High Performance Research Computing

A Resource for Research and Discovery



HPRC Galaxy Maroon, Silver, Kaiser



Texas A&M University

High Performance Research Computing

Your HPRC Galaxy Username and Password

- Your HPRC Galaxy login is the same as your TAMU account NetID and password
- Both state of Texas law and TAMU regulations prohibit the sharing and/or illegal use of computer passwords and accounts
- Don't write down passwords
- Don't choose easy to guess/crack passwords



For More Help...

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

Help us, help you -- we need more info

- Which Cluster
- •UserID/NetID
- Job id(s) if any
- Location of your jobfile, input/output files
- Application used if any
- Module(s) loaded if any
- •Error messages
- Steps you have taken, so we can reproduce the problem

Texas A&M University High Performance Research Computing

What is the Galaxy Project?

usegalaxy.org

Fools 🖒 📩	Galaxy is an open source, web-b	ased platform for data intensive	History	ଟ+⊡¢
search tools		lew to Galaxy start here or consult our	search datasets	00
Get Data	help resources. You can install ye and choose from thousands of to	our own Galaxy by following the tutorial ools from the Tool Shed.	S. cerevisiae chrM	
Collection Operations	THE REAL	James Taylor (1979-2020)	•	laxy instan ble workflo
Text Manipulation		believed that scientific progress can best be	 shared data 	
Filter and Sort		sustained through the	workflows	
Join, Subtract and Group	990 J	mentoring of students and		-
Datamash		junior faculty.	 many pop 	
GENOMIC FILE MANIPULATION		To ensure implementation of this vision		atic tools a
-ASTA/FASTQ		the Galaxy community has established a foundation—Junior Training and	available	
ASTQ Quality Control	Design by Rebekka Paisner	Educational Connections Hotspot	no progra	Imming
SAM/BAM		(JTech). JTech's mission is to (1) assis graduate students to participate in	knowledg	e required
BED		computational biology and data science	 try usega 	laxy.org to
/CF/BCF		conferences, and (2) organize and hos mentoring sessions between senior and	, ,	axy is a go
VCF/DCF		junior faculty members at high-profile	fit for you	, ,

Texas A&M University

A M

High Performance Research Computing

Galaxy 101

https://galaxyproject.org/learn

Galaxy COMMUNITY HUB

Search Galaxy

Q C Edit

Tutorials by Galaxy Training

Tutorials using Galaxy Main

Tutorials from Lewis-Sigler Institute @ Princeton

Network

Interactive Tours

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

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

Texas A&M University

- Genome Annotation
- Imaging

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

High Performance Research Computing

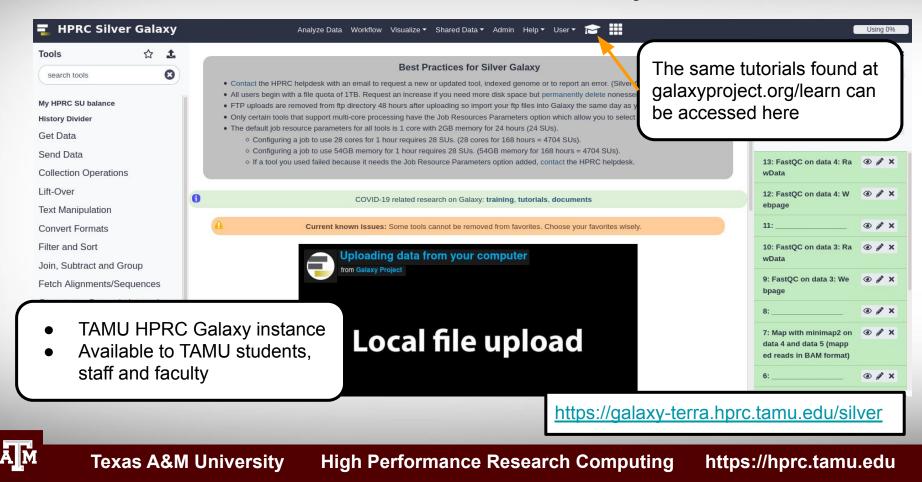
HPRC Galaxy Maroon, Silver, Kaiser



Texas A&M University

High Performance Research Computing

HPRC Silver Galaxy



HPRC Galaxy Notes

HPRC Galaxy Tutorial

Silver Galaxy slides 🖉

Maroon Galaxy slides

Silver Galaxy

Before you request an account on Silver Galaxy, you must do the following:

• Go to usegalaxy.org @ and get familiar with Galaxy. You can start with a free account and learn about Galaxy tools.

- . If you decide that Galaxy is a good choice for your research project then do the following

 - . After your HPRC account is approved, send an email to help@hprc.tamu.edu requesting an account on Silver 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 Silver Galaxy.

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

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 u

Permanently Delete unwanted files

In order to free disk space, you should permanently delete files that you have already deleted from the history.

Please make this part of your Galaxy work routine in order to free up disk space.

This is a two step process.

1. Click the X on the history item you want deleted.

2. At the top of the history panel, you will see a link to deleted files. Click that link and the

hprc.tamu.edu/wiki/SW:Galaxy

https://hprc.tamu.edu

ant remo

Galaxy

Contents [hide]

1 Galaxy

- 1.1 HPRC Galaxy Tutorial
- 1.2 Silver Galaxy
- 1.3 Maroon Galaxy Accounts
- 1.4 Set Your Default Account
- 1.5 Account Security
- 1.6 Permanently Delete unwanted files
- 1.7 FishCamp Galaxy Accounts
- 1.8 Uploading Files > 2GB via FTP to Maroon Galaxy
 - 1.8.1 From a Unix Computer (Mac or Linux)
 - 1.8.2 From one of your directories on Ada
 - 1.8.3 Using BitVise
 - 1.8.4 Using MobaXterm
 - 1.8.5 Using Filezilla
 - 1.8.6 Using WinSCP
 - 1.8.7 Reveille Galaxy
- 1.9 Requesting New Galaxy Tools
 - 1.9.1 When a Galaxy tool is Available
- 1.9.2 When a tool has no Galaxy interface
- 1.10 FAQ
- 1.11 Tool specific notes
 - 1.11.1 Fastq groomer
 - 1.11.2 Trinity
 - 1.11.2.1 Before you run a Trinity job
 - 1.11.2.2 If your Trinity job Fails
 - 1.11.3 RSEM
 - 1.11.4 BLAST
 - 1.11.5 bwa, bowtie, bowtie2 hisat2 1.11.6 HISAT2
- 1.12 Share your History

 1.12.1 with user(s) on the same Galaxy instance
 1.13 Manage your HPRC Accounts

Texas A&M University

High Performance Research Computing

New Galaxy Features

Added & Versions Options 1. You can easily switch to an older software version Switch to 0.7.15.2 Tools 2. You can add tools to your favorites to easily locate later T Favorite Show favorites #favorites Add to favorite Show Sections 3. Disk guota enforced for all users Map with BWA - map short reads (< Using 0% all users begin with 1TB disk guota а. 100 bp) against reference genome request an increase if needed but permanently deleted nonessential files first b. compressed (gzipped) file format is supported C. Tutorials are easy to follow along without leaving Silver Galaxy 4. M 5. Job Resource Parameters available for cores, memory and job time Recource Darameters Job Resource Parameters Specify job resource parameters Specify job resource parameters tools tools Memory (GB) Cores & Memory 28 cores & 54GB memory supporting supporting

Texas A&M University High P

Time (hours)

multi-core

Number of processing cores & max total job memory

High Performance Research Computing

single-core

processing

https://hprc.tamu.edu

Maximum job memory

Time (hours)

Maximum job time.

24

Silver Galaxy

- Try Galaxy at usegalaxy.org to see if it appropriate for your project
- Getting Access to HPRC Silver Galaxy
 - Available to Texas A&M students, staff and faculty with a NetID and an HPRC account
 - Apply for an HPRC account first
 - <u>https://hprc.tamu.edu/apply</u>
 - Then send an email request for a Silver Galaxy account
 - help@hprc.tamu.edu
 - Need to use \underline{VPN} when connecting to Galaxy from off campus
 - Login to Silver Galaxy using your TAMU NetID and password
- Read the Galaxy Usage Notes
 - <u>https://hprc.tamu.edu/wiki/SW:Galaxy</u>
- There are no backups of users' Galaxy files

Look for Additionally Available Tools https://toolshed.g2.bx.psu.edu

7923 valid tools on Oct 31, 2020	Repositories by Category				
Search Search for valid tools	search repository name, descri	iption Q			
/alid Galaxy Utilities	Name	Description	Repositories		
Tools	Assembly	Tools for working with assemblies	145		
Custom datatypes Repository dependency definitions	ChIP-seq	Tools for analyzing and manipulating ChIP-seq data.	73		
Tool dependency definitions All Repositories Browse by category Available Actions	Climate Analysis	Tools for analyzing climate data	5		
	Combinatorial Selections	Tools for combinatorial selection	9		
	Computational chemistry	Tools for use in computational chemistry	159		
Login to create a repository	Constructive Solid Geometry	Tools for constructing and analyzing 3-dimensional shapes and their properties	11		
	Convert Formats	Tools for converting data formats	124		
	Data Export	Tools for exporting data to various destinations	8		
	Data Managers	Utilities for Managing Galaxy's built-in data cache	77		
	Data Source	Tools for retrieving data from external data sources	88		
	Ecology	Tools related to ecological studies	45		
	Entomology	Tools that involve insect studies	4		

A M

Texas A&M University

High Performance Research Computing

Uploading files to your Galaxy History



Texas A&M University

High Performance Research Computing

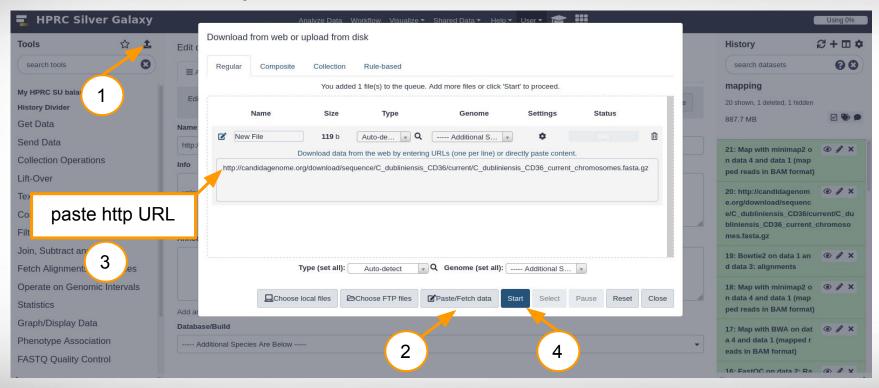
Get Data into Your Galaxy History



- Upload files < 2GB in size using Galaxy web interface
 - \circ select local file on your computer to upload
 - paste URL address of any size file
- Can retrieve data from external websites directly into your Galaxy history with 'Get Data' tools
 UCSC, BioMart, Ratmine, ...



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





High Performance Research Computing

Uploading a 2GB+ size file to Silver Galaxy

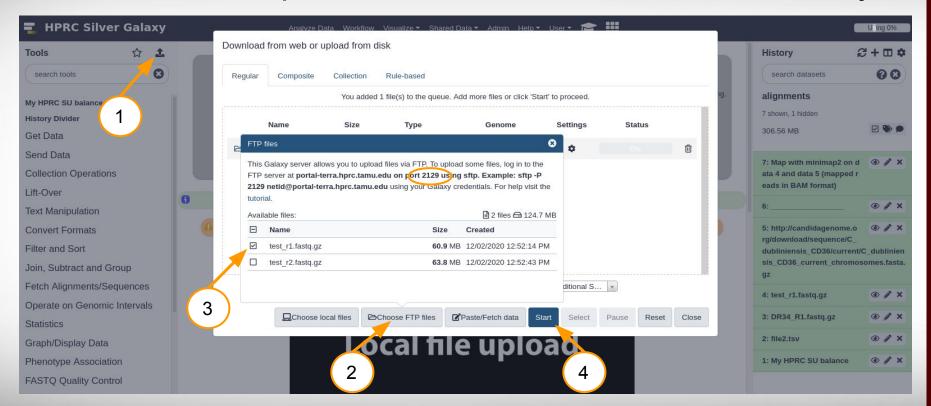
- Files larger than 2GB should be copied using ftp instead of the upload button
- There are three options for uploading files via ftp
 - a. Use the sftp command in a Unix terminal on your Mac or Linux desktop
 - b. Copy files from Terra \$SCRATCH directory to ftp directory using sftp on Terra
 - c. Use sftp on <u>BitVise</u> on your Windows desktop to copy from your desktop to Terra
- 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



High Performance Research Computing

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

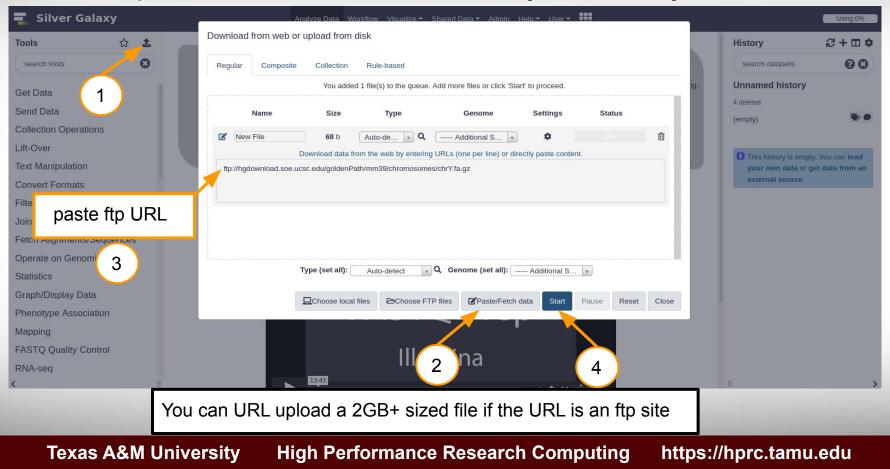


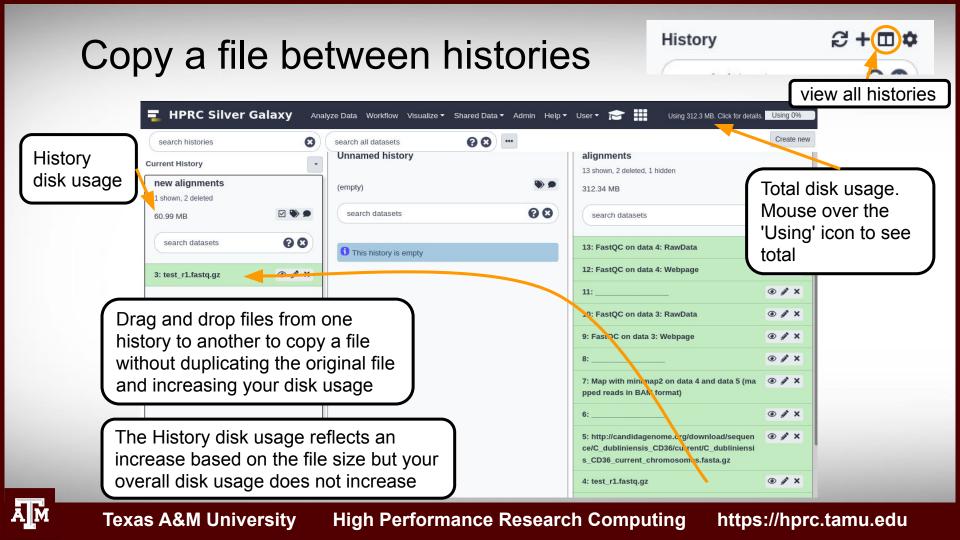


Texas A&M University

High Performance Research Computing

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





File and History Management



Texas A&M University

High Performance Research Computing

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.

11 shown, 3 deleted, 1 hidden	
4.58 GB	
5: https://sra-downloadb.be-md.nc	•
bi.nlm.nih.gov/sos1/sra-pub-run-2/ SRR504368/SRR504368.1 (fastq-dum	p)
581.0 MB	
format: fastqsanger.gz, database: ?	
Read 12038307 spots for /scratch/use	r/galaxy
Read 12038307 spots for /scratch/use /silver/files/000/dataset_135.dat	r/galaxy
/silver/files/000/dataset_135.dat	
/silver/files/000/dataset_135.dat Written 12038307 spots for /scratch/us	
/silver/files/000/dataset_135.dat Written 12038307 spots for /scratch/us /silver/files/000/dataset_135.dat	er/galaxy



High Performance Research Computing

Galaxy File Formats

- Many Galaxy tools require fastqsanger not just fastq
- Galaxy now auto detects and assigns fastqsanger format
- FastQC tool will check fastq format of a fastq file
 - MiSeq, HiSeq, NextSeq, NovaSeq all use 1.8+



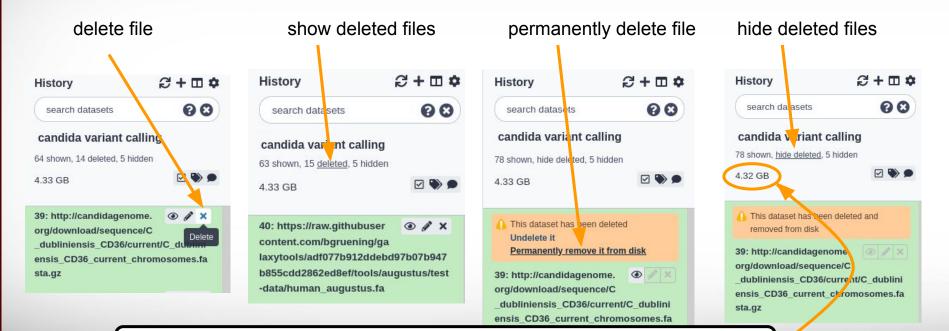


Texas A&M University

High Performance Research Computing

Permanently Delete Nonessential Files

deletes files from disk and reduces your Galaxy disk usage



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



Texas A&M University

High Performance Research Computing

Check Your HPRC SUs Balance

💶 HPRC Silv	ver Gala	axy	Analyze Data Workflow Visualize 🔻 Shared Data 👻 Help 👻 User 💌 📻 🏭		Using 0%
Tools	☆	1	My HPRC SU balance (Galaxy Version 1.0.0)	History	S+0 ¢
search tools		8	The solution of the solution	search datasets	00
My HPRC SU balance			I want to	multi QC	
History Divider			Show my current SU balance	13 shown, 2 deleted, 3 hidden	
Get Data			Job Resource Parameters	296.94 MB	2 🌑 🗩
Send Data			Use default job resource parameters	18:	● # ×
Collection Operatio	ons		✓ Execute		
Lift-Over				16: My HPRC SU balance	● / ×
Text Manipulation			3 This tool retreives a summary of your HPRC SU balance or allows the user to set the default account.	12: MultiQC on data 10, data 8, and data 6: Webp	⊙ # ×
Convert Formats			Run this tool selecting the option 'Show my current SU balance' to get a list of your project account numbers.	age	
Filter and Sort			In the following example, a default account is not set as both accounts have N in the Default column:	11: MultiQC on data 10, da	ata 8, 🗙
Join, Subtract and	Group			and data 6: Stats a list with 3 items	
Fetch Alignments/S	Sequence	s	You can also change your default HPRC project account	10: FastQC on data 3: Ra	
Operate on Genom	nic Interva	ls		wData	
Statistics			000000000001 N 5000.00 -4990.00 10.00 00000000002] N 100000.00 -24555.76 75444.24	9: FastQC on data 3: We	● # ×
Graph/Display Data	a			bpage	
Phenotype Associa	ation		If you are unable to run the HPRC SU balance tool then	8: FastQC on data 2: Ra wData	● / ×
FASTQ Quality Cor	ntrol		you most likely need to <u>renew</u> your HPRC account	7: FastOC on data 2: We	



High Performance Research Computing

History Divider

HPRC Silv	ver Galaxy	Analyze Data Workflow Visualize 🕶 Shared Data 🕶 Admin Help 🕶 User 🖝 📻 🇰		Using 0%	
Tools	☆ 🛓	Histor Divide (Color) Koris 100	History	2 + 🗆 🕈	
search tools	8	History Divider (Galaxy Version 1.0.0)	search datasets	00	
My HPRC SU balance History Divider Get Data	•	spacer character	alignments 13 shown, 1 hidden 312.34 MB	v 🔊 🗩	
Send Data Collection Operati	ons	✓ Execute What it does	13: FastQC on data 4: wData	Ra 💿 🌶 🗙	
Lift-Over Text Manipulation		This tool just adds a spacer between distinct jobs in your Galaxy history panel.	12: FastQC on data 4: ebpage		job with 2 output file
Convert Formats Filter and Sort		Used to add a spacer between output files of distinct job so you can see	11: 10: FastQC on data 3: wData	- • • × Ra • • ×	
Join, Subtract and Fetch Alignments/ Operate on Genoi	Sequences	which files were created by each job	9: FastQC on data 3: W bpage	le 💿 🧳 🗙	job with 2
Statistics Graph/Display Da		Run the History Divider tool between jobs	8: 7: Map with minimap2		output file
Phenotype Associ FASTQ Quality Co	ation	Jona	data 4 and data 5 (map ed reads in BAM forma 6:		job with 1 output file

ĀМ

Texas A&M University

High Performance Research Computing

Shared Data Libraries

Search	+ Folder + Datasets Exp	port to His	Data Libraries	Delete	Details	include deleted	
Libraries / C_dubliniensis_CD	36 data		Histories Workflows Visualizations				
Name ↓ ≜	Description	Da	Pages	Date Update	ed (UTC) Sta	te	
DR34_R1.fastq.gz	uploaded fastqsanger.gz file	fasto	qsanger.gz 162.4 MB	2020-11-09	08:28 PM	٢	😩 Manage

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



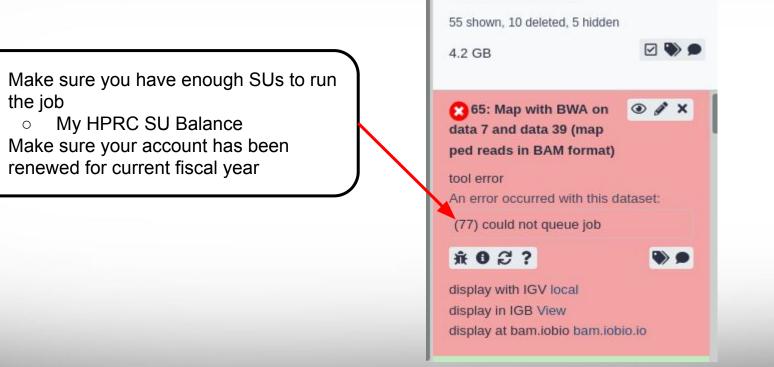
Debugging Failed Jobs



Texas A&M University

High Performance Research Computing

Failed to Queue Job



Unnamed history

Texas A&M University

High Performance Research Computing htt

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.	



High Performance Research Computing

Failed Job: memory limit reached

• Configure the job to use more Memory

Cores & Memory		Memory (G	B)
28 cores & 54GB memory	or	5	
Number of processing cores & max total job memory		Maximum jo	ob memory.



High Performance Research Computing

See Logs of Failed Jobs

	📮 HPRC Silver	Galaxy	Analyze Data Workflow Visualize 🕶 Shared Data 🕶 Help 👻 User 🕶 📻 🏭		Using 0%
	Tools	☆ 1	FastQC	History	2+□¢
	search tools	8	Dataset Information	search datasets	00
the bug icor	^י log by clicking ז ormation icon	g on	Number: 6 Name: FastQC on data 3: RawData Created: Sat Nov 21 19:49:42 2020 (UTC) Filesize: 0 bytes Dbkey: ?	mapping 20 shown, 1 deleted, 1 hidden 887.7 MB	v 🖏 🗩
	tderr file link tdout file link		Format: txt Job Information Galaxy Tool ID: toolshed psu.edu/repos/devteam/fastqc/fastqc/0.72+galaxy1	8: FastQC on data 3: Ra wData 7: FastQC on data 3: We	● / × ● / ×
cause of the the log mess the HPRC he information	Text Manipulation able to determ error after revi ages, send an elpdesk with er Galaxy you are name	ewing email ror	de: None tent API ID: 4er 32c2 (6) 555 3 ba0a8 (5) ID: 96228697dff25cef (1) 0a96847d-33b6-4842-90e7-56fd642505c7 ameters	bpage 6: FastQC on data 3: RawData tool error An error occurred with this $\hat{\mathbf{x}} \odot \hat{\mathbf{c}}$? test_r 2 3: test_r1.fastq.gz 2: DR34_R1.fastq.gz	
History	item number			1: http://candidagenome. org/download/sequence/	Sector Control of Cont

Texas A&M University

Ă M

High Performance Research Computing

usegalaxy.org Exercise

- 1. Add fasta file for S. cerevisiae chromosome I from genome.ucsc.edu to your current history and set the "Database/Build" attribute to the proper genome assembly
- 2. Run the "Compute Sequence Length" tool to get length of chromosome I
- 3. Add fasta file for chromosome M from genome.ucsc.edu to your history
- 4. Concatenate the two files into one file
- 5. Run the "Compute Sequence Length" tool on the concatenated file
- 6. Permanently delete the file you created in step 2.





Thank you.

Any questions?



Texas A&M University

High Performance Research Computing