HIGH PERFORMANCE RESEARCH COMPUTING

Introduction to HPRC Galaxy

November 21, 2025

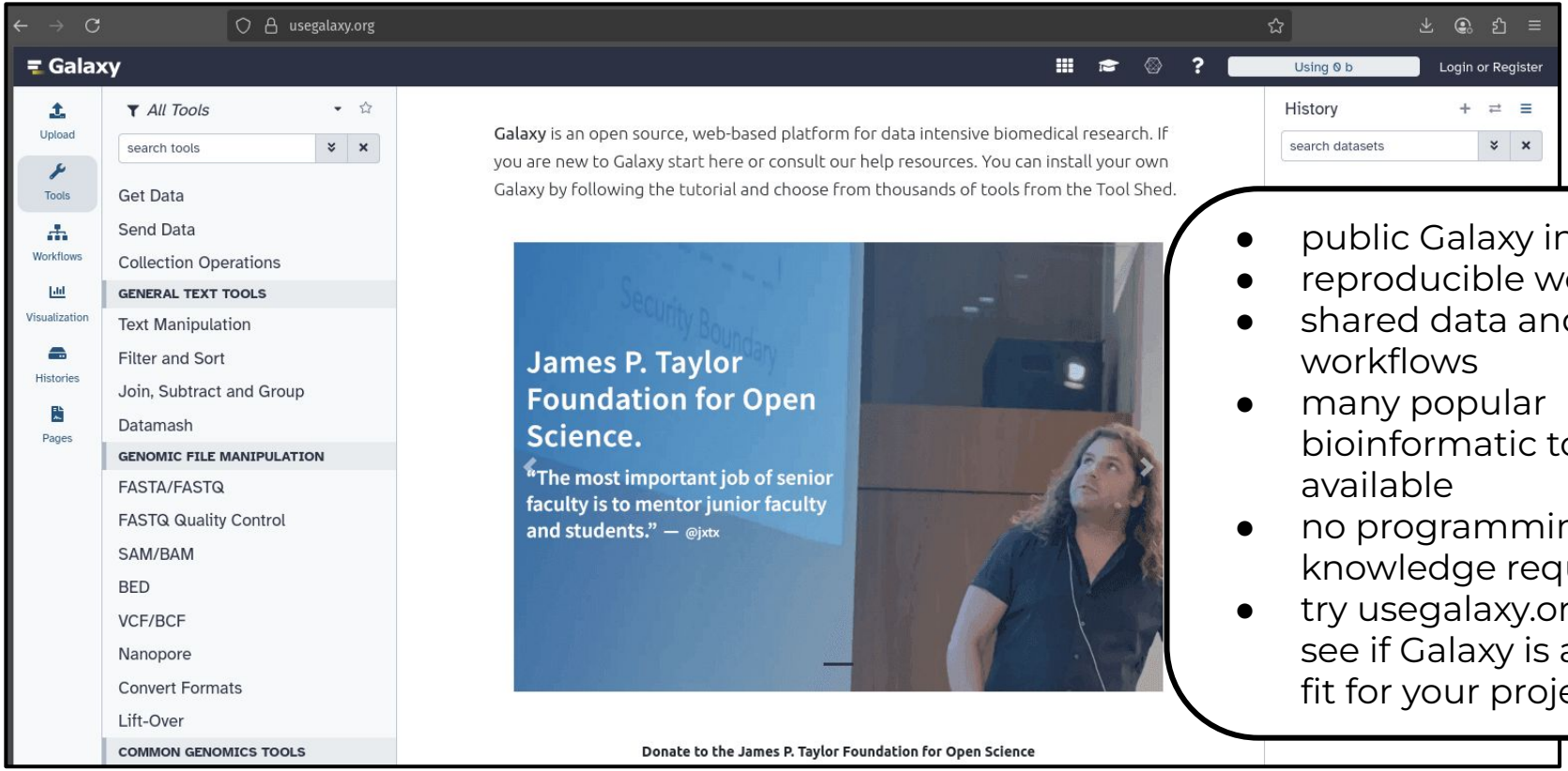
Michael Dickens

High Performance
Research Computing
DIVISION OF RESEARCH

Galaxy Training Objectives

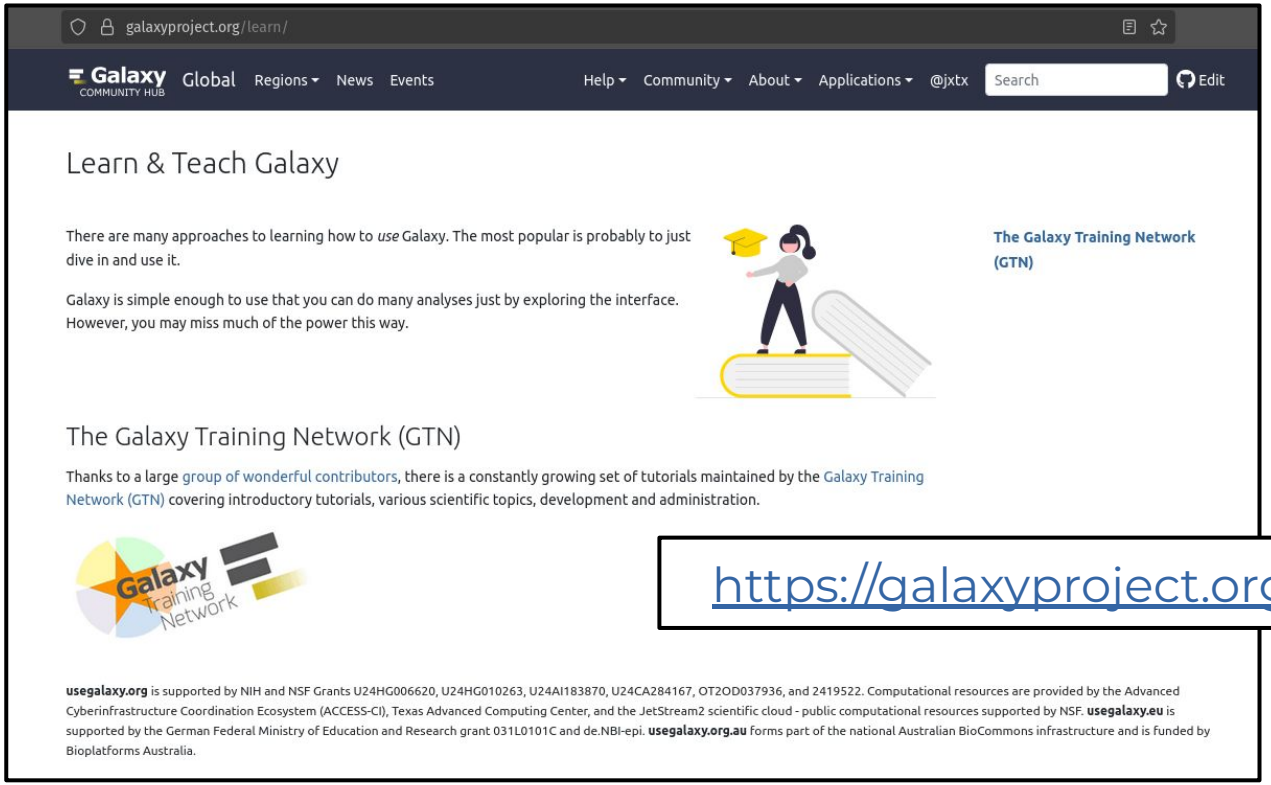
- usegalaxy.org
- upload and download data
- create new histories
- search and switch tool version
- view data
- view/change file attributes
- permanently delete files
- debug jobs
- copy datasets between histories
- run a variant calling analysis
- create a workflow

usegalaxy.org



- 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 Videos and Tutorials

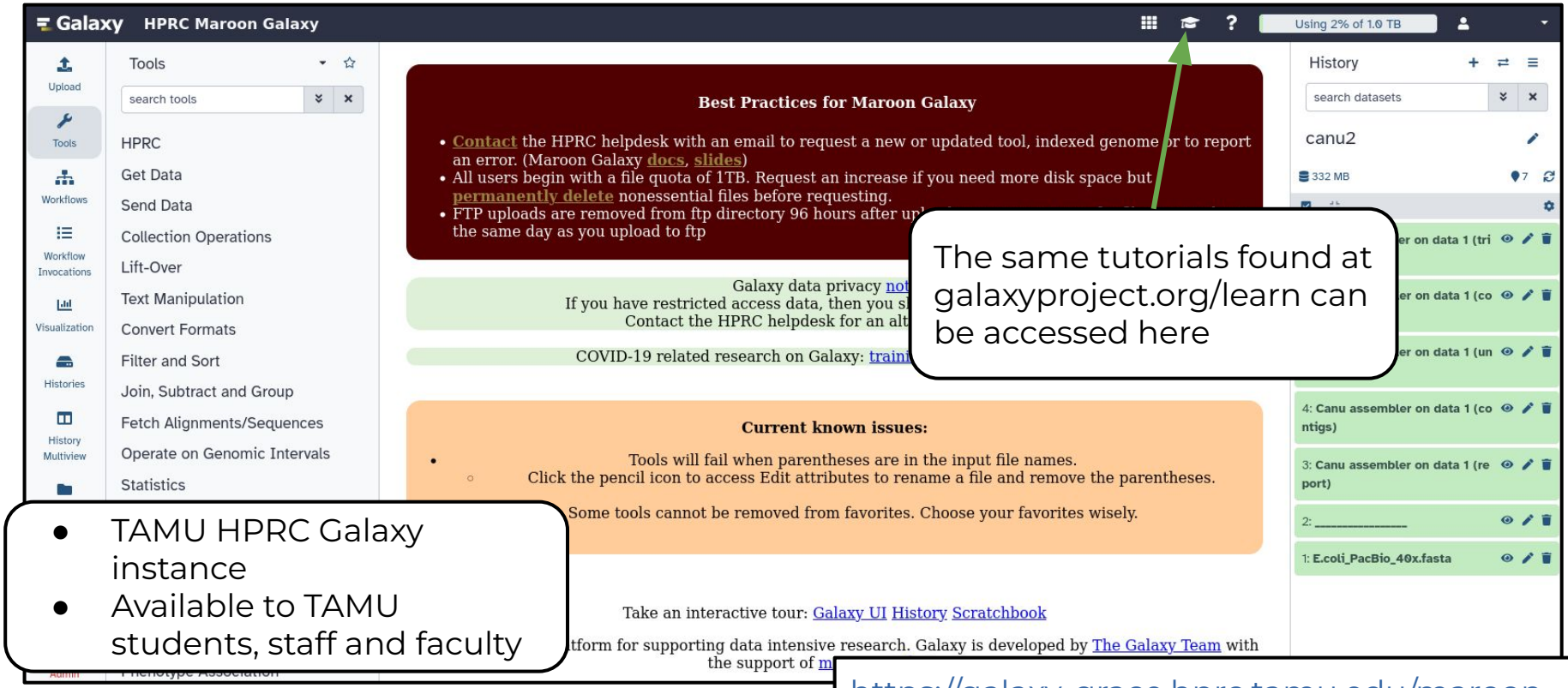


https://galaxyproject.org/learn

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
    - 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
  - hprc.tamu.edu/kb/User-Guides/Galaxy
- There are no backups of users' Galaxy files

HPRC Maroon Galaxy

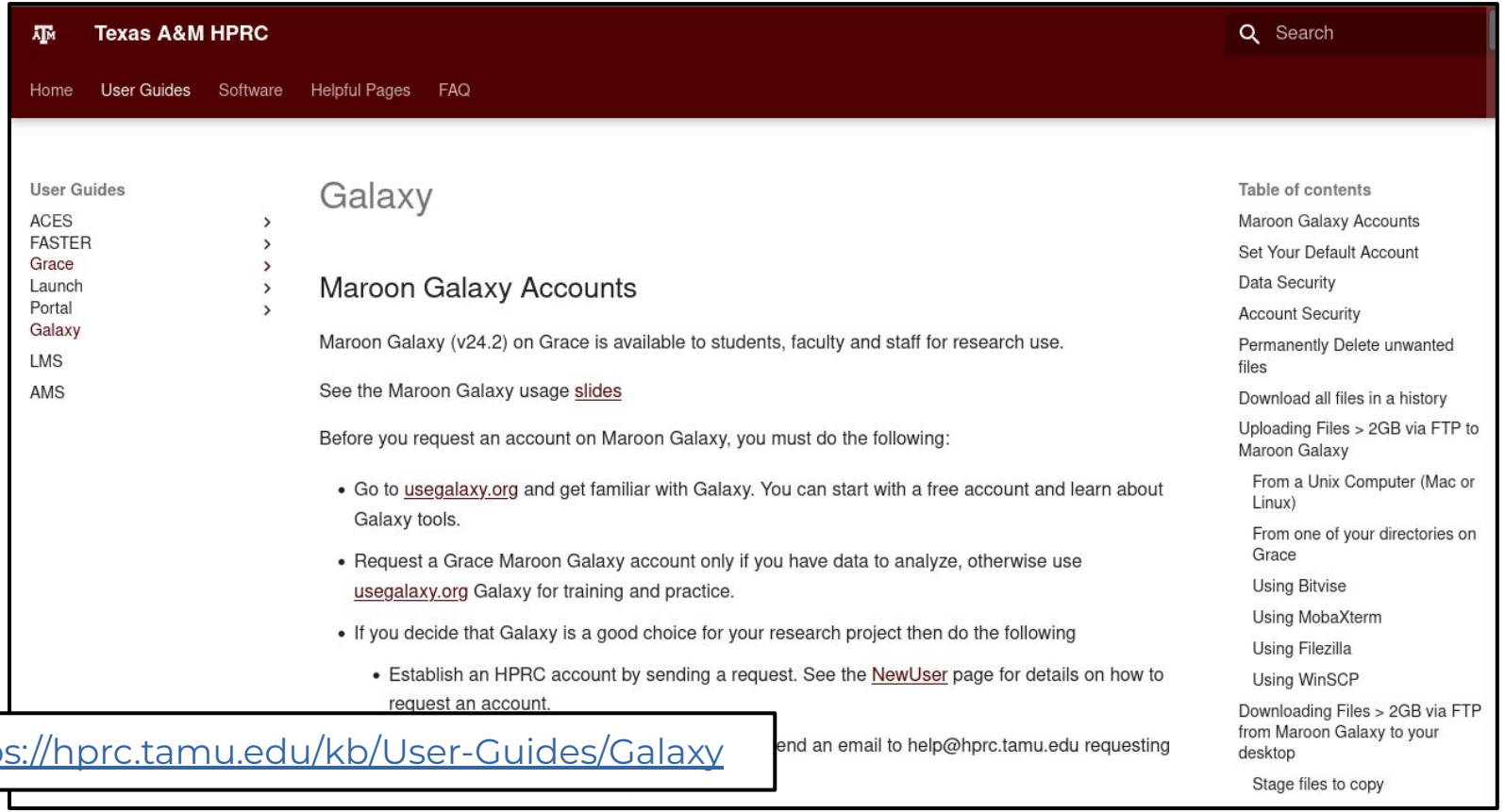


The same tutorials found at galaxyproject.org/learn can be accessed here

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

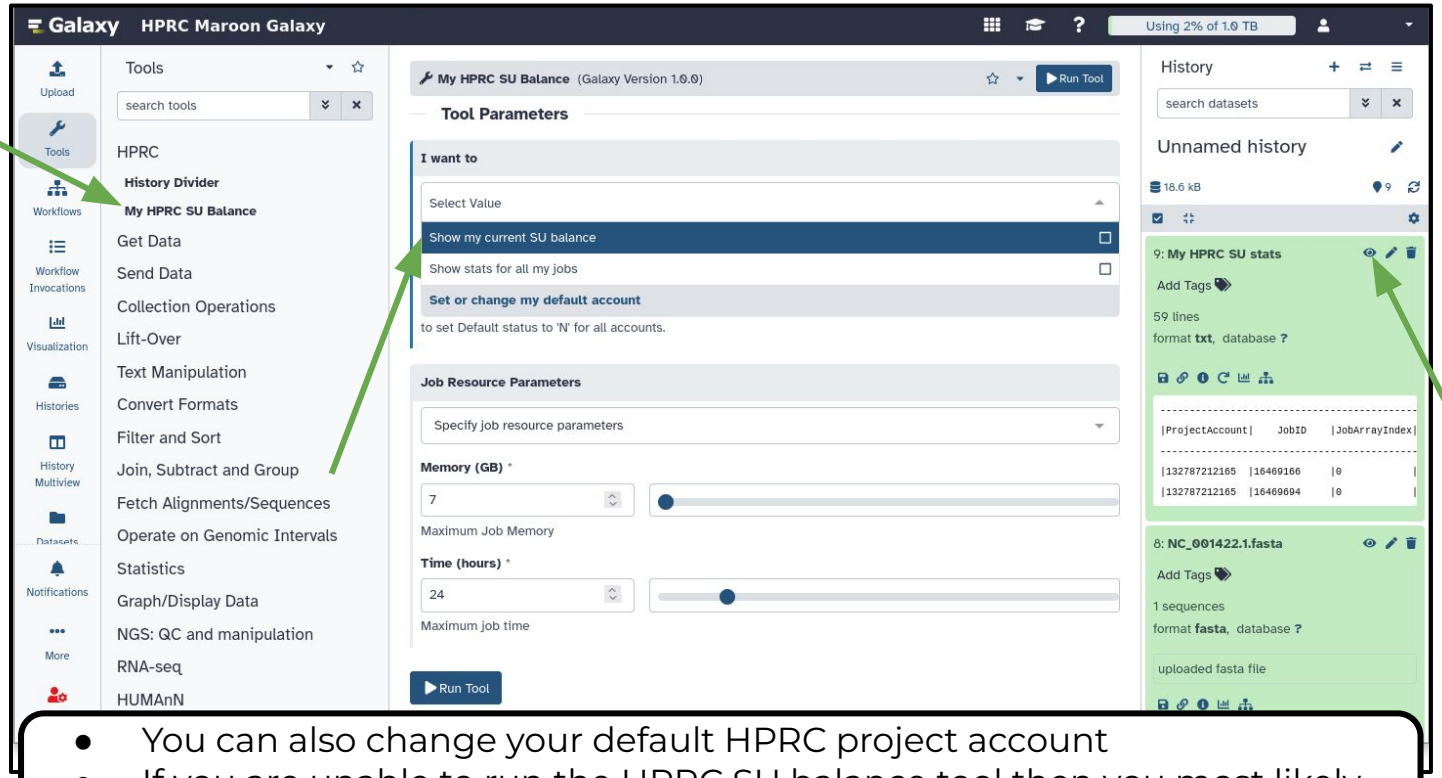
https://galaxy-grace.hprc.tamu.edu/maroon

HPRC Galaxy Documentation Knowledge Base



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

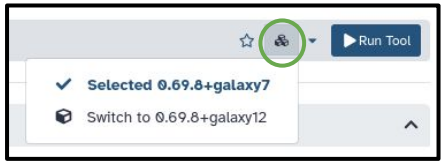
Check Your HPRC SUs Balance



- You can also change your default HPRC project account
- If you are unable to run the HPRC SU balance tool then you most likely need to renew your HPRC account or you are out of SUs

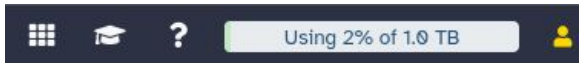
Galaxy Features

1. You can easily switch the software version

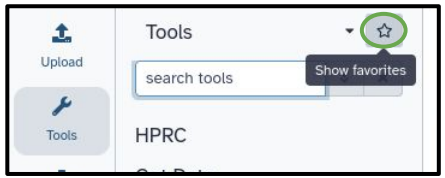


2. Tutorials are easy to follow along without leaving Maroon Galaxy

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



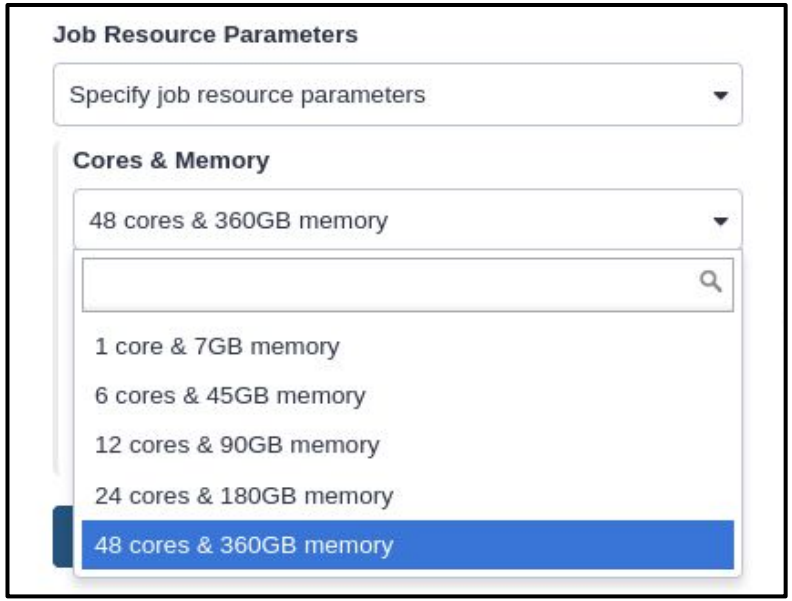
4. You can add tools to your favorites to easily locate



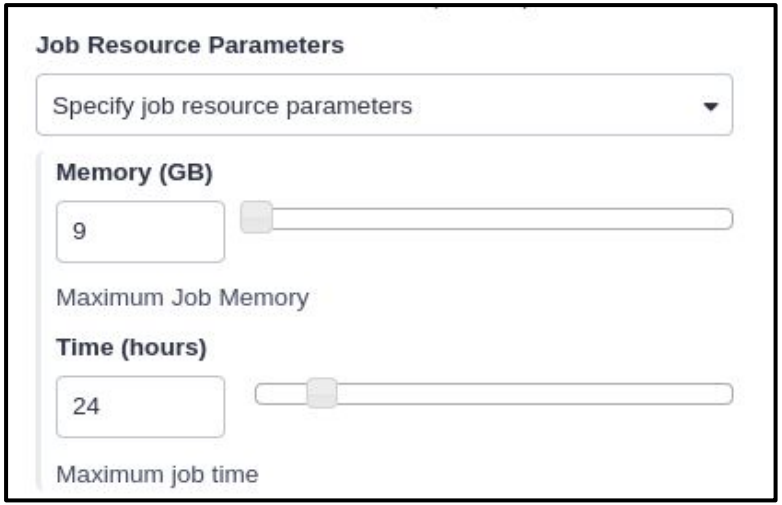
Galaxy Features cont.

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

tools supporting multi-core processing



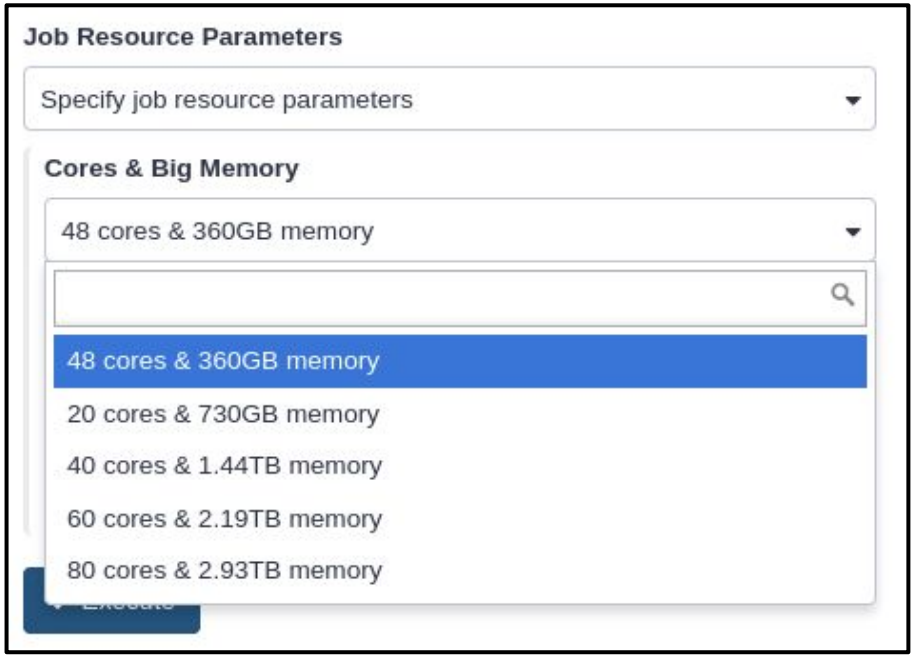
tools supporting single-core processing



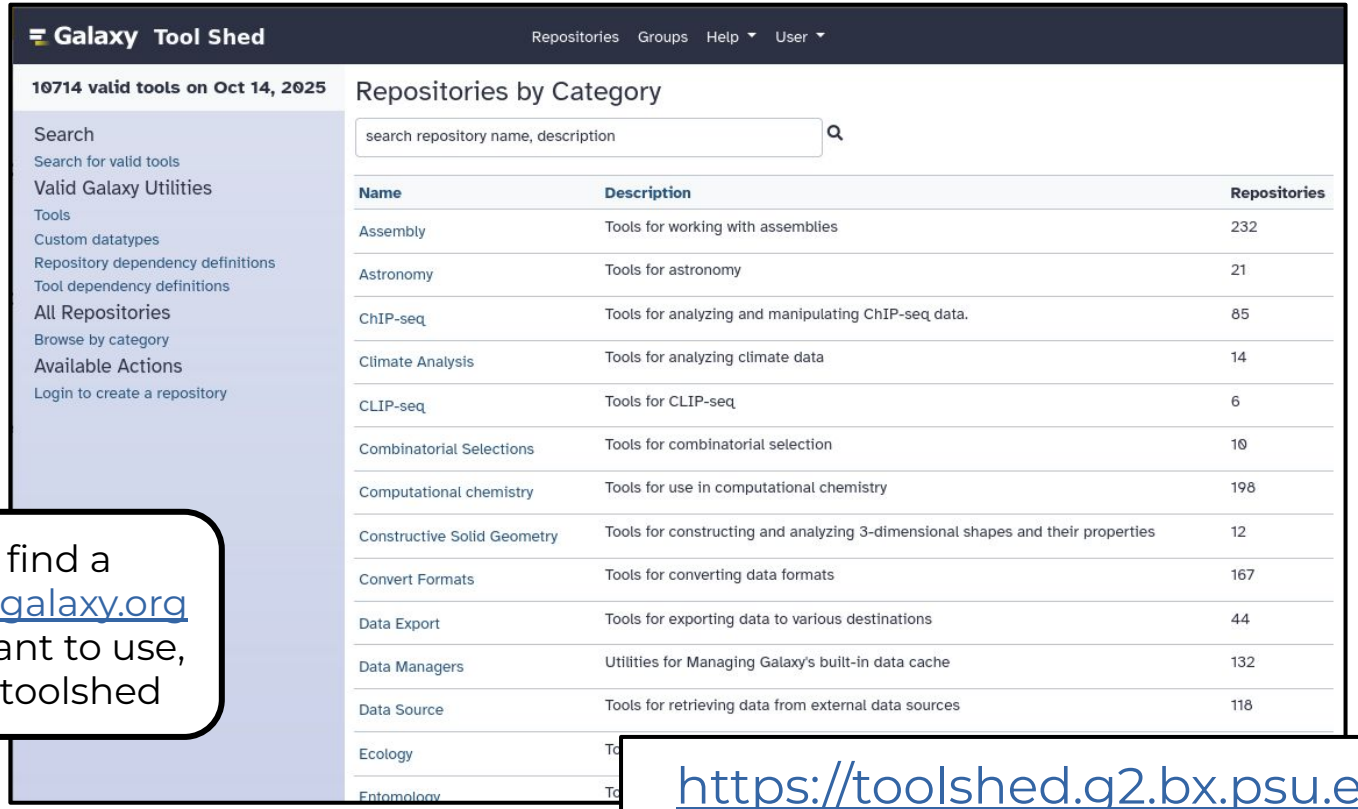
Galaxy Features cont.

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

tools supporting multi-core and requiring big memory processing



Look for Additionally Available Tools

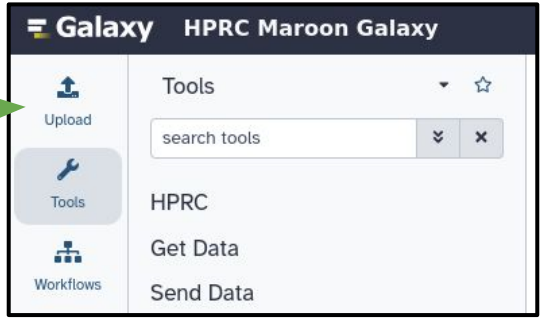


If you can't find a tool on usegalaxy.org that you want to use, search the toolshed

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

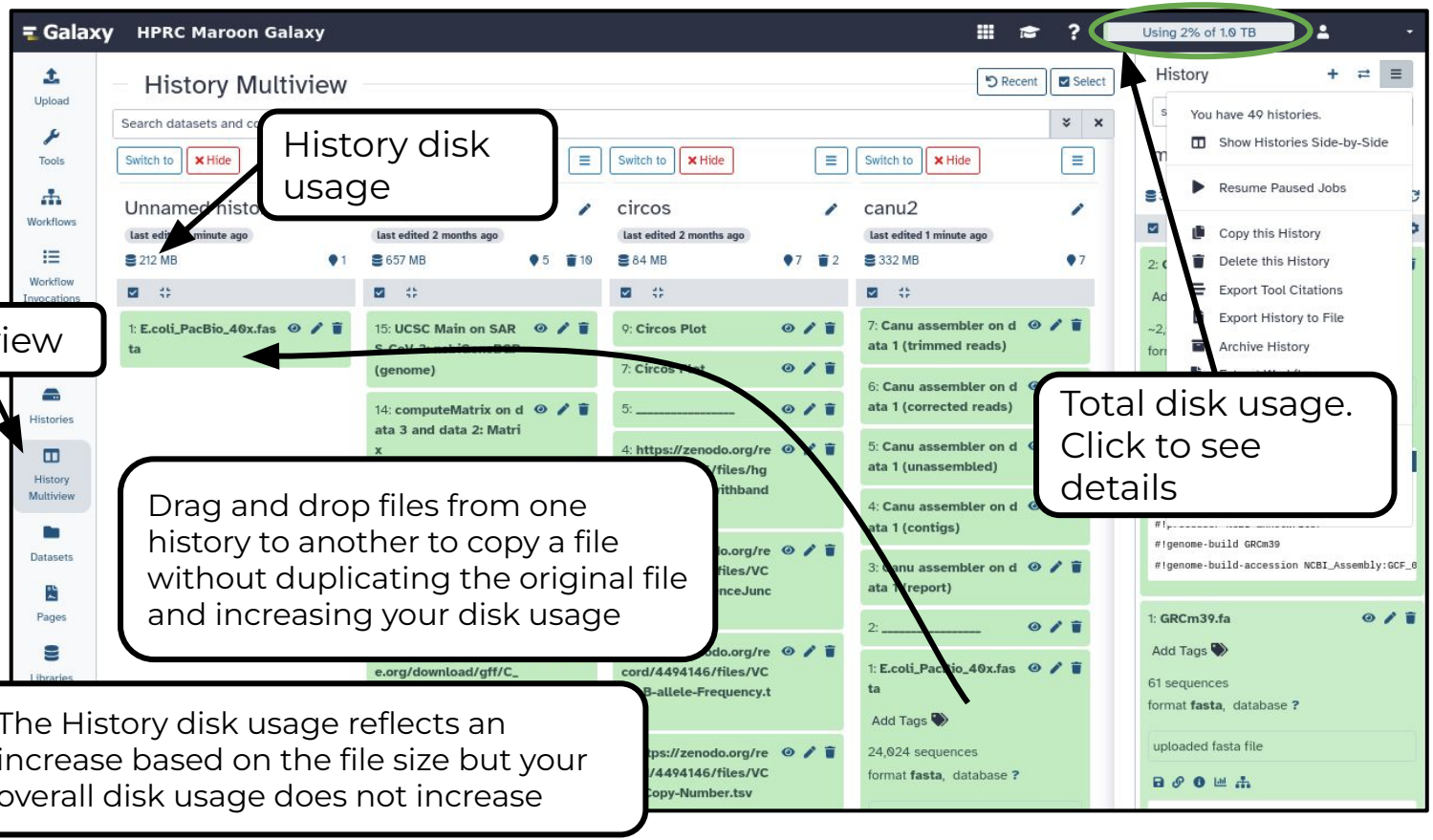
Get Data into Your Galaxy History

upload button



- 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



History disk usage

multiview

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

Total disk usage. Click to see details

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

Uploading a 2GB+ Size File to Maroon Galaxy

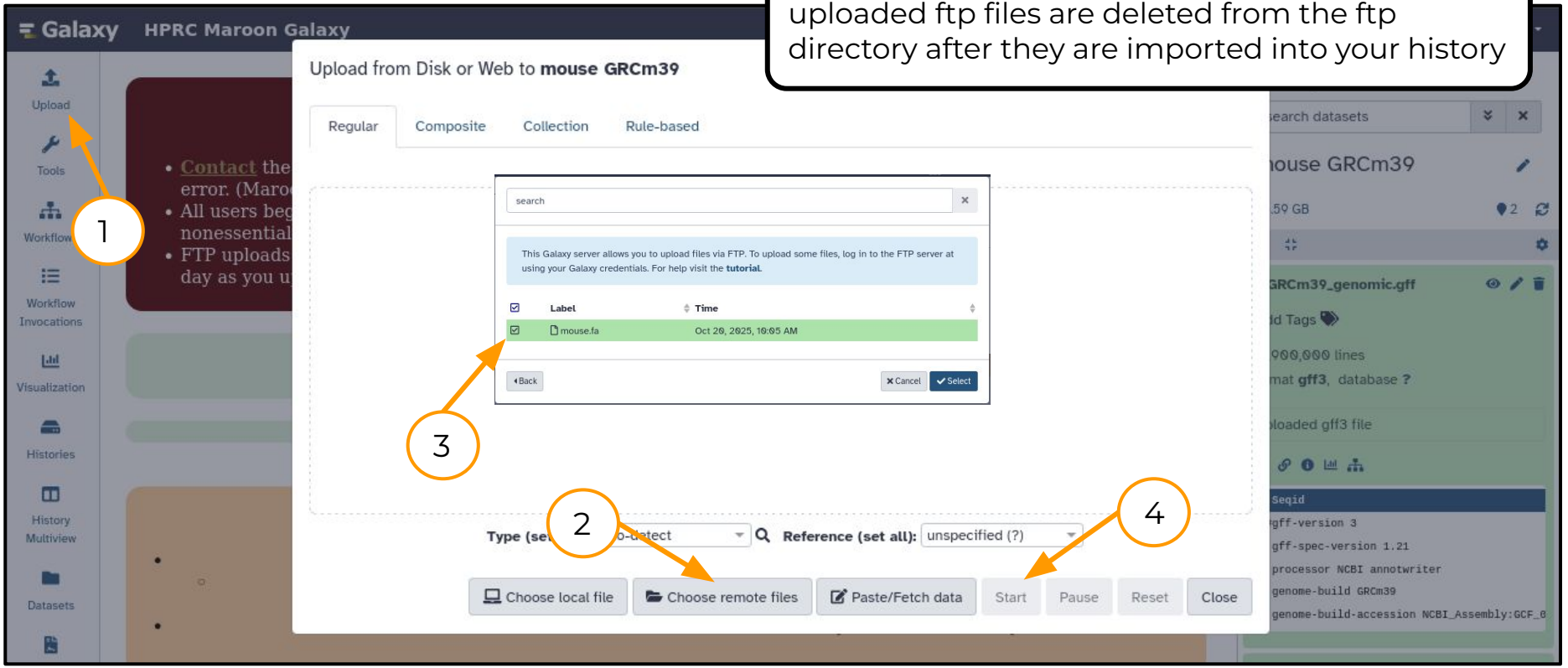
- Files larger than 2GB should be copied using ftp instead of the upload button which uses the http protocol which has limitations on file sizes for file transfers
- There are three options for uploading files via ftp
  a. Use the sftp command in a Unix terminal on your Mac or Linux desktop
  b. Use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

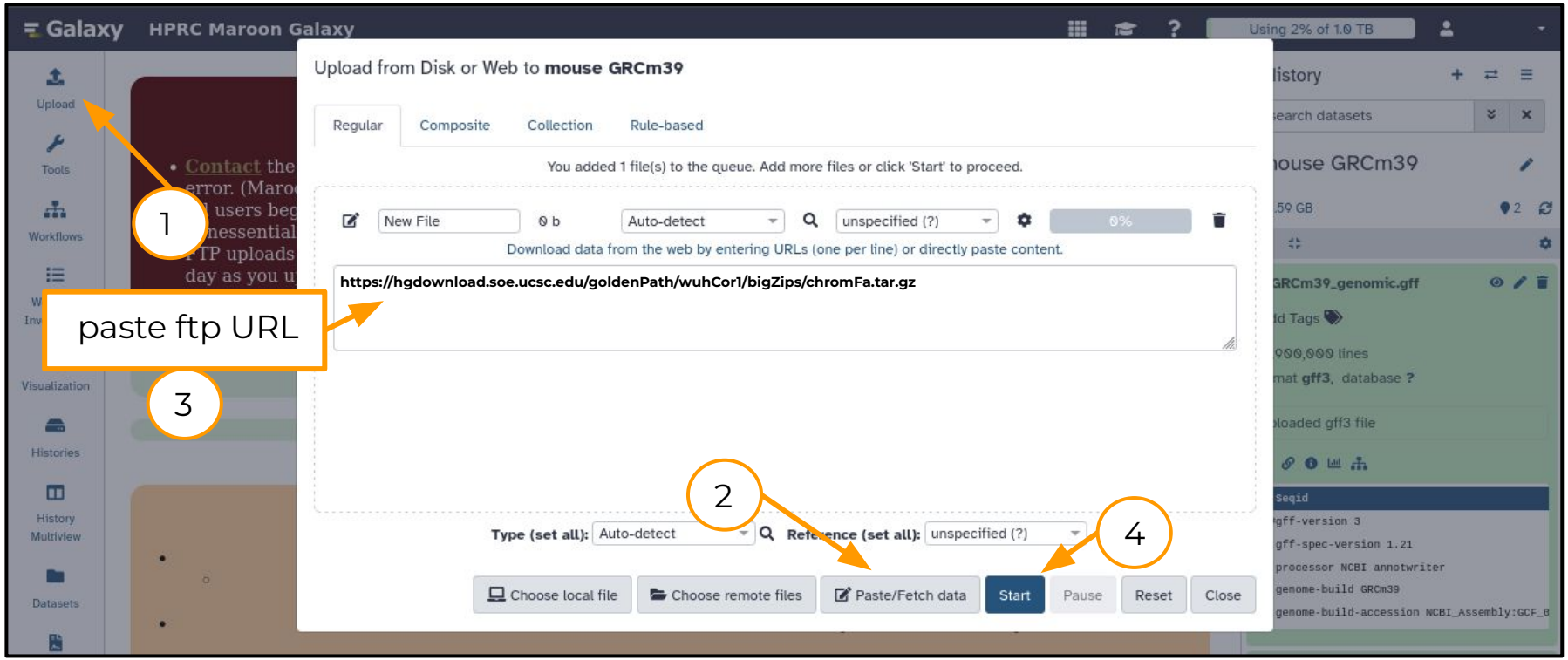
https://hprc.tamu.edu/kb/User-Guides/Galaxy/#uploading-files-2gb-via-ftp-to-maroon-galaxy

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



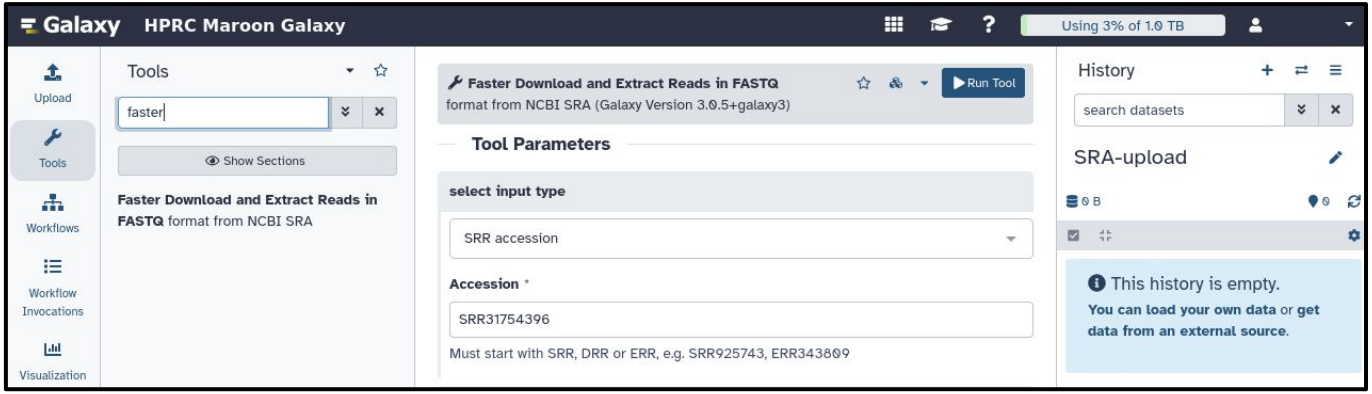
FTP Upload File Directly to History via URL



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

Get Data Exercise

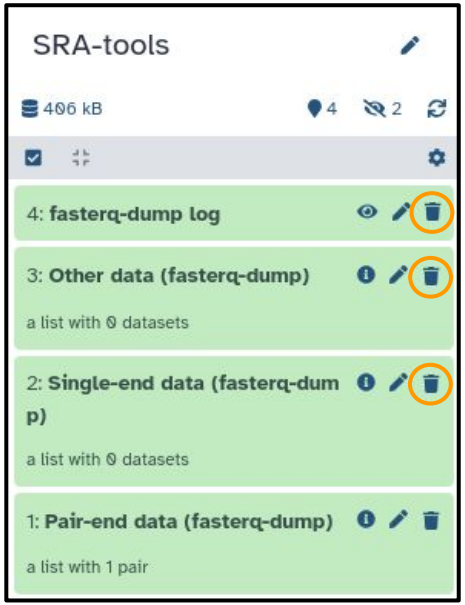
1. Create or rename a new history with the name SRA-upload
2. Use the "Faster Download and Extract Reads in FASTQ NCBI SRA" tool to download sequence reads for accession **SRR31754396** to your history



Permanently Delete Nonessential Files

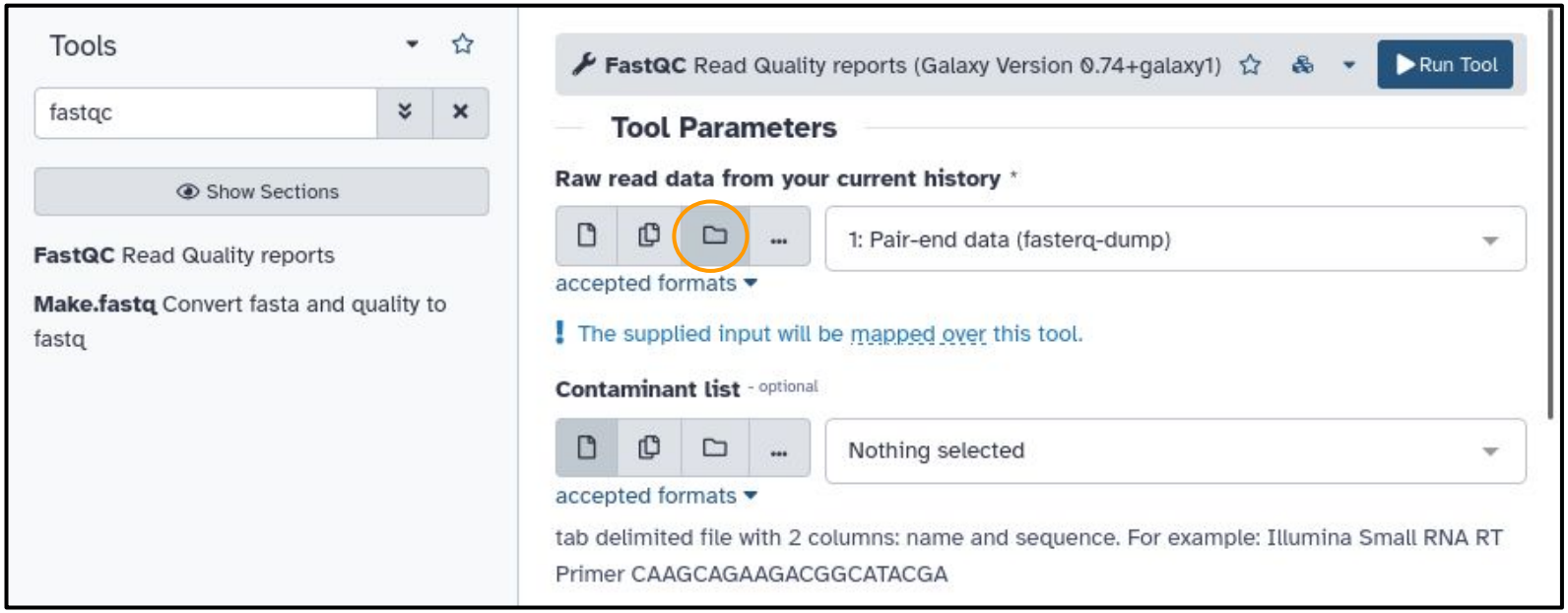
purge files from disk and reduce your Galaxy disk usage

Deleted non-essential files



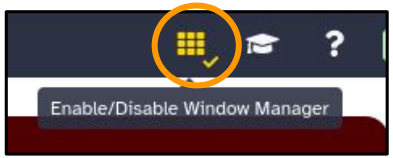
Purge deleted content

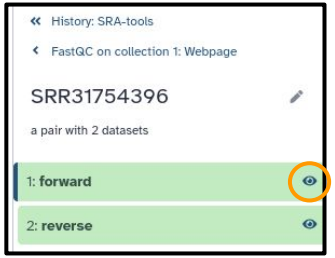
Run FastQC on List of Paired Files

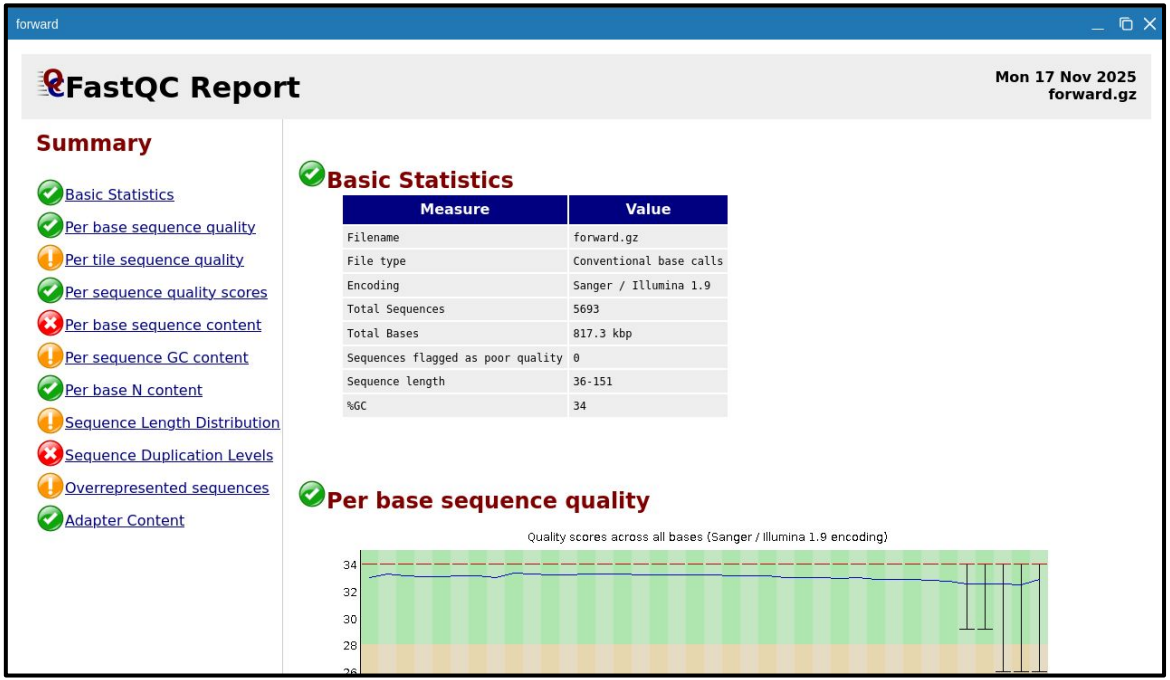


Review FastQC Output

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

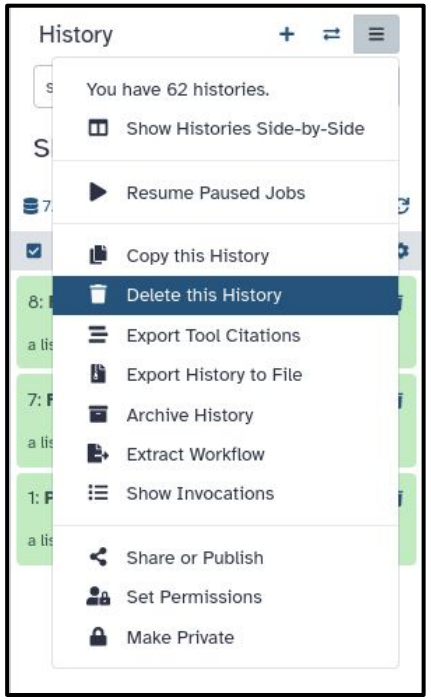


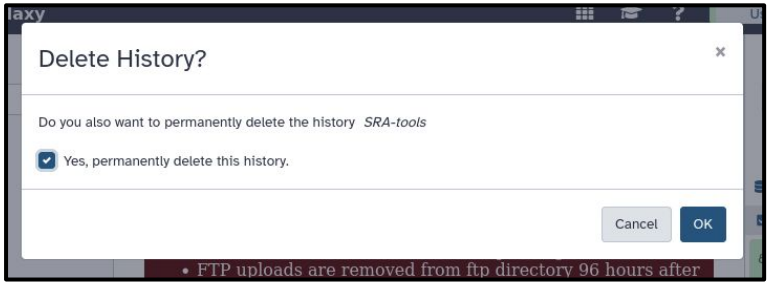




Permanently Delete History

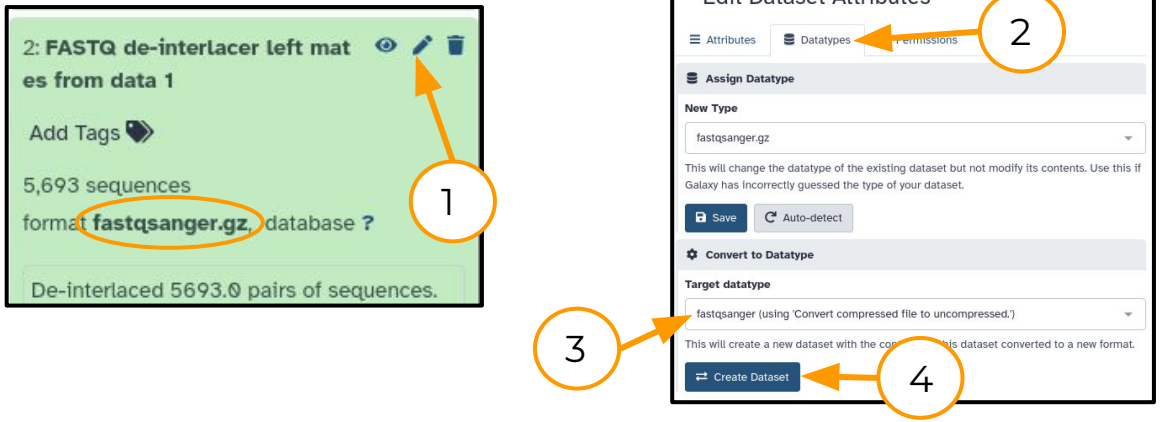
purge history from disk to reduce disk space





Converting and Using gzipped Files

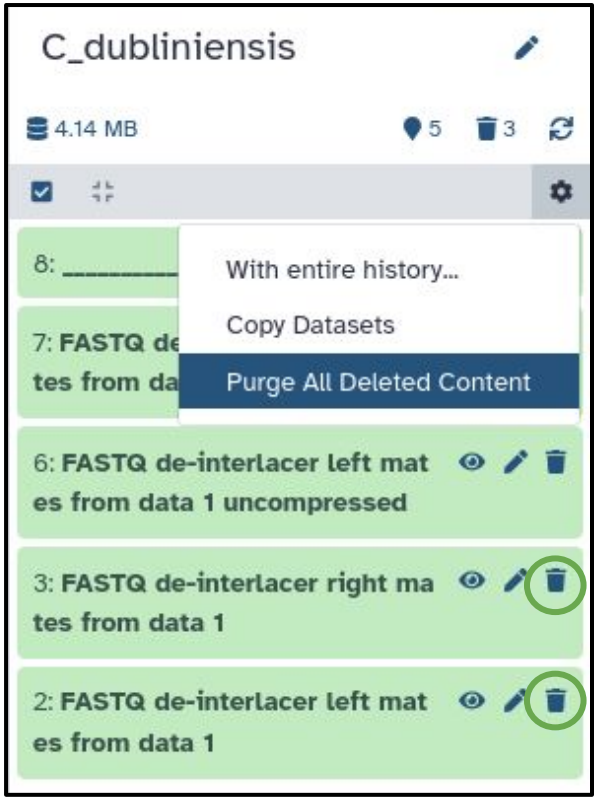
- Although Galaxy now supports uploading files in gzipped format, some Galaxy tools do not support gzipped format for input files.
- Files can be converted to uncompressed format but the original gzipped file should be permanently deleted in order not to have duplicated data that takes up disk space



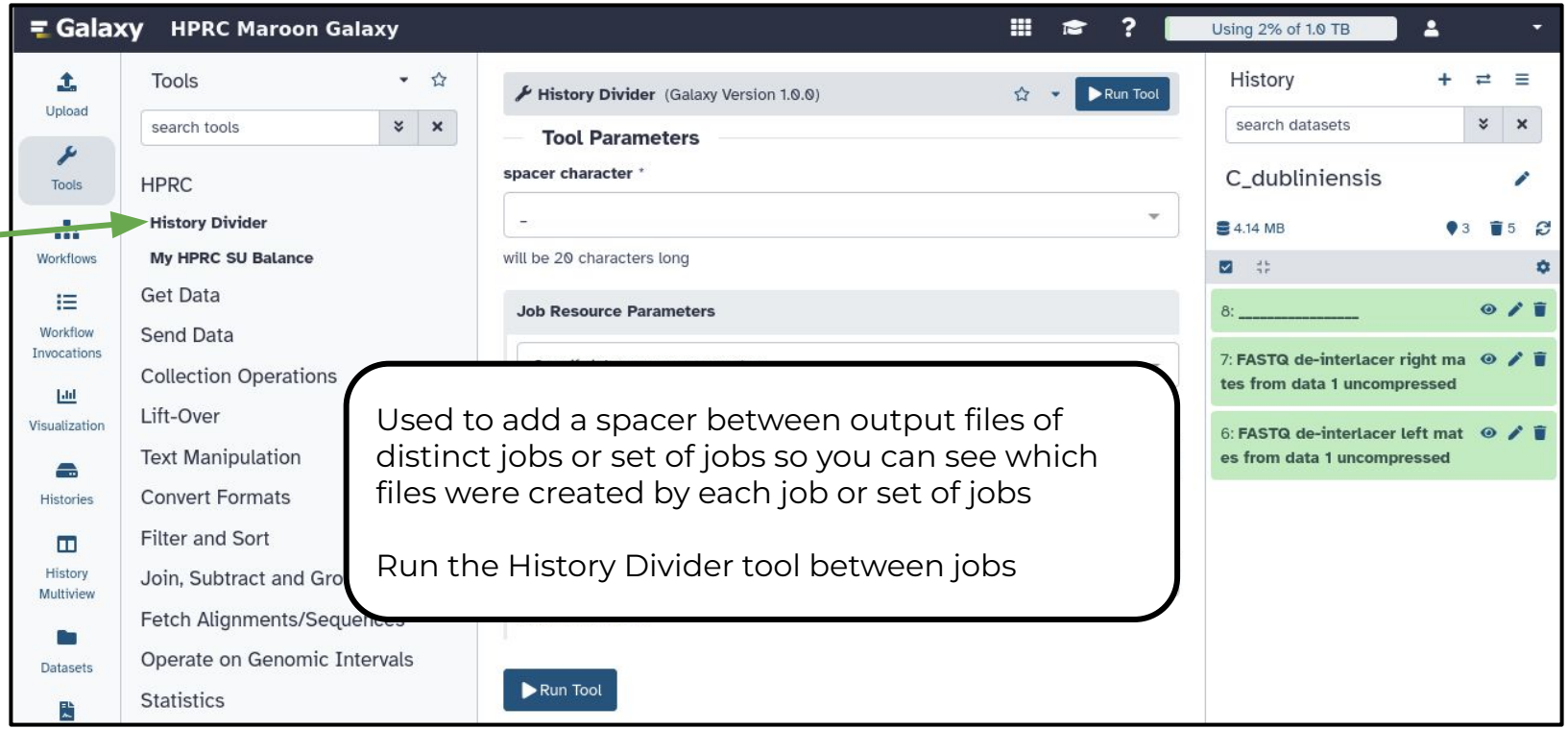
Convert the left and right compressed fastqsanger.gz files to uncompressed fastqsanger

Delete and Purge the Compressed Files

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



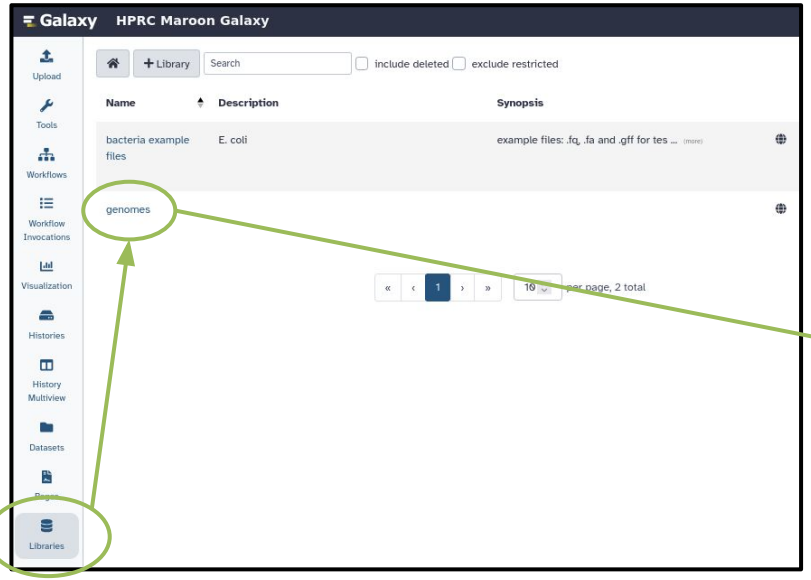
History Divider

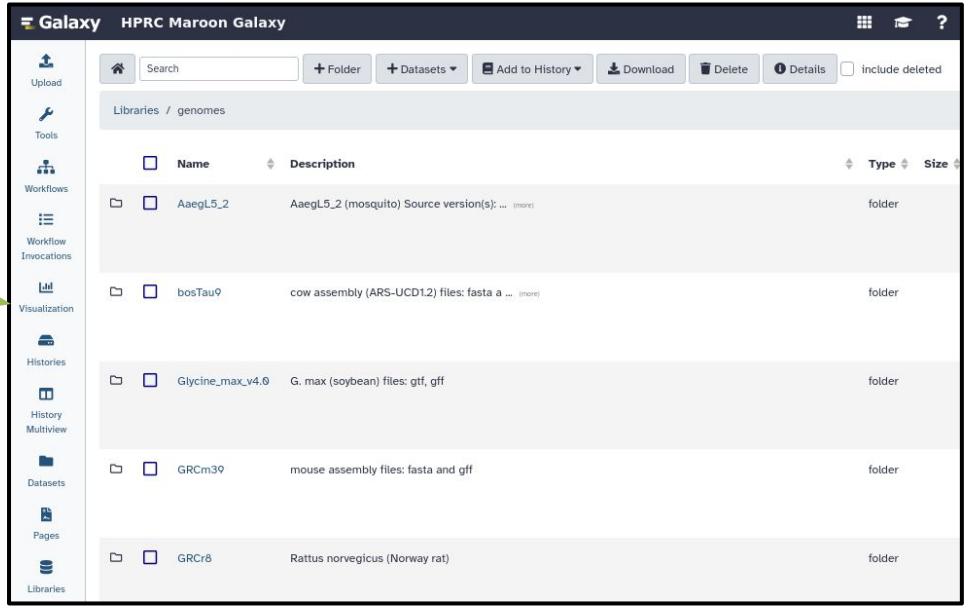


Used to add a spacer between output files of distinct jobs or set of jobs so you can see which files were created by each job or set of jobs

Run the History Divider tool between jobs

Shared Libraries



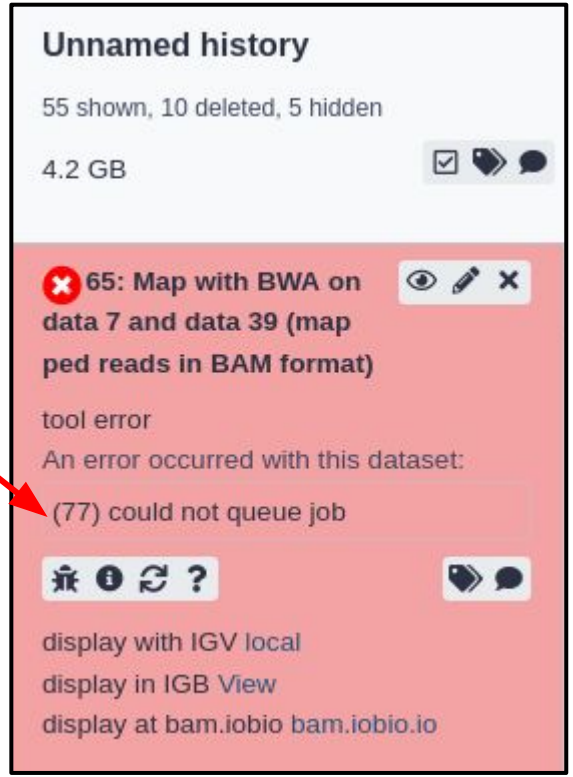


Files can be added to a 'Data Library' which you can share with your colleagues.
Send a request to the HPRC helpdesk if you would like a Data Library for your group

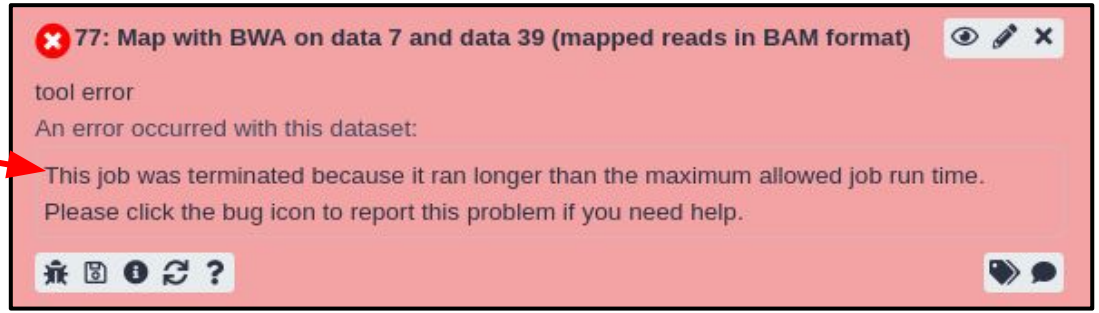
Debugging Failed Jobs

Failed Job: could not queue job

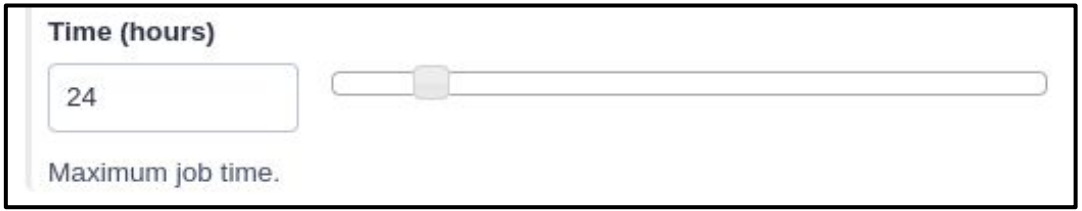
- Make sure you have enough SUs to run the job
  - My HPRC SU Balance
- Make sure your account has been renewed for the current fiscal year which starts September 1 each year
- 77 is the Galaxy job ID and not a fail code
- Check with the HPRC helpdesk if you have enough SUs and your job does not queue



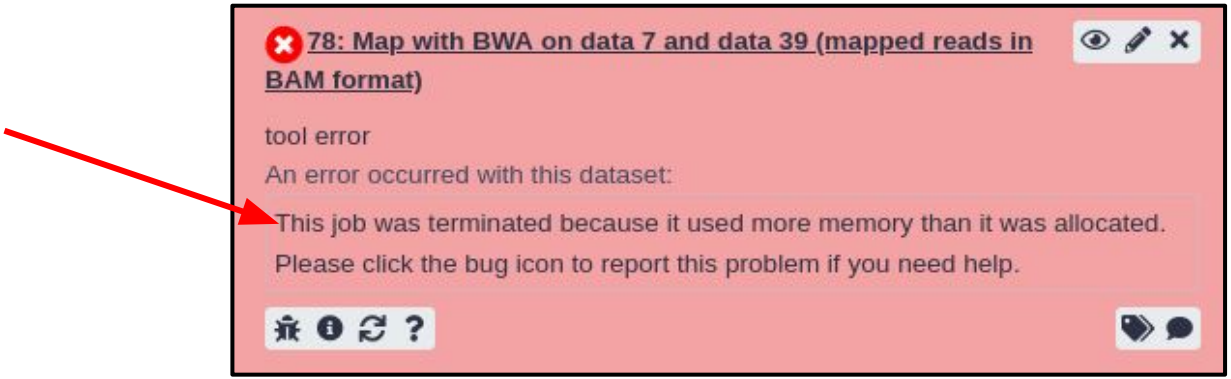
Failed Job: walltime limit reached



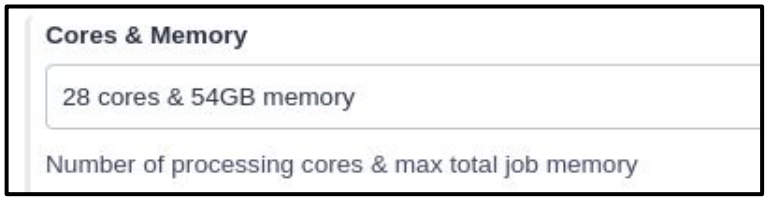
- Configure the job to use more Time



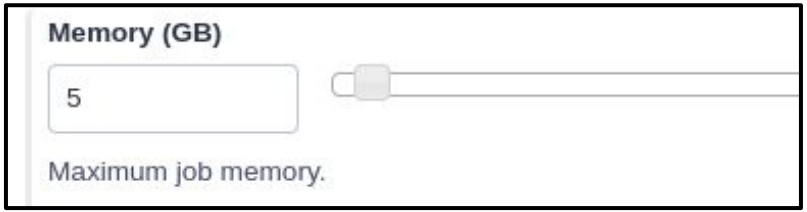
Failed Job: memory limit reached



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



or

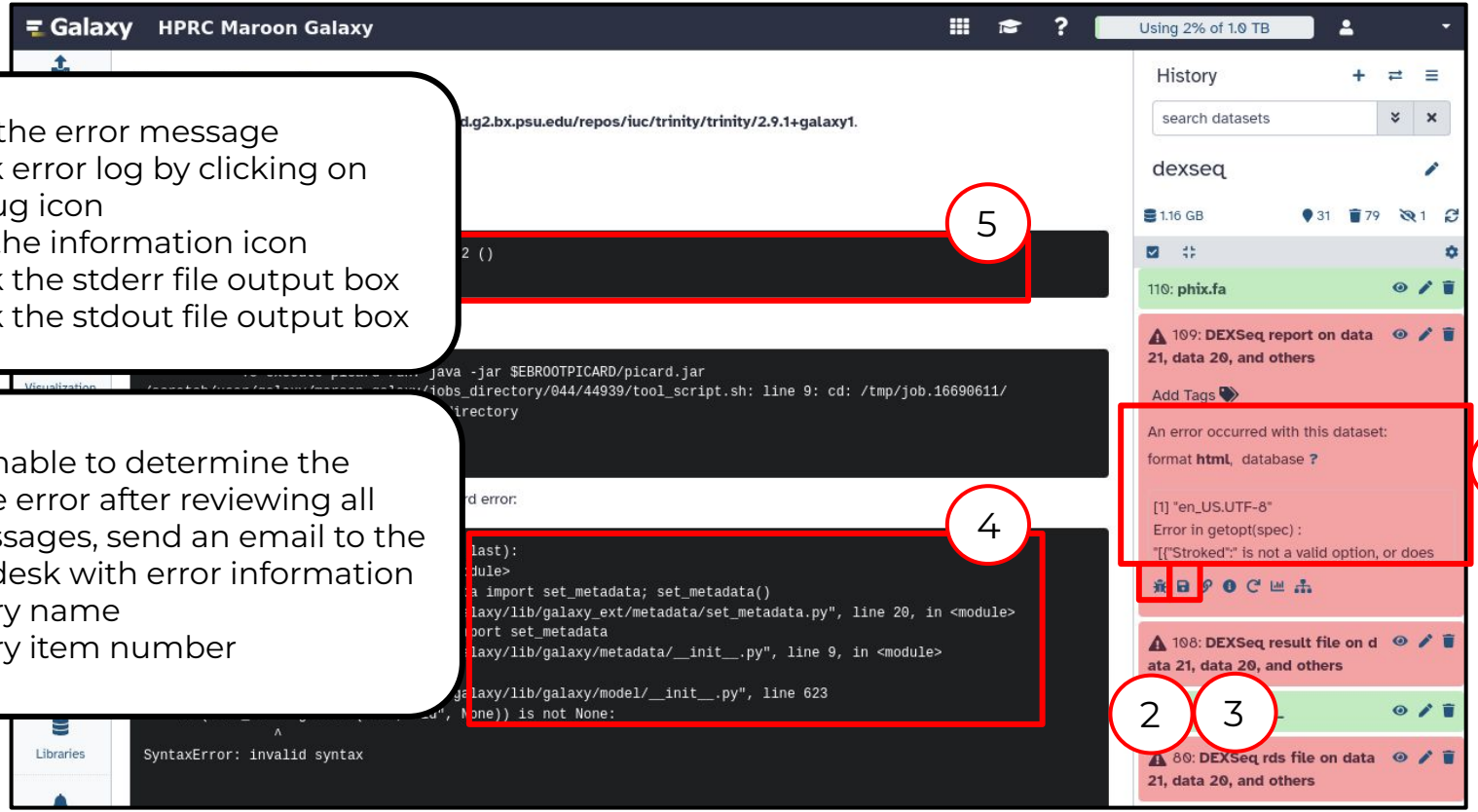


See Logs of Failed Jobs

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

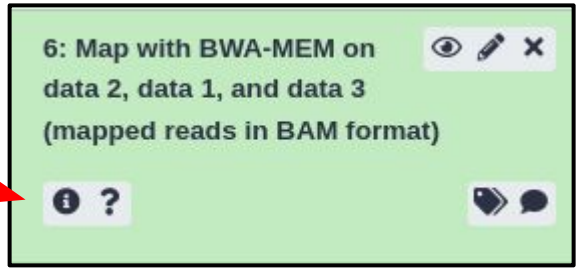
If you are unable to determine the cause of the error after reviewing all the log messages, send an email to the HPRC helpdesk with error information

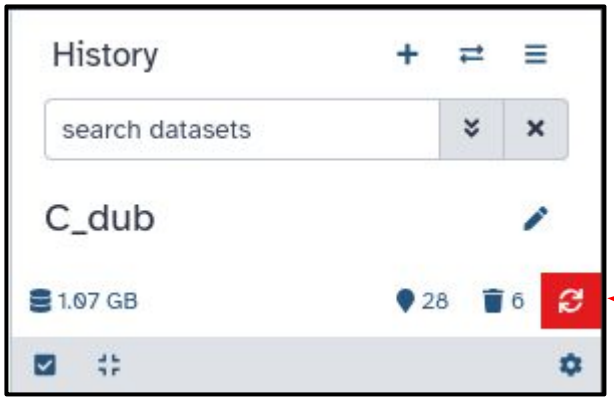
- History name
- History item number



Refresh Web Browser

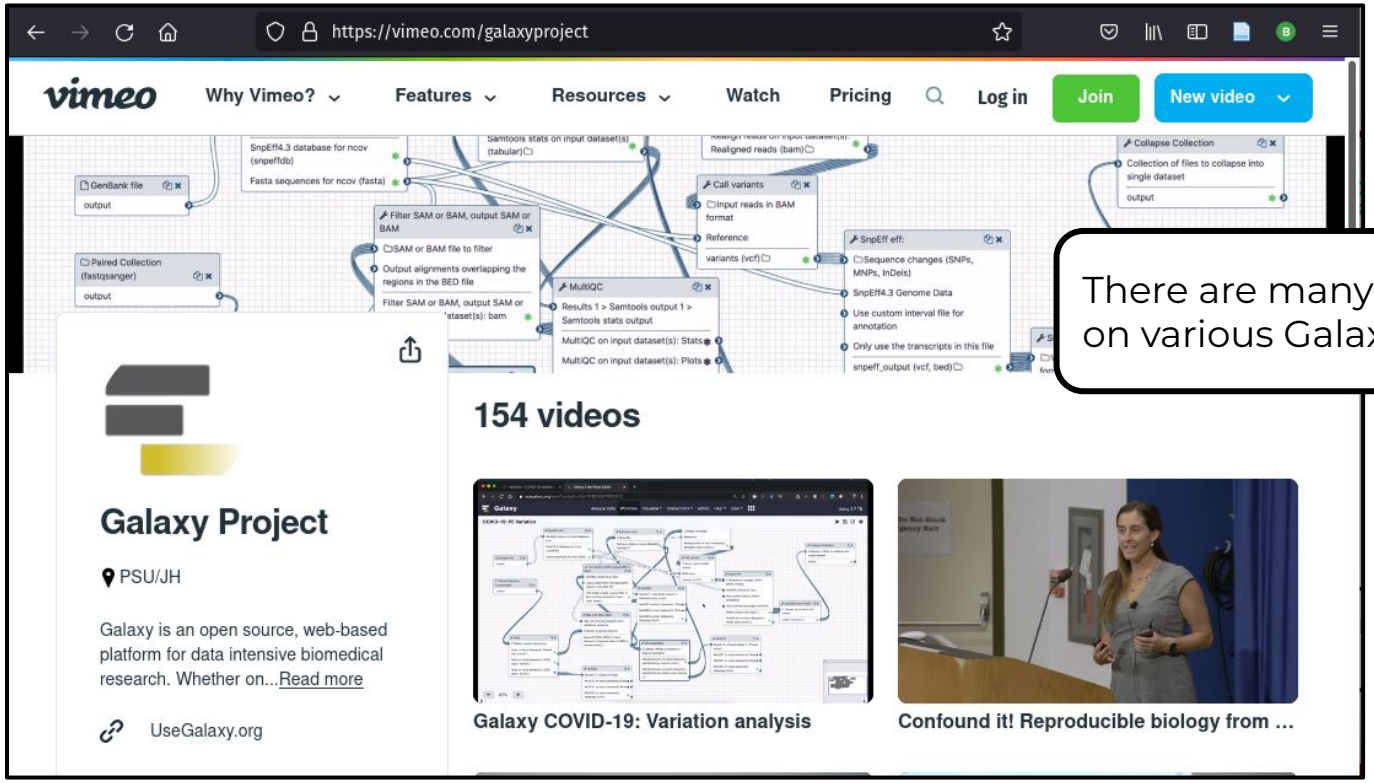
The disk image download button is missing





red refresh icon in history

Galaxy Videos https://vimeo.com/galaxyproject



There are many videos available on various Galaxy topics

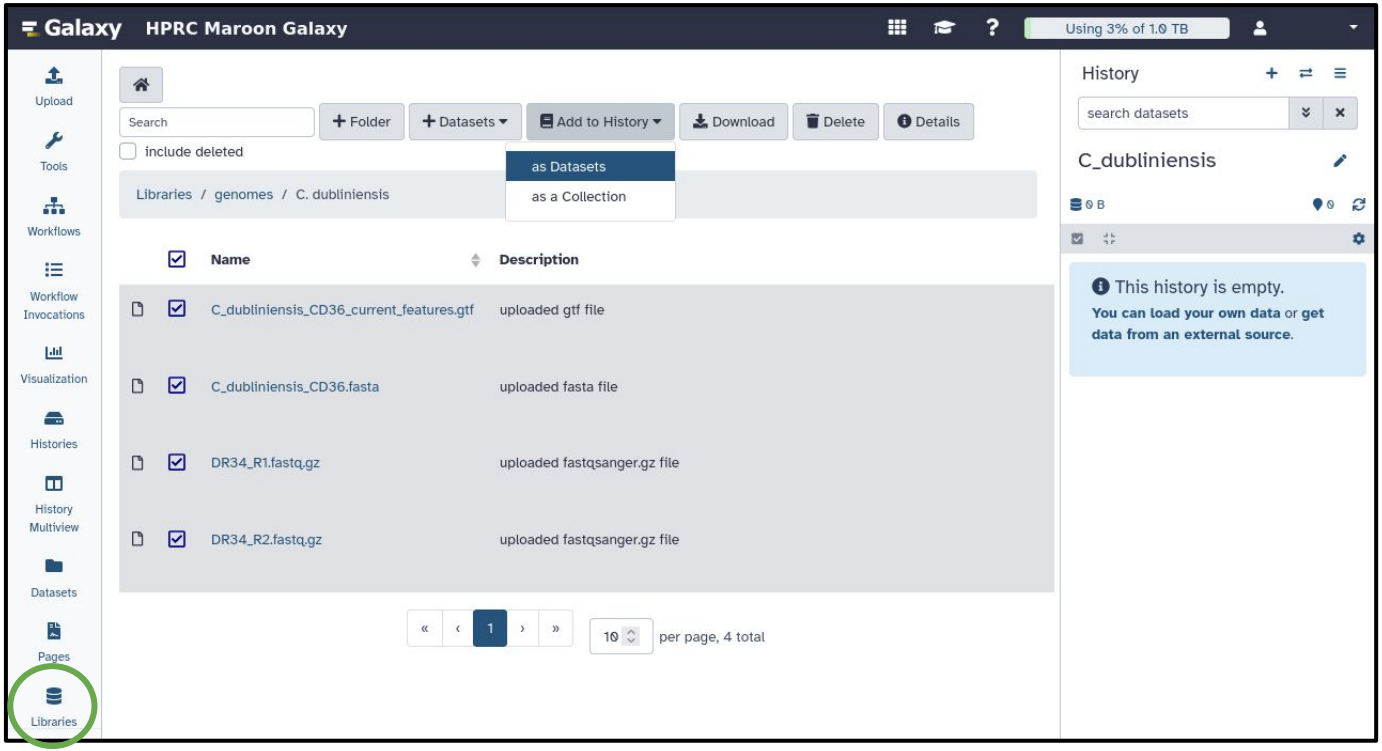
Sequence Variant Calling Exercise

Prep Reads for Alignment to Reference Genome

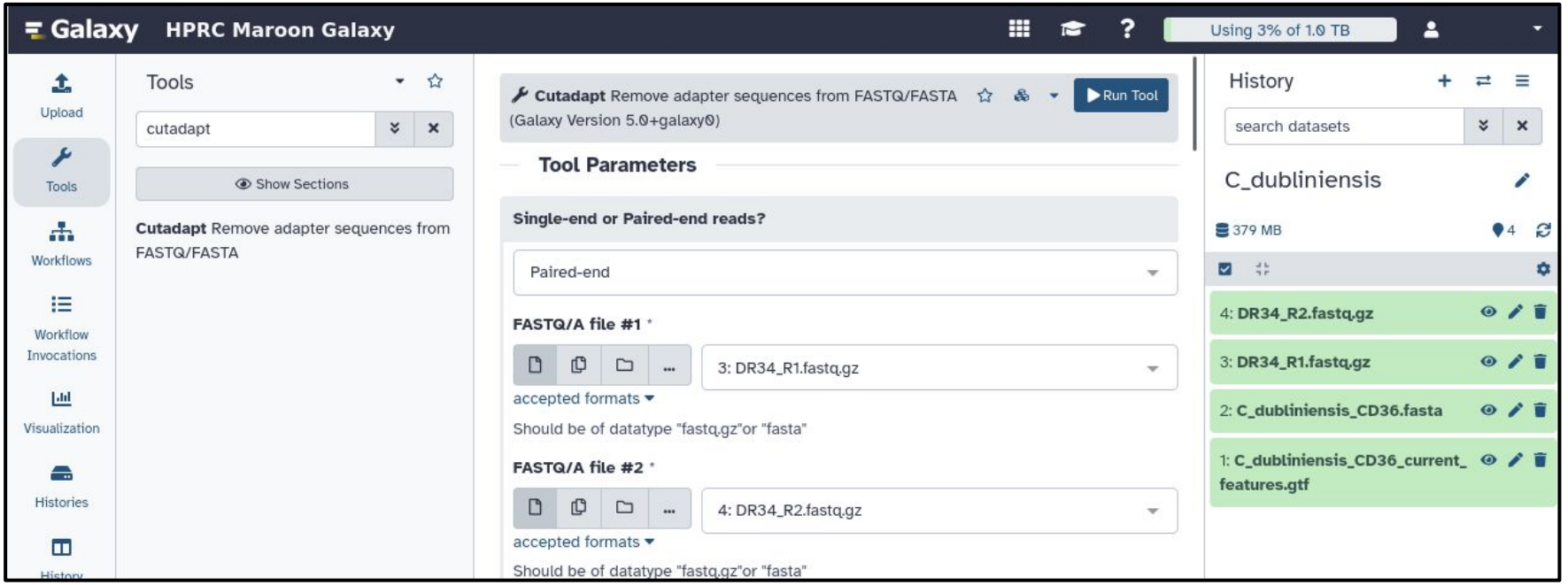
1. Create a history named C_dubliniensis
2. Import sequence reads and genome build from Libraries
3. Run FastQC on the paired end sequence read files
    - review the results
4. Trim reads with the cutadapt tool
5. Delete the un-trimmed input files after cutadapt has completed

Import Reference Genome

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

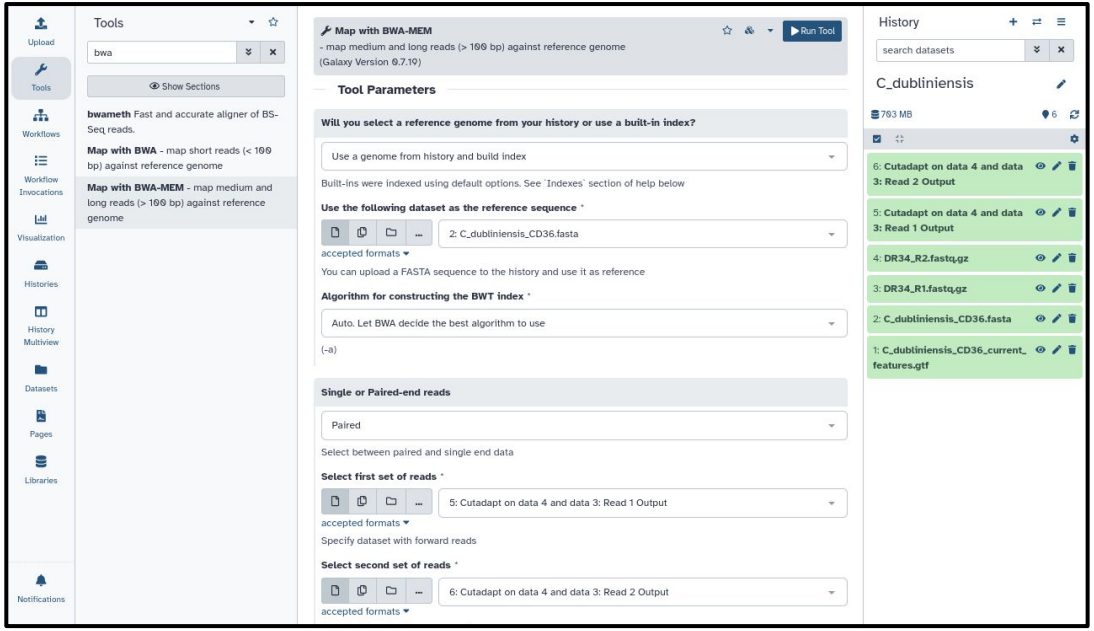


Trim Adapters with Cutadapt



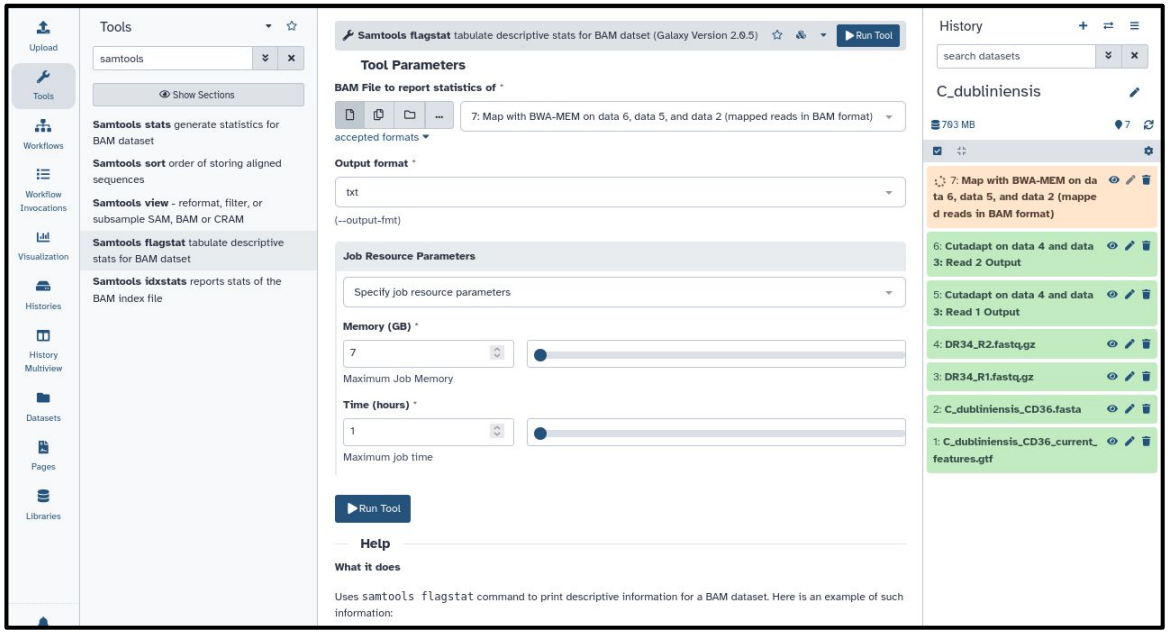
Align Sequences to Reference Genome

- 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)



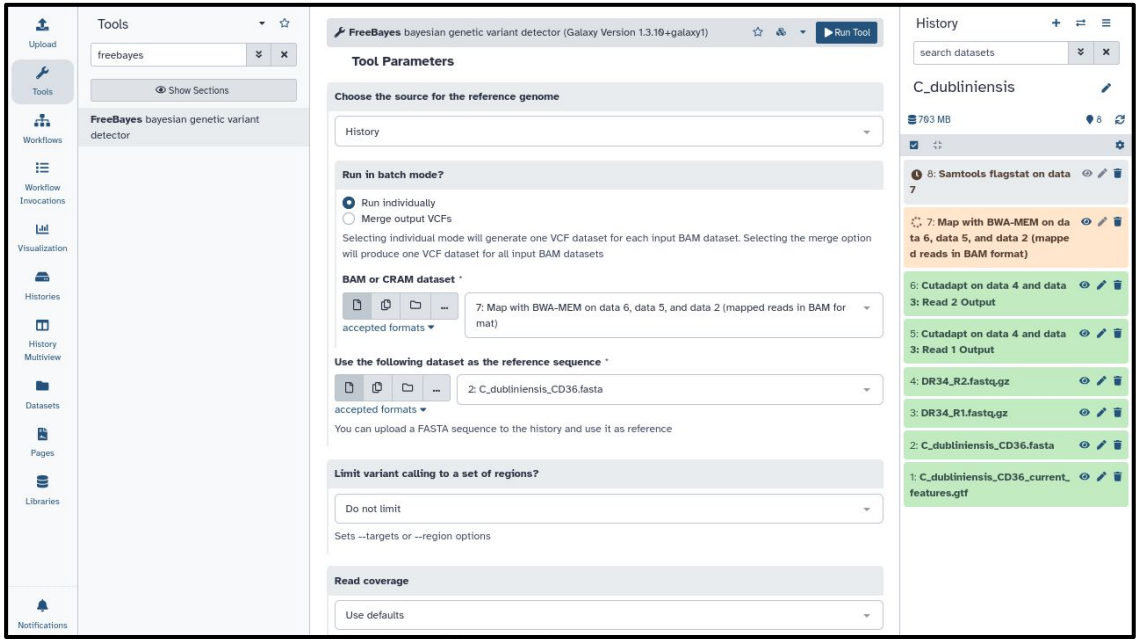
Run Samtools flagstat on the .bam Alignment File

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

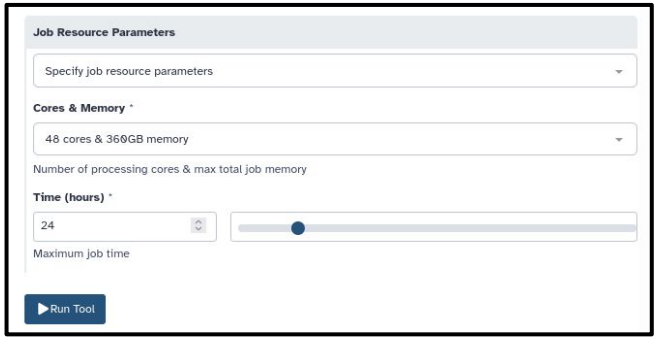


Call Variants with FreeBayes

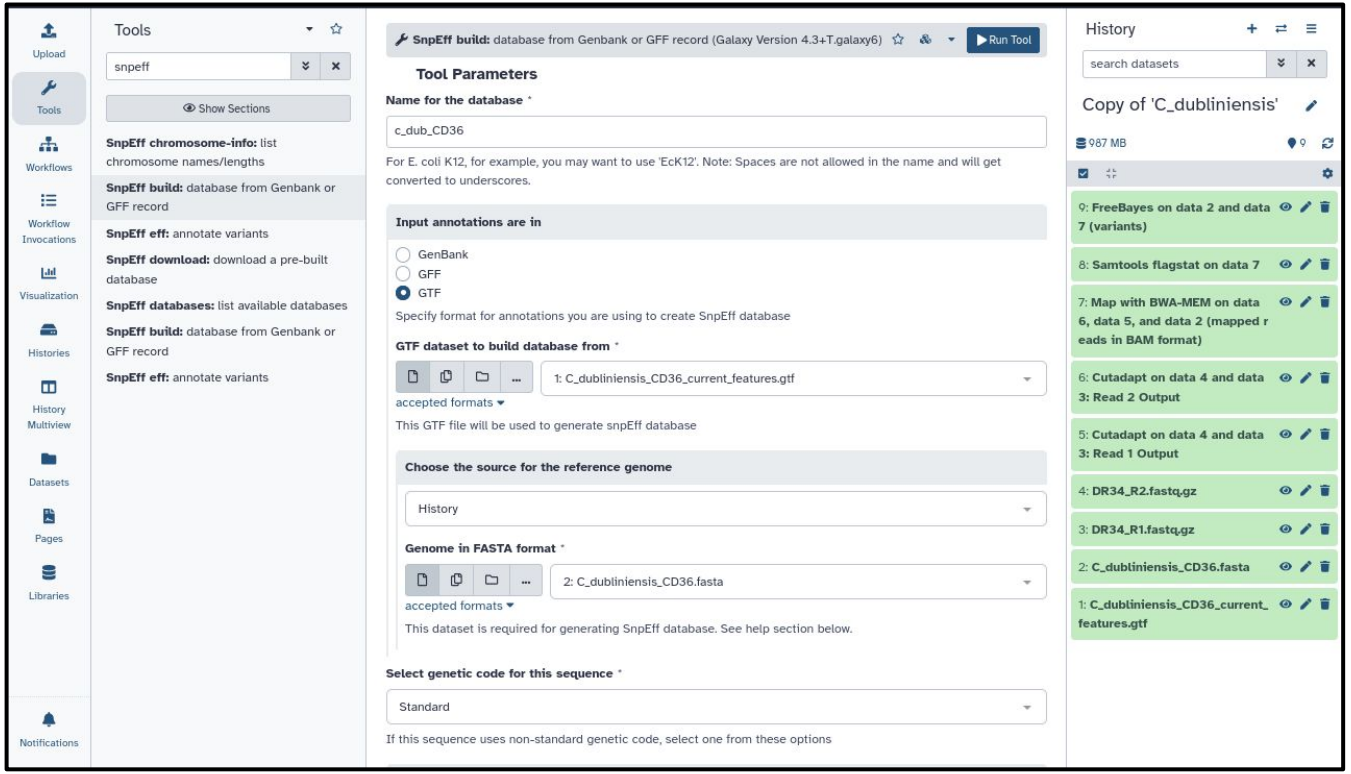
Use the reference genome fasta file from your history



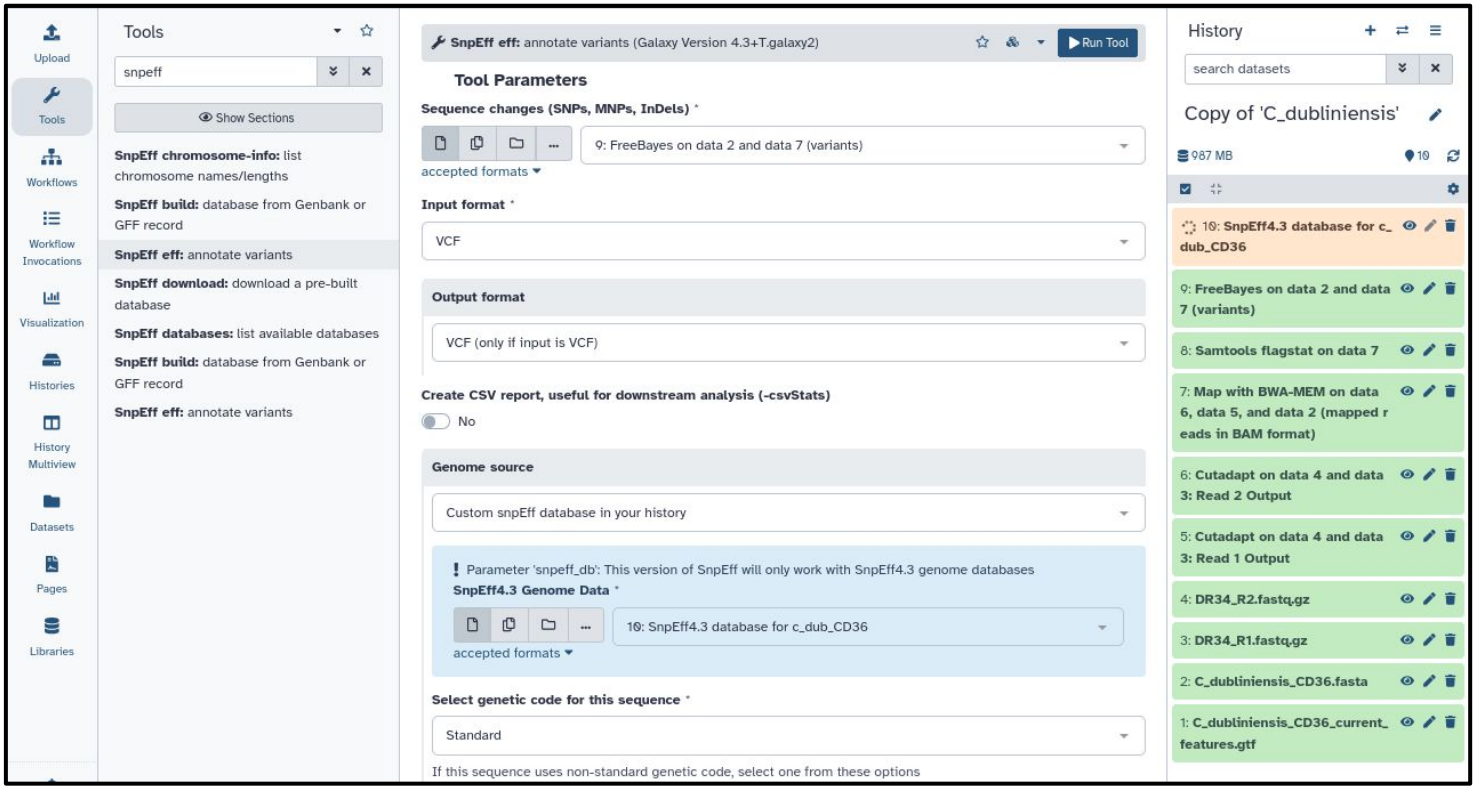
Select 48 cores & 360GB memory



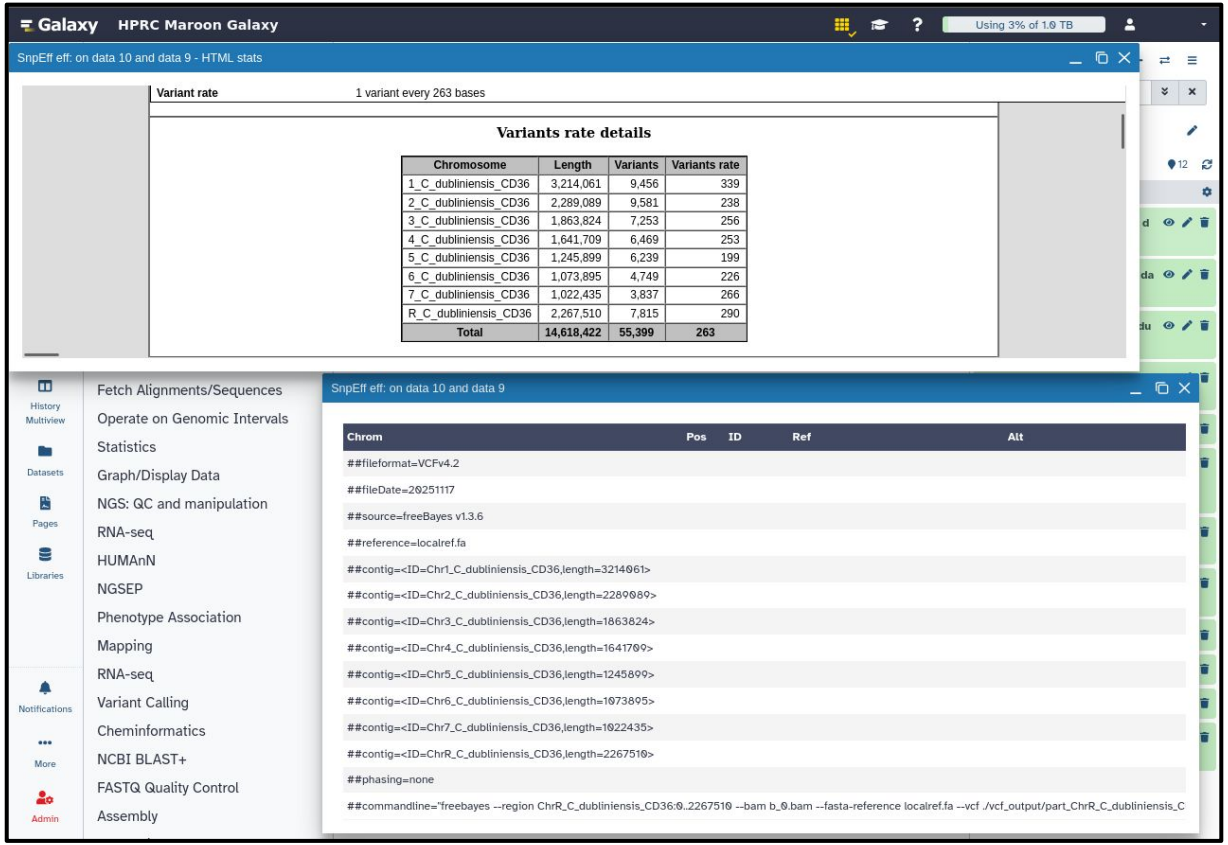
Build a snpEff Database



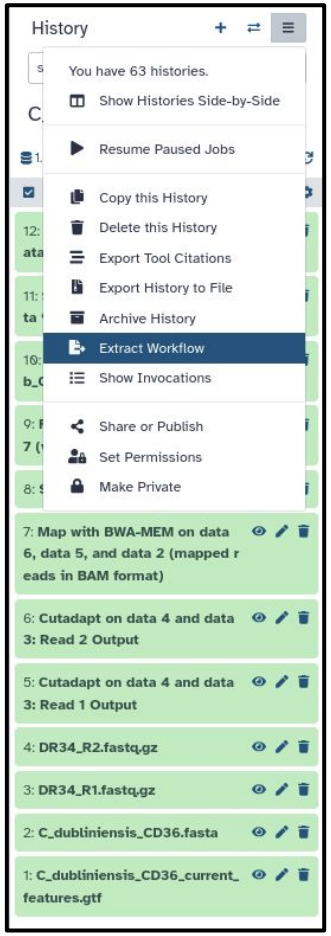
Annotate Variants with snpEff

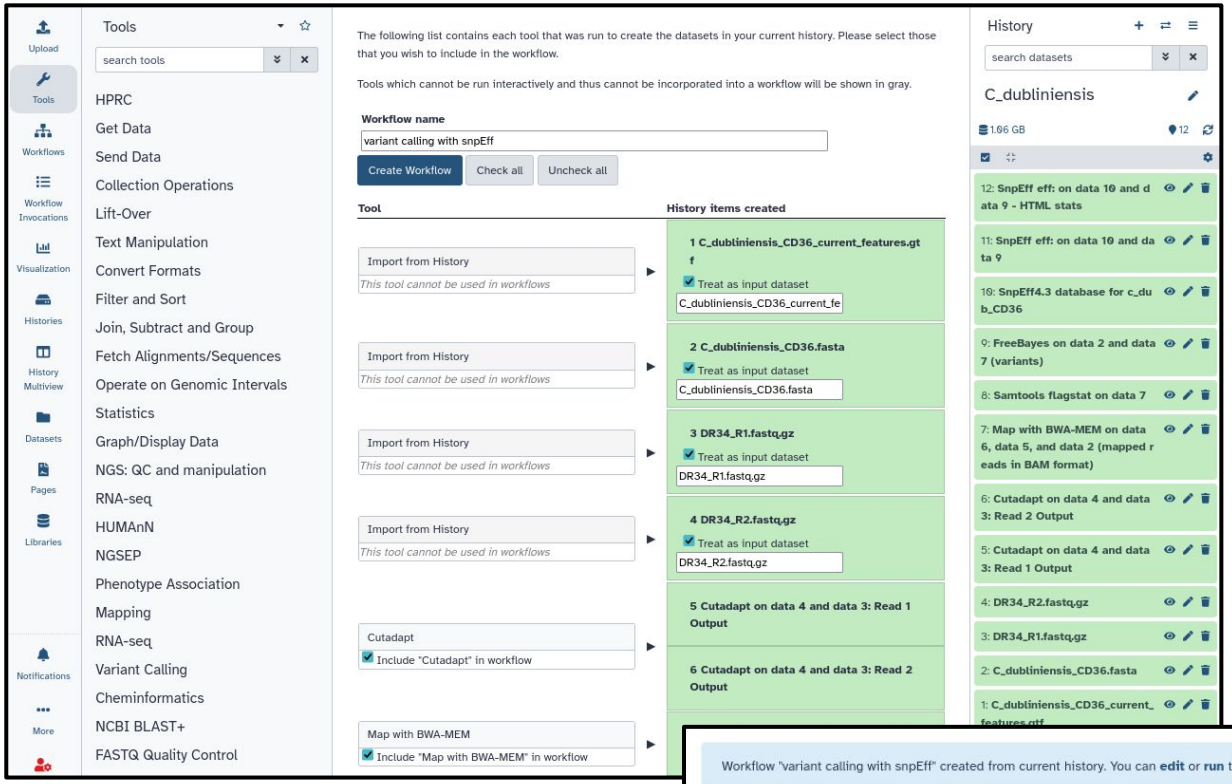


View snpEff Results

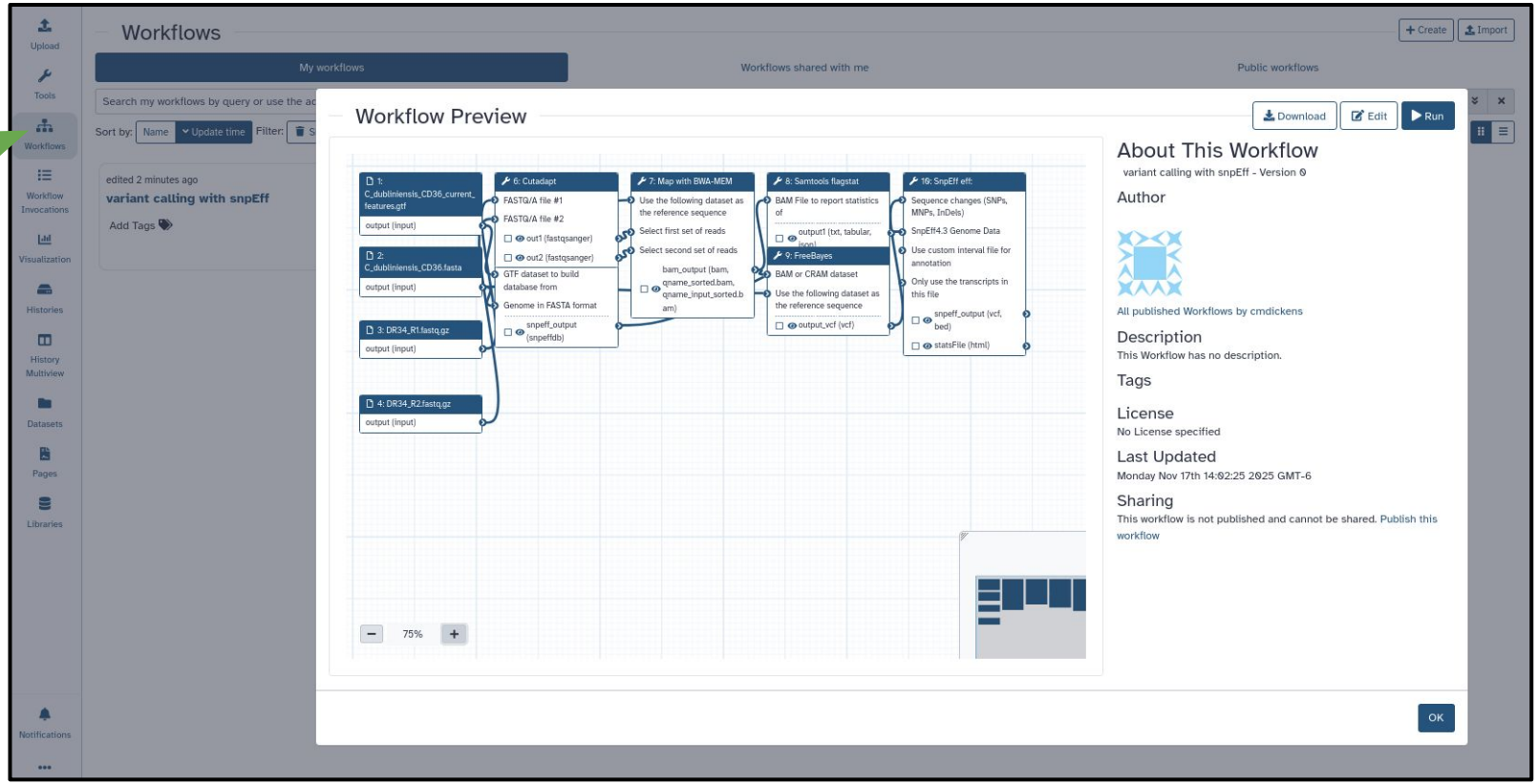


Extract the Workflow



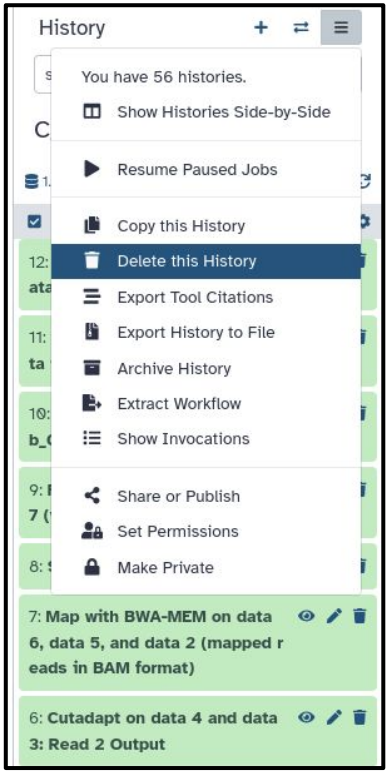


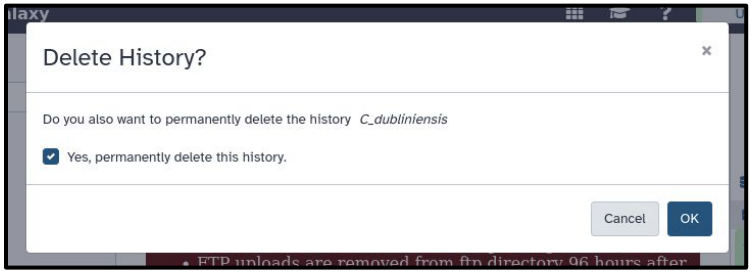
Review the snpEff Workflow



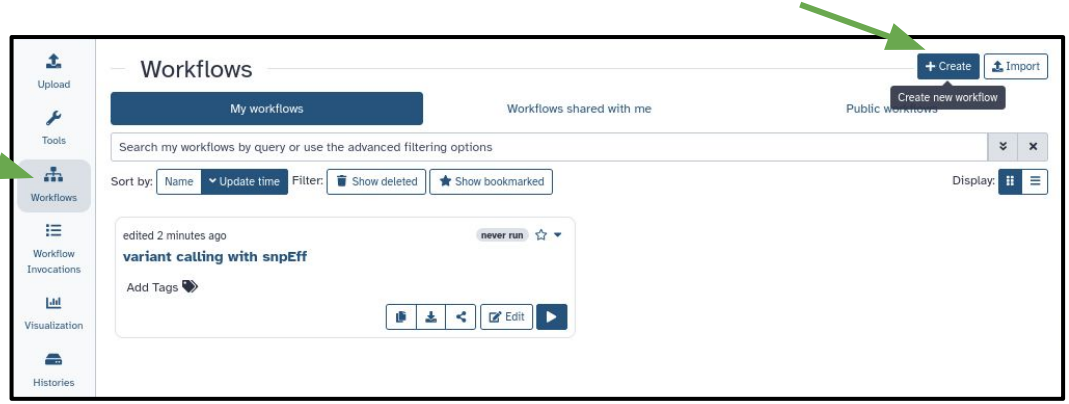
Permanently Delete History

Delete/purge the C_dubliniensis history to reduce disk usage after you successfully create the workflow

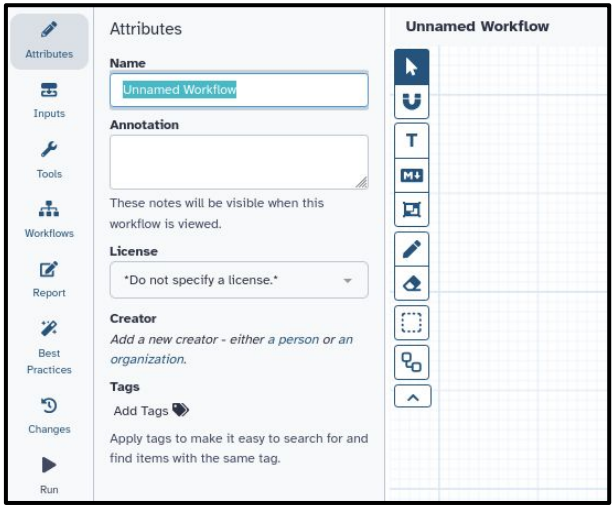




Create a Multi-step Workflow

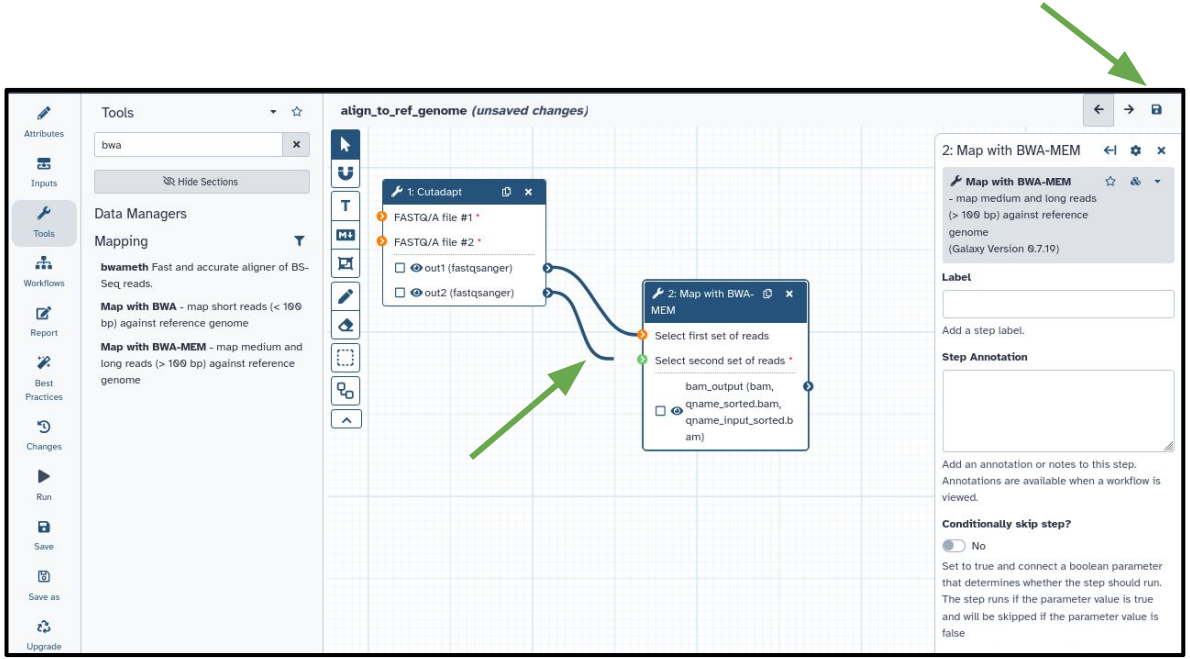


Name the new workflow



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



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

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.