 cores & 2.19TB memory
80 cores & 2.93TB memory

# Look for Additionally Available Tools



The screenshot shows the Galaxy Tool Shed interface. At the top, it says "Galaxy Tool Shed" and "10714 valid tools on Oct 14, 2025". The main content is titled "Repositories by Category" and includes a search bar. Below the search bar is a table with three columns: Name, Description, and Repositories. The table lists various categories and their corresponding tool counts.

Name	Description	Repositories
Assembly	Tools for working with assemblies	232
Astronomy	Tools for astronomy	21
ChIP-seq	Tools for analyzing and manipulating ChIP-seq data.	85
Climate Analysis	Tools for analyzing climate data	14
CLIP-seq	Tools for CLIP-seq	6
Combinatorial Selections	Tools for combinatorial selection	10
Computational chemistry	Tools for use in computational chemistry	198
Constructive Solid Geometry	Tools for constructing and analyzing 3-dimensional shapes and their properties	12
Convert Formats	Tools for converting data formats	167
Data Export	Tools for exporting data to various destinations	44
Data Managers	Utilities for Managing Galaxy's built-in data cache	132
Data Source	Tools for retrieving data from external data sources	118
Ecology	Tools for ecology	1
Entomology	Tools for entomology	1

If you can't find a tool on [usegalaxy.org](https://usegalaxy.org) that you want to use, search the toolshed

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

# 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, ...

# Copy a File Between Histories

The screenshot displays the Galaxy HPRC Maroon Galaxy interface. At the top right, a status bar indicates "Using 2% of 1.0 TB". The main area is titled "History Multiview" and shows a grid of history items. A sidebar on the left contains navigation options like "Upload", "Tools", "Workflows", "Workflow Invocations", "Histories", "History Multiview", "Datasets", "Pages", and "Libraries". A right-hand panel titled "History" shows a list of 49 histories and options to manage them.

**History disk usage**: A callout points to the "Using 2% of 1.0 TB" status bar.

**multiview**: A callout points to the "History Multiview" sidebar option.

**Drag and drop files from one history to another to copy a file without duplicating the original file and increasing your disk usage**: A callout points to a file in the multiview grid.

**Total disk usage. Click to see details**: A callout points to the "Using 2% of 1.0 TB" status bar.

**The History disk usage reflects an increase based on the file size but your overall disk usage does not increase**: A callout points to the "Using 2% of 1.0 TB" status bar.

# 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/kb/User-Guides/Galaxy/#uploading-files-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 after they are imported into your history

Galaxy HPRC Maroon Galaxy

Upload from Disk or Web to **mouse GRCm39**

Regular Composite Collection Rule-based

search

This Galaxy server allows you to upload files via FTP. To upload some files, log in to the FTP server at using your Galaxy credentials. For help visit the [tutorial](#).

<input checked="" type="checkbox"/>	Label	Time
<input checked="" type="checkbox"/>	mouse.fa	Oct 20, 2025, 10:05 AM

Back Cancel Select

Type (select)  Reference (set all):

Choose local file Choose remote files Paste/Fetch data Start Pause Reset Close

- 1: Upload button
- 2: Choose remote files button
- 3: Table containing file information
- 4: Start button

# FTP Upload File Directly to History via URL

Galaxy HPRC Maroon Galaxy

Upload from Disk or Web to **mouse GRCh39**

Regular Composite Collection Rule-based

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

New File 0 b Auto-detect unspecified (?) 0%

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

`https://hgdownload.soe.ucsc.edu/goldenPath/wuhCor1/bigZips/chromFa.tar.gz`

Type (set all): Auto-detect Reference (set all): unspecified (?)

Choose local file Choose remote files Paste/Fetch data Start Pause Reset Close

1

2

3

4

paste ftp URL

history

search datasets

mouse GRCh39

59 GB

GRCh39\_genomic.gff

100,000 lines

format gff3, database ?

loaded gff3 file

Seqid

gff-version 3

gff-spec-version 1.21

processor NCBI annotwriter

genome-build GRCh39

genome-build-accession NCBI\_Assembly:GCF\_8

You can URL upload small or 2GB+ sized file if the URL is an ftp or http(s) site

# Get Data Exercise

1. Create or rename a new history with the name SRA-upload
2. Use the "Faster Download and Extract Reads in FASTQ NCBI SRA" tool to download sequence reads for accession **SRR31754396** to your history

The screenshot displays the Galaxy HPRC Maroon Galaxy interface. The top navigation bar shows 'Galaxy HPRC Maroon Galaxy' and 'Using 3% of 1.0 TB'. The left sidebar contains navigation options: Upload, Tools, Workflows, Workflow Invocations, and Visualization. The main content area is divided into three sections:

- Tools:** A search bar contains the text 'faster'. Below it is a 'Show Sections' button. The tool 'Faster Download and Extract Reads in FASTQ format from NCBI SRA' is selected.
- Tool Parameters:**
  - select input type:** A dropdown menu is set to 'SRR accession'.
  - Accession \*:** A text input field contains 'SRR31754396'. Below it is a note: 'Must start with SRR, DRR or ERR, e.g. SRR925743, ERR343809'.
- History:** A search bar for datasets is present. Below it, a history named 'SRA-upload' is shown, which is currently empty. A message states: 'This history is empty. You can load your own data or get data from an external source.'

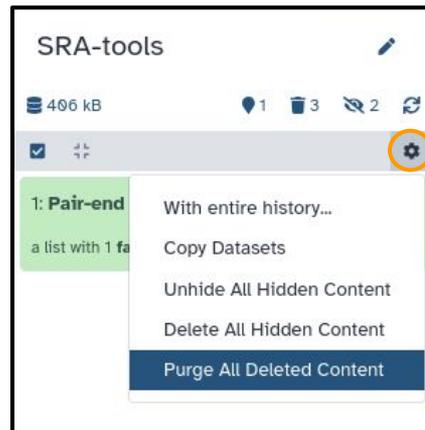
# Permanently Delete Nonessential Files

purge files from disk and reduce your Galaxy disk usage

Deleted non-essential files



Purge deleted content



# Run FastQC on List of Paired Files

**Tools** ▼ ☆

fastqc ▼ ✕

Show Sections

**FastQC** Read Quality reports

**Make.fastq** Convert fasta and quality to fastq

**FastQC** Read Quality reports (Galaxy Version 0.74+galaxy1) ☆ 🔗 ▼ ▶ Run Tool

**Tool Parameters**

**Raw read data from your current history \***

📄 📄 📁 ⋮ 1: Pair-end data (fasterq-dump) ▼

accepted formats ▼

! The supplied input will be mapped over this tool.

**Contaminant list** - optional

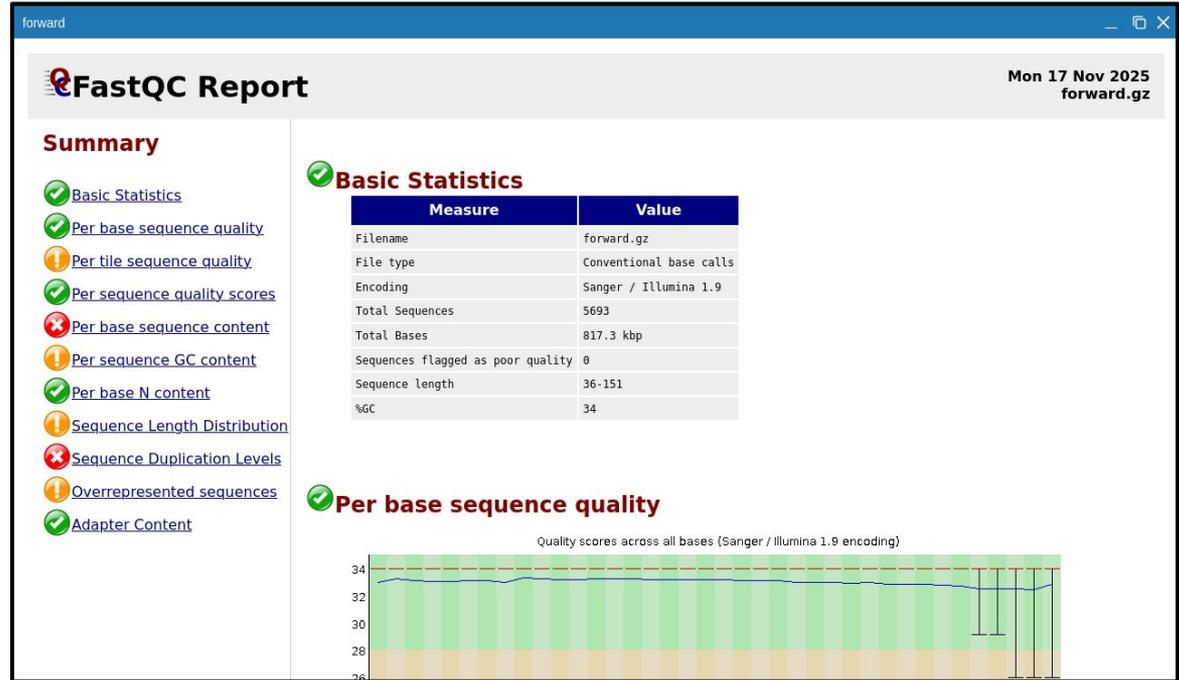
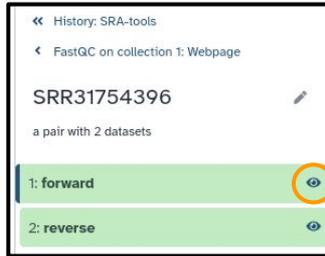
📄 📄 📁 ⋮ Nothing selected ▼

accepted formats ▼

tab delimited file with 2 columns: name and sequence. For example: Illumina Small RNA RT  
Primer CAAGCAGAAGACGGCATACGA

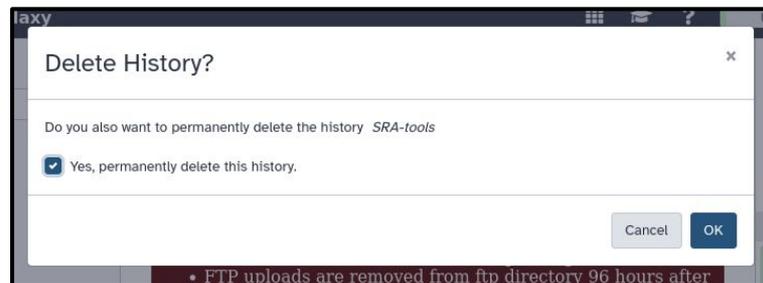
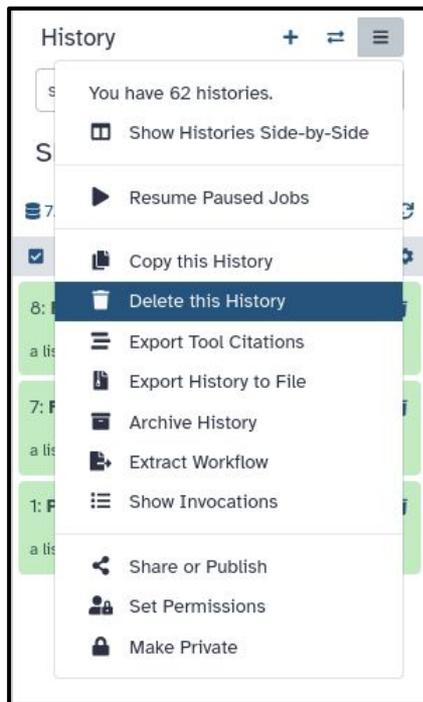
# Review FastQC Output

- Select "FastQC on collection 1: Webpage"
- SRR31754396
- Enable Window Manager
- Click the eye icon for the forward reads



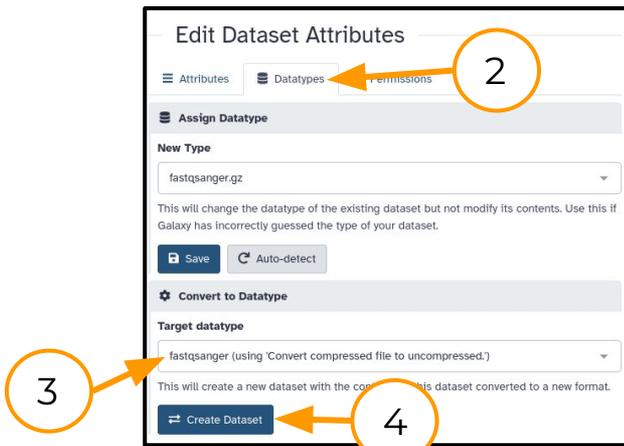
# Permanently Delete History

purge history from disk to reduce disk space

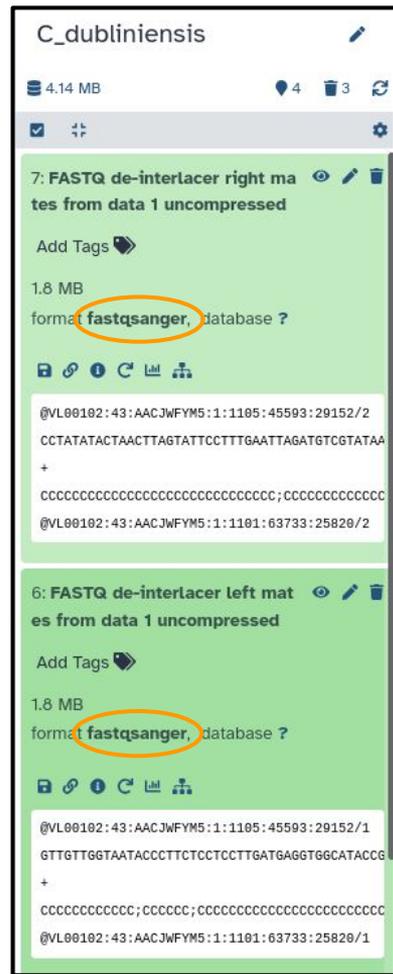


# Converting and Using gzipped Files

- 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 in order not to have duplicated data that takes up disk space

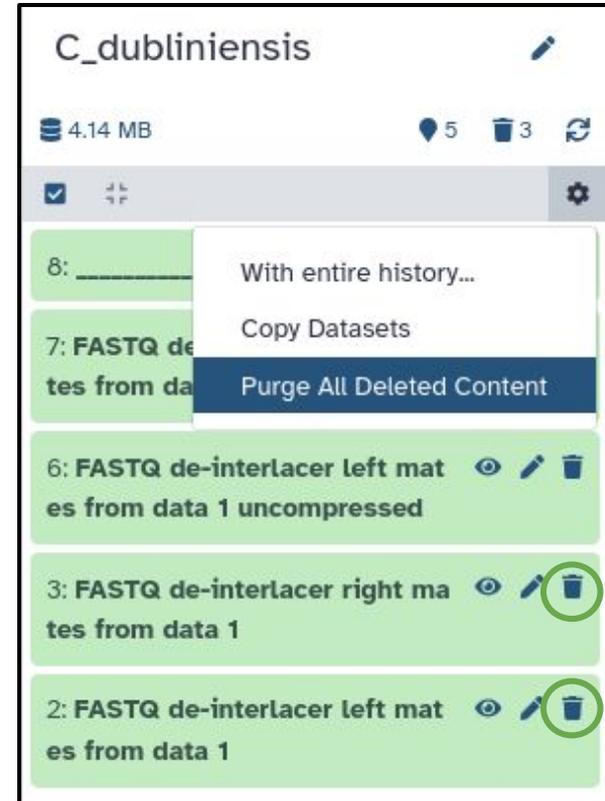


Convert the left and right compressed fastqsanger.gz files to uncompressed fastqsanger



# Delete and Purge the Compressed Files

- Permanently delete (purge) both compressed fastqsanger.gz files after the decompressed fastqsanger files have been created



# History Divider

The screenshot displays the Galaxy web interface for 'HPRC Maroon Galaxy'. The left sidebar contains navigation options: Upload, Tools, Workflows, Workflow Invocations, Visualization, Histories, History Multiview, and Datasets. The 'Tools' section is expanded, showing a search bar and a list of tool categories including HPRC, My HPRC SU Balance, Get Data, Send Data, Collection Operations, Lift-Over, Text Manipulation, Convert Formats, Filter and Sort, Join, Subtract and Gro, Fetch Alignments/Sequences, Operate on Genomic Intervals, and Statistics. The 'History Divider' tool is highlighted with a green arrow. The main panel shows the 'History Divider' tool configuration, including a 'Run Tool' button, 'Tool Parameters' (with a 'spacer character' field set to '-'), and 'Job Resource Parameters'. The right panel shows the 'History' section with a search bar and a list of datasets, including 'C\_dublinsiensis' and 'FASTQ de-interlacer' outputs.

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

Run the History Divider tool between jobs

# Shared Libraries

The screenshot shows the Galaxy HPRC Maroon Galaxy interface. On the left sidebar, the 'Libraries' icon is circled in green. In the main content area, a table lists data libraries. The 'genomes' entry in the table is also circled in green. A green arrow points from the 'genomes' entry to the right-hand screenshot.

Name	Description	Synopsis
bacteria example files	E. coli	example files: fq, fa and gff for tes ... (more)
genomes		

The screenshot shows the Galaxy HPRC Maroon Galaxy interface with the 'genomes' library selected. The main content area displays a table of data libraries within the 'genomes' folder.

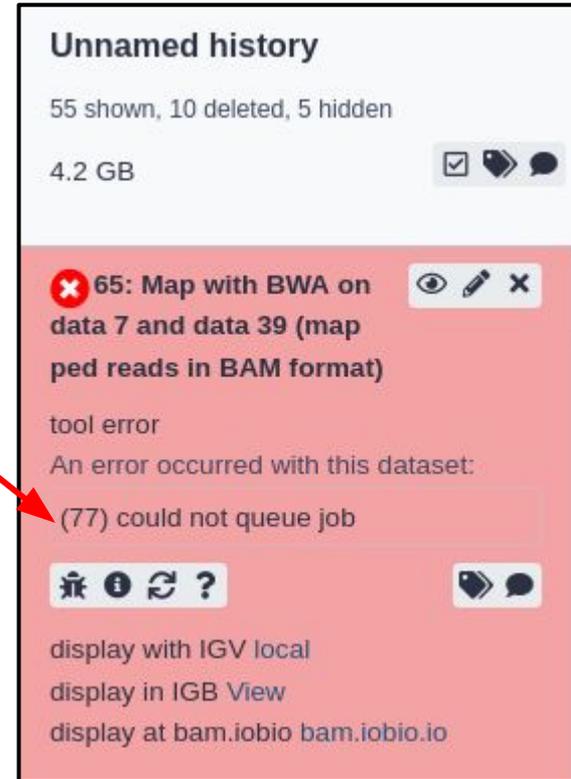
Name	Description	Type	Size
AaegL5_2	AaegL5_2 (mosquito) Source version(s): ... (more)	folder	
bosTau9	cow assembly (ARS-UCD1.2) files: fasta a ... (more)	folder	
Glycine_max_v4.0	G. max (soybean) files: gtf, gff	folder	
GRCm39	mouse assembly files: fasta and gff	folder	
GRCr8	Rattus norvegicus (Norway rat)	folder	

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 Job: could not 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:

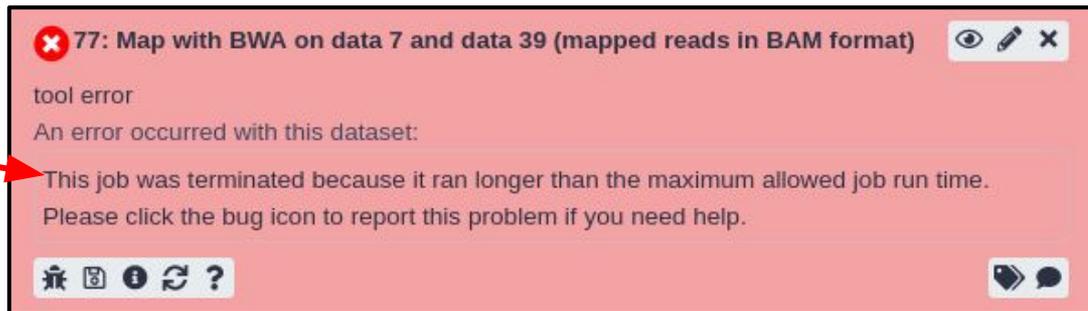
(77) could not queue job

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.

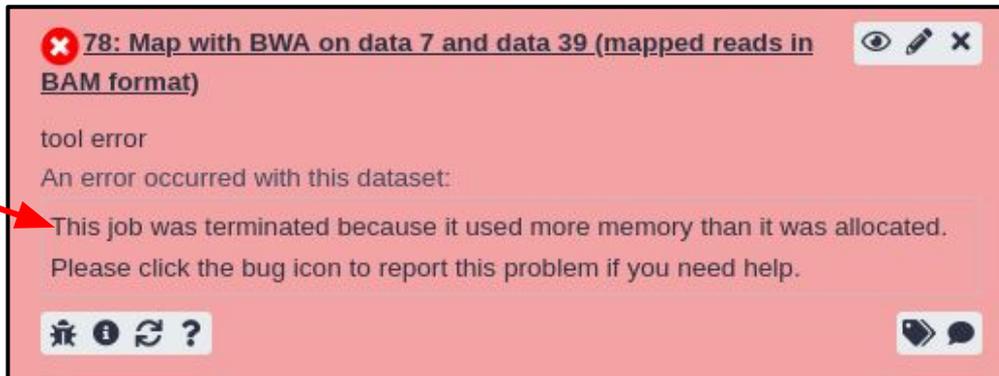
- Configure the job to use more Time

**Time (hours)**



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

- History name
- History item number

The screenshot shows the Galaxy HPRC Maroon interface. The top navigation bar includes the Galaxy logo, the text 'HPRC Maroon Galaxy', and a storage indicator 'Using 2% of 1.0 TB'. The main content area displays a job log for a failed job. The log text includes:  
java -jar \$EBROOTPICARD/picard.jar  
/jobs\_directory/044/44939/tool\_script.sh: line 9: cd: /tmp/job.16690611/  
directory  
d error:  
last):  
odule>  
a import set\_metadata; set\_metadata()  
laxy/lib/galaxy\_ext/metadata/set\_metadata.py", line 20, in <module>  
port set\_metadata  
laxy/lib/galaxy/metadata/\_\_init\_\_.py", line 9, in <module>  
galaxy/lib/galaxy/model/\_\_init\_\_.py", line 623  
None)) is not None:  
SyntaxError: invalid syntax

The right-hand 'History' panel shows a list of datasets. The selected dataset is '110: phix.fa'. Below it, an error message is displayed in a red box, with a bug icon (1) and an information icon (1) next to it. The error message reads:  
An error occurred with this dataset:  
format html, database ?  
[1] "en\_US.UTF-8"  
Error in getopt(spec):  
["(Stroked:": is not a valid option, or does

Below the error message, there are two green boxes representing output files: '80: DEXSeq rds file on data 21, data 20, and others' (2) and '108: DEXSeq result file on data 21, data 20, and others' (3).

# Refresh Web Browser

The disk image download button  is missing

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



History

search datasets

C\_dub

1.07 GB

28

6



red refresh icon in history

# Galaxy Videos

<https://vimeo.com/galaxyproject>

The screenshot displays the Vimeo interface for the Galaxy Project channel. At the top, the Vimeo navigation bar includes the logo, menu items like 'Why Vimeo?', 'Features', 'Resources', 'Watch', 'Pricing', 'Log in', and buttons for 'Join' and 'New video'. The main content area shows a grid of Galaxy workflow diagrams with various tool icons and data flow arrows. Below the grid, a video player is visible, and a list of 154 videos is shown. Two video thumbnails are highlighted: 'Galaxy COVID-19: Variation analysis' and 'Confound it! Reproducible biology from ...'. A callout box on the right contains the text: 'There are many videos available on various Galaxy topics'.

# Sequence Variant Calling Exercise

# Prep Reads for Alignment to Reference Genome

1. Create a history named C\_dublinsiensis
2. Import sequence reads and genome build from Libraries
3. Run FastQC on the paired end sequence read files
  - review the results
4. Trim reads with the cutadapt tool
5. Delete the un-trimmed input files after cutadapt has completed

# Import Reference Genome

- Import *Candida dubliniensis* fasta reference file from "Libraries"
  - *C. dubliniensis* is a diploid fungal genome

Galaxy HPRC Maroon Galaxy

Using 3% of 1.0 TB

Upload  
Tools  
Workflows  
Workflow Invocations  
Visualization  
Histories  
History Multiview  
Datasets  
Pages  
Libraries

Search + Folder + Datasets Add to History Download Delete Details

include deleted

Libraries / genomes / C. dubliniensis

as Datasets  
as a Collection

<input checked="" type="checkbox"/>	Name	Description
<input checked="" type="checkbox"/>	C_dubliniensis_CD36_current_features.gtf	uploaded gtf file
<input checked="" type="checkbox"/>	C_dubliniensis_CD36.fasta	uploaded fasta file
<input checked="" type="checkbox"/>	DR34_R1.fastq.gz	uploaded fastqsanger.gz file
<input checked="" type="checkbox"/>	DR34_R2.fastq.gz	uploaded fastqsanger.gz file

« < 1 > » 10 per page, 4 total

History  
search datasets  
C\_dubliniensis  
This history is empty.  
You can load your own data or get data from an external source.

# Trim Adapters with Cutadapt

The screenshot displays the Galaxy web interface for the Cutadapt tool. The top navigation bar shows 'Galaxy HPRC Maroon Galaxy' and 'Using 3% of 1.0 TB'. The left sidebar contains navigation icons for Upload, Tools, Workflows, Workflow Invocations, Visualization, Histories, and History.

The main content area shows the 'Cutadapt Remove adapter sequences from FASTQ/FASTA' tool configuration. The tool parameters are as follows:

- Single-end or Paired-end reads?**: Paired-end
- FASTQ/A file #1 \***: 3: DR34\_R1.fastq.gz
- FASTQ/A file #2 \***: 4: DR34\_R2.fastq.gz

The right sidebar shows the 'History' panel with a search bar and a list of datasets:

- 379 MB
- 4 datasets
- 4: DR34\_R2.fastq.gz
- 3: DR34\_R1.fastq.gz
- 2: C\_dublinsiensis\_CD36.fasta
- 1: C\_dublinsiensis\_CD36\_current\_features.gtf

# Align Sequences to Reference Genome

- Align sequence reads using bwa-mem to the *Candida dubliniensis* reference fasta file
  - select "Use a genome from history and build index"
  - use 24 cores & 180GB
  - (use all cores and memory for large genomes)

The screenshot displays the Galaxy web interface for the 'Map with BWA-MEM' tool. The tool parameters are configured as follows:

- Tool Parameters:** Will you select a reference genome from your history or use a built-in index? Use a genome from history and build index.
- Use the following dataset as the reference sequence:** 2: C\_dubliniensis\_CD36.fasta
- Algorithm for constructing the BWT index:** Auto. Let BWA decide the best algorithm to use.
- Single or Paired-end reads:** Paired
- Select first set of reads:** 5: Cutadapt on data 4 and data 3: Read 1 Output
- Select second set of reads:** 6: Cutadapt on data 4 and data 3: Read 2 Output

The history panel on the right shows the dataset 'C\_dubliniensis' with a size of 783 MB and several output files:

- 6: Cutadapt on data 4 and data 3: Read 2 Output
- 5: Cutadapt on data 4 and data 3: Read 1 Output
- 4: DR34\_R2.fastq.gz
- 3: DR34\_R1.fastq.gz
- 2: C\_dubliniensis\_CD36.fasta
- 1: C\_dubliniensis\_CD36\_current\_features.gtf

# Run Samtools flagstat on the .bam Alignment File

- Run the Samtools flagstat tool on the alignment .bam file
  - keep Memory at 7GB and change Time to 1 hour
  - review the alignment percentage

The screenshot displays the Galaxy web interface for running the Samtools flagstat tool. The interface is divided into three main sections:

- Left Panel (Tools):** A sidebar with navigation icons and a search bar. The search results show several tools, with "Samtools flagstat" selected and highlighted in light blue. Other tools listed include "Samtools stats", "Samtools sort", "Samtools view", and "Samtools idxstats".
- Center Panel (Tool Parameters):** The configuration page for "Samtools flagstat". It includes:
  - Tool Parameters:** A section for "BAM File to report statistics of" with a dropdown menu showing "7: Map with BWA-MEM on data 6, data 5, and data 2 (mapped reads in BAM format)".
  - Output format:** A dropdown menu set to "txt".
  - Job Resource Parameters:** A section for "Memory (GB)" set to 7 and "Time (hours)" set to 1. Both have sliders and input fields.
  - Run Tool:** A blue button at the bottom of the configuration section.
- Right Panel (History):** A list of previous jobs. The top job is highlighted in orange and matches the current tool configuration: "7: Map with BWA-MEM on data 6, data 5, and data 2 (mapped reads in BAM format)". Below it are other jobs like "6: Cutadapt on data 4 and data 3: Read 2 Output", "5: Cutadapt on data 4 and data 3: Read 1 Output", "4: DR34\_R2.fastq.gz", "3: DR34\_R1.fastq.gz", "2: C\_dubliniensis\_CD36.fasta", and "1: C\_dubliniensis\_CD36\_current\_features.gtf".

# Call Variants with FreeBayes

Use the reference genome fasta file from your history

The screenshot shows the Galaxy web interface for the FreeBayes tool. The tool parameters are configured as follows:

- Tool Parameters:** FreeBayes bayesian genetic variant detector (Galaxy Version 1.3.10+galaxy1)
- Choose the source for the reference genome:** History
- Run in batch mode?:** Run individually (selected)
- BAM or CRAM dataset \*:** 7: Map with BWA-MEM on data 6, data 5, and data 2 (mapped reads in BAM format)
- Use the following dataset as the reference sequence \*:** 2: C\_dublinsiensis\_CD36.fasta
- Limit variant calling to a set of regions?:** Do not limit
- Read coverage:** Use defaults

The History panel on the right shows a list of datasets, with 'C\_dublinsiensis' at the top.

Select 48 cores & 360GB memory

The Job Resource Parameters section is configured as follows:

- Specify job resource parameters:** (Dropdown menu)
- Cores & Memory \*:** 48 cores & 360GB memory
- Time (hours) \*:** 24
- Maximum job time:** (Slider)
- Run Tool:** (Button)

# Build a snpEff Database

The screenshot displays the Galaxy web interface for the 'SnpEff build' tool. The interface is organized into several key sections:

- Left Sidebar:** Contains navigation icons for Upload, Tools, Workflows, Workflow Invocations, Visualization, Histories, History Multiview, Datasets, Pages, Libraries, and Notifications.
- Top Navigation:** Shows the tool name 'SnpEff build: database from Genbank or GFF record (Galaxy Version 4.3+T.galaxy6)' and a 'Run Tool' button.
- Tool Parameters:**
  - Name for the database:** A text input field containing 'c\_dub\_CD36'.
  - Input annotations are in:** Radio buttons for GenBank, GFF, and GTF (selected).
  - GTF dataset to build database from:** A dropdown menu showing '1: C\_dubliniensis\_CD36\_current\_features.gtf'.
  - Choose the source for the reference genome:** A dropdown menu showing 'History'.
  - Genome in FASTA format:** A dropdown menu showing '2: C\_dubliniensis\_CD36.fasta'.
  - Select genetic code for this sequence:** A dropdown menu showing 'Standard'.
- Right Panel (History):** Titled 'History', it shows a search bar and a list of previous tool runs. The top entry is 'Copy of 'C\_dubliniensis'' with a size of 987 MB. Below it are several other runs, including 'FreeBayes on data 2 and data 7 (variants)', 'Samtools flagstat on data 7', 'Map with BWA-MEM on data 6, data 5, and data 2 (mapped reads in BAM format)', 'Cutadapt on data 4 and data 3: Read 2 Output', 'Cutadapt on data 4 and data 3: Read 1 Output', 'DR34\_R2.fastq.gz', 'DR34\_R1.fastq.gz', 'C\_dubliniensis\_CD36.fasta', and 'C\_dubliniensis\_CD36\_current\_features.gtf'.

# Annotate Variants with snpEff

**Tools**

Upload

Tools

Workflows

Workflow Invocations

Visualization

Histories

History Multiview

Datasets

Pages

Libraries

**Tools**

snpeff

Show Sections

**SnpEff chromosome-info:** list chromosome names/lengths

**SnpEff build:** database from Genbank or GFF record

**SnpEff eff:** annotate variants

**SnpEff download:** download a pre-built database

**SnpEff databases:** list available databases

**SnpEff build:** database from Genbank or GFF record

**SnpEff eff:** annotate variants

**SnpEff eff: annotate variants (Galaxy Version 4.3+T.galaxy2)**

Run Tool

**Tool Parameters**

Sequence changes (SNPs, MNPs, InDels) \*

9: FreeBayes on data 2 and data 7 (variants)

accepted formats

**Input format \***

VCF

**Output format**

VCF (only if input is VCF)

**Create CSV report, useful for downstream analysis (-csvStats)**

No

**Genome source**

Custom snpEff database in your history

! Parameter 'snpeff\_db': This version of SnpEff will only work with SnpEff4.3 genome databases

**SnpEff4.3 Genome Data \***

10: SnpEff4.3 database for c\_dub\_CD36

accepted formats

**Select genetic code for this sequence \***

Standard

If this sequence uses non-standard genetic code, select one from these options

**History**

search datasets

Copy of 'C\_dubliniensis'

987 MB

10

10: SnpEff4.3 database for c\_dub\_CD36

9: FreeBayes on data 2 and data 7 (variants)

8: Samtools flagstat on data 7

7: Map with BWA-MEM on data 6, data 5, and data 2 (mapped reads in BAM format)

6: Cutadapt on data 4 and data 3: Read 2 Output

5: Cutadapt on data 4 and data 3: Read 1 Output

4: DR34\_R2.fastq.gz

3: DR34\_R1.fastq.gz

2: C\_dubliniensis\_CD36.fasta

1: C\_dubliniensis\_CD36\_current\_features.gtf

# View snpEff Results

Galaxy HPRC Maroon Galaxy

Using 3% of 1.0 TB

Variant rate 1 variant every 263 bases

Variants rate details

Chromosome	Length	Variants	Variants rate
1_C_dubliniensis_CD36	3,214,061	9,456	339
2_C_dubliniensis_CD36	2,289,089	9,581	238
3_C_dubliniensis_CD36	1,863,824	7,253	256
4_C_dubliniensis_CD36	1,641,709	6,469	253
5_C_dubliniensis_CD36	1,245,899	6,239	199
6_C_dubliniensis_CD36	1,073,895	4,749	226
7_C_dubliniensis_CD36	1,022,435	3,837	266
R_C_dubliniensis_CD36	2,267,510	7,815	290
<b>Total</b>	<b>14,618,422</b>	<b>55,399</b>	<b>263</b>

Fetch Alignments/Sequences  
Operate on Genomic Intervals  
Statistics  
Graph/Display Data  
NGS: QC and manipulation  
RNA-seq  
HUMAN  
NGSEP  
Phenotype Association  
Mapping  
RNA-seq  
Variant Calling  
Cheminformatics  
NCBI BLAST+  
FASTQ Quality Control  
Assembly

Snpeff: on data 10 and data 9

```
##fileformat=VCFv4.2
##fileDate=20251117
##source=freeBayes v1.3.6
##reference=localref.fa
##contig=<ID=Chr1_C_dubliniensis_CD36,length=3214061>
##contig=<ID=Chr2_C_dubliniensis_CD36,length=2289089>
##contig=<ID=Chr3_C_dubliniensis_CD36,length=1863824>
##contig=<ID=Chr4_C_dubliniensis_CD36,length=1641709>
##contig=<ID=Chr5_C_dubliniensis_CD36,length=1245899>
##contig=<ID=Chr6_C_dubliniensis_CD36,length=1073895>
##contig=<ID=Chr7_C_dubliniensis_CD36,length=1022435>
##contig=<ID=ChrR_C_dubliniensis_CD36,length=2267510>
##phasing=none
##commandline="freebayes --region ChrR_C_dubliniensis_CD36:9.2267510 --bam b_9.bam --fasta-reference localref.fa --vcf ./vcf_output/part_ChrR_C_dubliniensis_C
```

# Extract the Workflow

History

- You have 63 histories.
- Show Histories Side-by-Side
- Resume Paused Jobs
- Copy this History
- Delete this History
- Export Tool Citations
- Export History to File
- Archive History
- Extract Workflow**
- Show Invocations
- Share or Publish
- Set Permissions
- Make Private

7: Map with BWA-MEM on data 6, data 5, and data 2 (mapped reads in BAM format)

6: Cutadapt on data 4 and data 3: Read 2 Output

5: Cutadapt on data 4 and data 3: Read 1 Output

4: DR34\_R2.fastq.gz

3: DR34\_R1.fastq.gz

2: C\_dublinsiensis\_CD36.fasta

1: C\_dublinsiensis\_CD36\_current\_features.gtf

Tools

search tools

Workflow name: variant calling with snpEff

Create Workflow Check all Uncheck all

Tool	History items created
Import from History <i>This tool cannot be used in workflows</i>	1 C_dublinsiensis_CD36_current_features.gtf <input checked="" type="checkbox"/> Treat as input dataset C_dublinsiensis_CD36_current_fe
Import from History <i>This tool cannot be used in workflows</i>	2 C_dublinsiensis_CD36.fasta <input checked="" type="checkbox"/> Treat as input dataset C_dublinsiensis_CD36.fasta
Import from History <i>This tool cannot be used in workflows</i>	3 DR34_R1.fastq.gz <input checked="" type="checkbox"/> Treat as input dataset DR34_R1.fastq.gz
Import from History <i>This tool cannot be used in workflows</i>	4 DR34_R2.fastq.gz <input checked="" type="checkbox"/> Treat as input dataset DR34_R2.fastq.gz
Cutadapt <input checked="" type="checkbox"/> Include "Cutadapt" in workflow	5 Cutadapt on data 4 and data 3: Read 1 Output
Map with BWA-MEM <input checked="" type="checkbox"/> Include "Map with BWA-MEM" in workflow	6 Cutadapt on data 4 and data 3: Read 2 Output

The following list contains each tool that was run to create the datasets in your current history. Please select those that you wish to include in the workflow.

Tools which cannot be run interactively and thus cannot be incorporated into a workflow will be shown in gray.

History

search datasets

C\_dublinsiensis

106 GB

- 12: SnpEff eff: on data 10 and data 9 - HTML stats
- 11: SnpEff eff: on data 10 and data 9
- 10: SnpEff4.3 database for C\_dublinsiensis\_CD36\_current\_fe
- 9: FreeBayes on data 2 and data 7 (variants)
- 8: Samtools flagstat on data 7
- 7: Map with BWA-MEM on data 6, data 5, and data 2 (mapped reads in BAM format)
- 6: Cutadapt on data 4 and data 3: Read 2 Output
- 5: Cutadapt on data 4 and data 3: Read 1 Output
- 4: DR34\_R2.fastq.gz
- 3: DR34\_R1.fastq.gz
- 2: C\_dublinsiensis\_CD36.fasta
- 1: C\_dublinsiensis\_CD36\_current\_features.gtf

Workflow "variant calling with snpEff" created from current history. You can edit or run the workflow.

# Review the snpEff Workflow

Workflows

My workflows

Workflows shared with me

Public workflows

Search my workflows by query or use the advanced search

Sort by: Name Update time Filter: [ ]

edited 2 minutes ago

**variant calling with snpEff**

Add Tags

**Workflow Preview**

1: C:\dublinensis\_CD36\_current\_features.gtf  
output (Input)

2: C:\dublinensis\_CD36.fasta  
output (Input)

3: DR34\_R1.fastq.gz  
output (Input)

4: DR34\_R2.fastq.gz  
output (Input)

6: Cutadapt  
FASTQ/A file #1  
FASTQ/A file #2  
out1 (fastqanger)  
out2 (fastqanger)

7: Map with BWA-MEM  
Use the following dataset as the reference sequence  
Select first set of reads  
Select second set of reads  
bam\_output (bam, qname\_sorted.bam, qname\_input\_sorted.bam)  
bam\_output (bam, qname\_sorted.bam, qname\_input\_sorted.bam)

8: Samtools flagstat  
BAM File to report statistics of  
output! (txt, tabular, local)

9: FreeBayes  
BAM or CRAM dataset  
Use the following dataset as the reference sequence  
output\_vcf (vcf)

10: SnpEff eff  
Sequence changes (SNPs, MNPs, InDels)  
SnpEff4.3 Genome Data  
Use custom interval file for annotation  
Only use the transcripts in this file  
snpEff\_output (vcf, bed)  
statsFile (html)

**About This Workflow**

variant calling with snpEff - Version 0

**Author**

cmdickens

All published Workflows by cmdickens

**Description**

This Workflow has no description.

**Tags**

**License**

No License specified

**Last Updated**

Monday Nov 17th 14:02:25 2025 GMT-6

**Sharing**

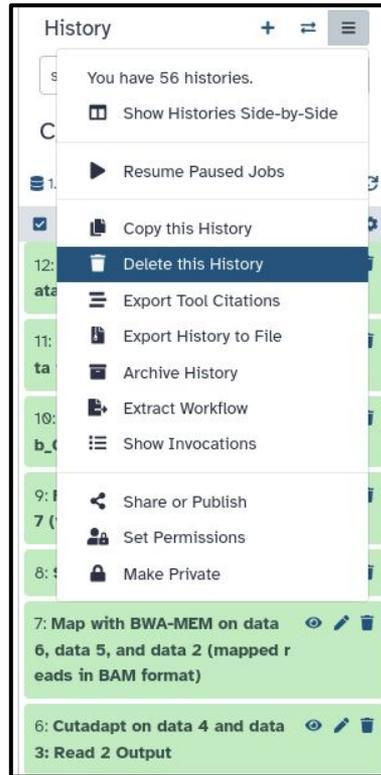
This workflow is not published and cannot be shared. Publish this workflow

Download Edit Run

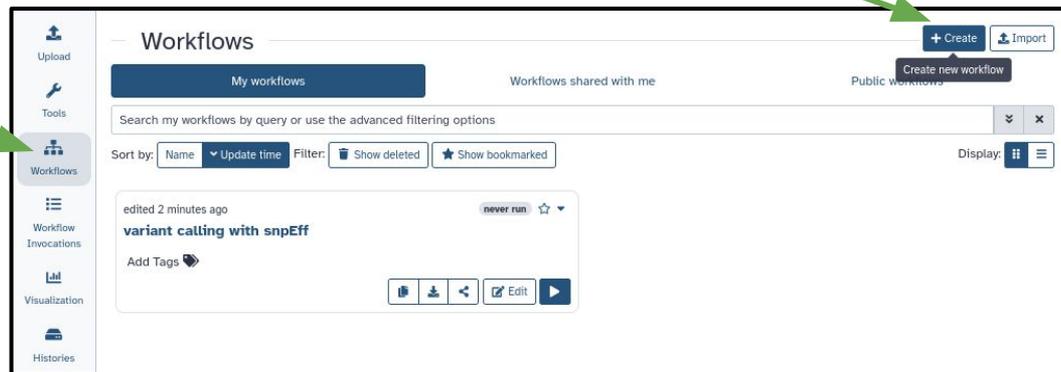
OK

# Permanently Delete History

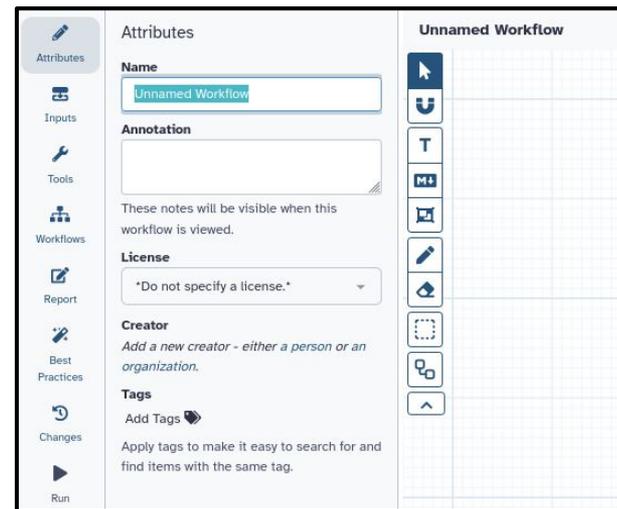
Delete/purge the C\_dublinsiensis history to reduce disk usage after you successfully create the workflow



# Create a Multi-step Workflow



Name the new workflow



# Build a Multi-step Workflow

- Search and select a tool to add it to the workflow
- Click and drag the output of one tool to the input of another tool
- Save the workflow

The screenshot displays the Galaxy workflow editor interface. The workflow is titled "align\_to\_ref\_genome (unsaved changes)". It contains two steps:

- 1: Cutadapt**: Takes two FASTQ/A files as input and produces two output files (out1 and out2).
- 2: Map with BWA-MEM**: Takes the output of the first step as input and produces a BAM file (bam\_output) and a sorted BAM file (qname\_sorted.bam).

The workflow is visualized as a sequence of steps connected by lines. A green arrow points to the connection between the output of the first step and the input of the second step. Another green arrow points to the second step in the top right corner. The right sidebar shows the configuration for the selected step, including a label, step annotation, and a "Conditionally skip step?" option.



**HIGH PERFORMANCE  
RESEARCH COMPUTING**  
TEXAS A&M UNIVERSITY

<https://hprc.tamu.edu>

HPRC Helpdesk:

help@hprc.tamu.edu

Phone: 979-845-0219

*Take our short course survey!*



HPRC Survey

[https://u.tamu.edu/hprc\\_shortcourse\\_survey](https://u.tamu.edu/hprc_shortcourse_survey)

Help us help you. Please include details in your request for support, such as, **Cluster** (ACES, FASTER, Grace, Launch), NetID (UserID), Job information (**JobID**(s)), Location of your jobfile, input/output files, Application, Module(s) loaded, Error messages, etc), and Steps you have taken, so we can reproduce the problem.