

High Performance Research Computing

A Resource for Research and Discovery



TEXAS A&M
UNIVERSITY.

Introduction to HPRC Galaxy



Your Login 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
- Change passwords frequently



For More Help...

Website: hprc.tamu.edu

Email: help@hprc.tamu.edu

Telephone: (979) 845-0219

Visit us in person: Henderson Hall, Room 114A

Help us, help you -- we need more info

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



What is the Galaxy Project?

usegalaxy.org

The screenshot shows the Galaxy Project website interface. On the left is a sidebar with a 'Tools' section containing a search bar and a list of tool categories such as 'Get Data', 'Collection Operations', 'Text Manipulation', 'Datamash', 'Convert Formats', 'Filter and Sort', 'Join, Subtract and Group', 'Fetch Alignments/Sequences', 'NGS: QC and manipulation', 'NGS: Deep Tools', 'NGS: Mapping', 'NGS: RNA Analysis', 'NGS: SAMtools', 'NGS: BamTools', 'NGS: Picard', 'NGS: VCF Manipulation', 'NGS: Peak Calling', 'NGS: Variant Analysis', 'NGS: RNA Structure', 'NGS: Du Novo', 'NGS: Gemini', 'NGS: Assembly', 'NGS: Chromosome Conformation', 'NGS: Motif', 'Operate on Genomic Intervals', 'Statistics', 'Graph/Display Data', 'Phenotype Association', 'BEDTools', 'Genome Diversity', 'EMBOSS', 'Regional Variation', 'FASTA manipulation', 'Multiple Alignments', 'Metagenomic Analysis', and 'Multiple regression'. The main content area features a central banner for 'Public Galaxy Servers and still counting' with a '080+' logo. Below this banner are logos for Penn State, Johns Hopkins University, Oregon Health & Science University, TACC, and CyVerse. A tweet from @galaxyproject is displayed, mentioning 'How to use new tagging on individual datasets' and 'Vimeo @Vimeo'. The right sidebar shows a 'History' section with a search bar and a message: 'This history is empty. You can lead your own data or get data from an external source.' The top navigation bar includes 'Analyze Data', 'Workflow', 'Shared Data', 'Visualization', 'Help', 'Login or Register', and 'Using 0%'.

HPRC Galaxy is not a public Galaxy server

- reproducible workflow
- shared data and workflows
- many popular bioinformatic tools are available
- no programming knowledge required



The Galaxy Team is a part of the Center for Comparative Genomics and Bioinformatics, at Penn State, the Department of Biology and at Johns Hopkins University and the Computational Biology Program at Oregon Health & Science University.

This instance of Galaxy is utilizing infrastructure generously provided by the CyVerse at the Texas Advanced Computing Center, with support from the National Science Foundation.

The Galaxy Project is supported in part by NSF, NHGRI, The Huck Institutes of the Life Sciences, The Institute for CyberScience at Penn State, and Johns Hopkins University.



Galaxy 101

galaxyproject.org/tutorials

The screenshot shows the Galaxy Project website interface. At the top, there is a navigation bar with the Galaxy logo and menu items: Use, Community, Education, Deploy & Develop, and Support. A search bar labeled 'Search Galaxy' and an 'Edit' button are also present. The main content area features a notice about the wiki being reorganized, followed by a 'Table of contents' section with a bulleted list of tutorial topics. Below this is an 'Additional resources' section with two links. At the bottom, there is a footer with support information and an 'OPEN CHAT' button.

Galaxy
COMMUNITY HU

Use ▾ Community ▾ Education ▾ Deploy & Develop ▾ Support ▾

Search Galaxy 🔍 Edit

:exclamation: *The Galaxy wiki is in the process of being reorganized and rewritten. This github-based site will serve as a temporary repository for our newest pages. The old wiki is [here](#).*

[Table of contents](#)

[Additional resources](#)

Table of contents

- **Galaxy 101** - explains basics of Galaxy use. This tutorial takes you through downloading human genome annotation from UCSC Table Browser and manipulation of these data to count the number of single nucleotide polymorphisms in exons of protein-coding genes. It end with the creation of a workflow and its use for the analysis of new data. This tutorial consists of two parts:
 - [Part 1](#)
 - [Part 2](#)
- **Introduction to NGS technologies** - a quick overview of next-generation sequencing technologies currently present on the market.
- **Processing many samples at once** - this tutorial explains new feature of Galaxy interface - *dataset collections*. Dataset collections allow you to easily manipulate hundreds of samples in just a few clicks.
- **Diploid variant calling** - this tutorial demonstrates the use of Galaxy for finding sequence variants in diploid genomes.
- **Reference-based RNAseq** - an in depth overview of reference-based RNAseq analysis including methodological details and hand-on explanation of differential gene expression analysis.

Additional resources

- [Freiburg group tutorials](#) - an ever growing collection of Galaxy tutorials that serves as a constant inspiration to this page.
- [Galaxy NGS101](#) - a collection of video tutorials detailing various stages of NGS analysis.

The Galaxy Project is supported in part by NSF, NHGRI, The Huck Institutes of the Life Sciences, The Institute for CyberScience at Penn State, and Johns Hopkins University.

[OPEN CHAT](#)



HPRC Maroon Galaxy

Galaxy Analyze Data Workflow Shared Data Visualization Admin Help User Using 31.0 GB

Tools

search tools

My HPRC SU balance

History divider

Get Data

Text Manipulation

Datamash

Statistics

Filter and Sort

Join, Subtract and Group

Convert Formats

Extract Features

Fetch Alignments/Sequences

Operate on Genomic Intervals

Graph/Display Data

NCBI SRA Tools

Protein tools

Sequence Alignment

FASTA Tools

deepTools

BLAST+

EMBOSS

NGSEP

TAMU HPRC NGS TOOLBOX

NGS: QC and manipulation

NGS: Mapping

NGS: SAMtools

NGS: Picard Tools

NGS: BAMtools

NGS: BEDTools

NGS: Variant Analysis

NGS: de novo assembly

NGS: RNA-seq

NGS: ChIP-seq

NGS: ChV Tools

NGS: Population Analysis

NGS: Metagenomics

Welcome to the HPRC Maroon Galaxy

Notice that Maroon Galaxy will be offline on Tuesday Nov 7 from 9:00 am to 5:00 pm for maintenance. All jobs running at that time will be stopped and you will need to restart your jobs after the maintenance is complete.

Please contact the HPRC helpdesk to request a new tool or indexed genome, report errors or if you just have questions about using Galaxy.

Newest tools added to Maroon Galaxy

My HPRC SU balance Fastq de-interlacer Salmon 0.7.2 StringTie 1.3.3 PICAPD 2.8.3 deepTools 2.5.1 edgeR 3.14.0
Trinity 2.4.0 MACS2 2.1.1.20160309 GATK 3.6 Exonerate 2.4.0 **BUSCO 3.0.2b** InterProScan 5.25-64.0

History

search datasets

hisat2 picard

4 shown

632.4 MB

4: **EstimateLibraryComplexity on data 1: Library complexity report**

3: **EstimateLibraryComplexity on data 1: Library complexity report**

2: **EstimateLibraryComplexity on data 1: Library complexity report**

1: hisat2.bam

<https://hprcgalaxy.tamu.edu/maroon/>

HPRC 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 Ada account
 - Apply for an Ada account first
 - <https://hprc.tamu.edu/apply>
 - Then send an email request for a Maroon Galaxy account
 - help@hprc.tamu.edu
- Read the Galaxy Usage Notes
 - <https://hprc.tamu.edu/wiki/Ada:Galaxy>
- There are no backups of users' Galaxy files
 - You can export a Galaxy history to a single file which can be uploaded to the same or a different Galaxy instance
 - Some custom tools may not be available in all Galaxies



Look for Additionally Available Tools

toolshed.g2.bx.psu.edu

 **Galaxy Tool Shed** Repositories Groups Help User

4807 valid tools on May 16, 2017

Search

- [Search for valid tools](#)
- [Search for workflows](#)

Valid Galaxy Utilities

- [Tools](#)
- [Custom datatypes](#)
- [Repository dependency definitions](#)
- [Tool dependency definitions](#)

All Repositories

- [Browse by category](#)

Available Actions

- [Login to create a repository](#)

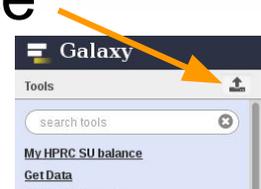
Repositories by Category

Name	Description	Repositories
Assembly	Tools for working with assemblies	88
ChIP-seq	Tools for analyzing and manipulating ChIP-seq data.	47
Combinatorial Selections	Tools for combinatorial selection	8
Computational chemistry	Tools for use in computational chemistry	30
Constructive Solid Geometry	Tools for constructing and analyzing 3-dimensional shapes and their properties	12
Convert Formats	Tools for converting data formats	79
Data Export	Tools for exporting data to various destinations	2

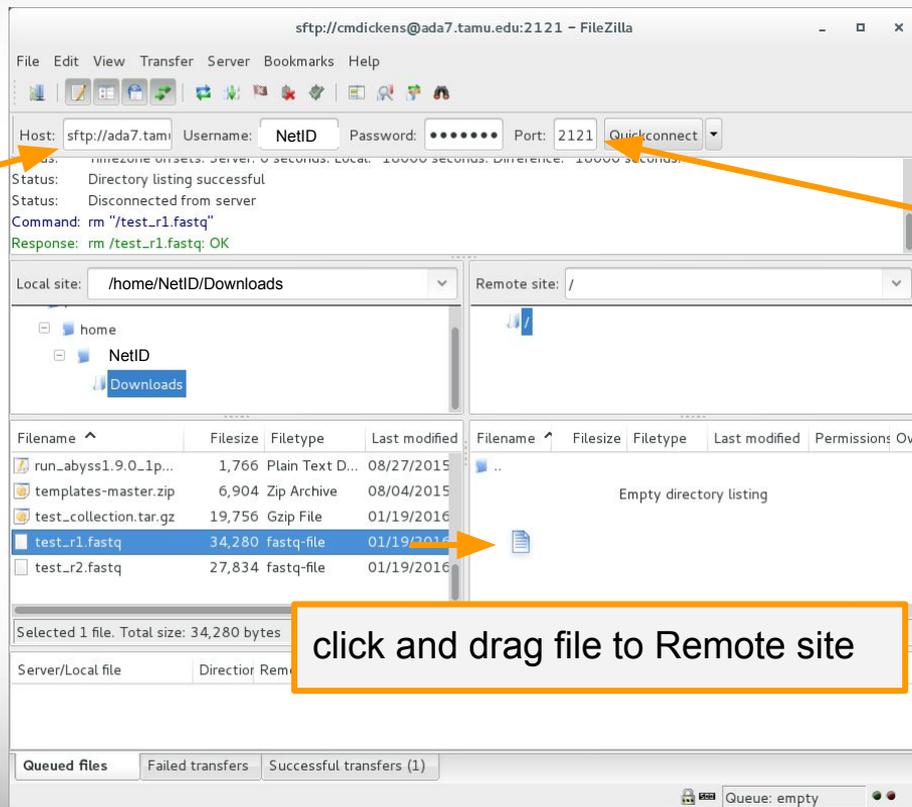


Get Data into Your Galaxy History

- Upload files < 2GB using Galaxy web interface
 - select local file to upload or paste URL
- Upload large files > 2GB
 - using sftp UNIX utility
 - using sftp with locally installed Filezilla or WinSCP
 - paste URL if it is a FTP URL
- Can retrieve data from external websites directly into your Galaxy history with 'Get Data' tools
 - UCSC, BioMart, Ratmine, ...



Copy a 2GB+ desktop file to Galaxy with Filezilla



Host: sftp://ada7.tamu.edu

Port: 2121

for file sizes > 2GB

click and drag file to Remote site

After transferring file to the 'Remote site' directory, go to Galaxy upload file interface to see your ftp transferred file (next slide)

Add Your FTP Uploaded 2GB+ File to Your History

Galaxy

Analyze Data Workflow Shared Data Visualization Admin Help User Using 29.2 GB

Tools

search tools

My HPRC SU balance

Get Data

Text Manipulation

Datamash

Statistics

Filter and Sort

Join, Subtract and Group

Convert Formats

Extract Features

Fetch Alignments/Sequences

History divider

Operate on Genomic Intervals

Graph/Display Data

NCBI SRA Tools

Protein tools

Sequence Alignment

FASTA Tools

deepTools

BLAST+

EMBOSS

NGSEP

TAMU HPRC NGS TOOLBOX

NGS: QC and manipulation

NGS: Mapping

Download data directly from web or upload files from your disk

You can Drag & Drop files into this box.

FTP files

This Galaxy server allows you to upload files via FTP. To upload some files, log in to the FTP server at 'ada7.tamu.edu or login7-ib, port 2121. If you are currently not on ada7, use server name ada7.tamu.edu. If you are on ada7, use server name login7-ib,' using your Galaxy credentials (email address and password).

Available files: 1 files 34.3 KB

<input type="checkbox"/>	Name	Size	Created
<input type="checkbox"/>	test_r1.fastq	34.3 KB	05/24/2017 04:46:18 PM

Type (set all):

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

HTTP URL Upload < 2GB File or Direct Paste

Download data directly from web or upload files from your disk

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

Name	Size	Type	Genome	Settings	Status
New File	74 b	fasta	unspecified (?)		100%

You can tell Galaxy to download data from web by entering URL in this box (one per line). You can also directly paste the contents of a file.

`http://hgdownload.soe.ucsc.edu/goldenPath/sacCer3/chromosomes/chrXVI.fa.gz`

New File	0.3 KB	Auto-detect	unspecified (?)		
----------	--------	-------------	-----------------	--	--

You can tell Galaxy to download data from web by entering URL in this box (one per line). You can also directly paste the contents of a file.

```
@ERR504787.5.1.M00368.15.000000000-A0HKH.1.2.16161.12630-1 length=100
TATTTAAGTGACCAAGGAATGACTCCCAATCATGGCTGATCAACTCCAAAATTTCTGCAACAGCTCGCTGAAATATCTGCAAAATGCCCTTGTGGAA
LEPFC0787.5.1.M00368.15.000000000-A0HKH.1.2.16161.12630-1 length=100
```

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

FTP URL Upload 2GB+ File via URL

Galaxy

Analyze Data Workflow Shared Data Visualization Admin Help User

Using 29.2 GB

Tools

search tools

My HPRC SU balance

History divider

Get Data

Text Manipulation

Datamash

Statistics

Filter and Sort

Join, Subtract

Convert Form

Extract Features

Fetch Alignments/Sequences

Operate on Genomic Intervals

Graph/Display Data

NCBI SRA Tools

Protein tools

Sequence Alignment

FASTA Tools

deepTools

BLAST+

EMBOSS

NGSEP

TAMU HPRC NGS TOOLBOX

Download data directly from web or upload files from your disk

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

Name	Size	Type	Genome	Settings	Status
New File	0.2 KB	vcf	Human Feb. 2009 (G...		

You can tell Galaxy to download data from web by entering URL in this box (one per line). You can also directly paste the contents of a file.

```
ftp://ftp-trace.ncbi.nlm.nih.gov/giab/ftp/data/AshkenazimTrio/HG002_NA24385_son/NIST_HiSeq_HG002_Homogeneity-10953946/HG002Run01-11419412/HG002run1_S1.vcf
```

ftp://

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

Type (set all): Auto-detect

Genome (set all): unspecified (?)

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

Shared Data Libraries

The screenshot shows the Galaxy web interface in Mozilla Firefox. The browser address bar displays <https://hprcgalaxy.tamu.edu/maroon/library/index>. The Galaxy navigation bar includes links for Analyze Data, Workflow, Shared Data (selected), Visualization, Admin, Help, and User. The page title is "Data Library 'fastq files'".

A context menu is open over the "mm10_SRR1261611.fastq" dataset, listing options: Data Libraries, Data Libraries Beta, Published Histories, Published Workflows, Published Visualizations, and Published Pages.

Name	Message	Type	Date uploaded	File size
<input type="checkbox"/> mm10_SRR1261611.fastq	Genomic Ampl		Fri Jan 22 22:42:57 2016 (UTC)	4.5 MB

For selected datasets:

TIP: You can download individual library datasets by selecting "Download this dataset" from the context menu (triangle) next to each dataset's name.

TIP: Several compression options are available for downloading multiple library datasets simultaneously:

- gzip: Recommended for fast network connections
- bzip2: Recommended for slower network connections (smaller size but takes longer to compress)
- zip: Not recommended but is provided as an option for those who cannot open the above formats

Files can be added to a 'Data Library' which you can share with your colleagues. Send a request to help@hprc.tamu if you would like to create a Data Library

Galaxy File Formats

- Fastq Groomer tool will convert fastq to fastqsanger
 - Will convert quality scores to fastqsanger encoding
 - Solexa to fastqsanger
 - Illumina 1.3 - 1.7 to fastqsanger
 - converted file will be an exact copy if the fastq file is already Illumina 1.8+
- If your fastq file is already in fastqsanger format, you can set format in file attributes instead of making a copy which will save you time and also save disk space

Setting Format to Fastqsanger on Upload

Galaxy

Analyze Data Workflow Shared Data Visualization Admin Help User Using 31.0 GB

Tools

search tools

My HPRC SU balance

History divider

Get Data

Text Manipulation

Datamash

Statistics

Filter and Sort

Join, Subtract and Group

Convert Formats

Extract Features

Fetch Alignments/Sequences

Operate on Genomic Intervals

Graph/Display Data

NCBI SRA Tools

Protein tools

Sequence Alignment

FASTA Tools

deepTools

BLAST+

EMBOSS

NGSEP

TAMU HPRC NGS TOOLBOX

Download data directly from web or upload files from your disk

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

Name	Size	Type	Genome	Settings	Status
U00096.2.fa	4.7 MB	Auto-detect	unspecified (?)		

Type (set all):

Genome (set all):

fastq

- fastq
- fastqcssanger
- fastqillumina
- fastqsanger**
- fastqsolexa

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

Setting Format to Fastqsanger in File Attributes

The screenshot displays the Galaxy web interface. At the top, the navigation bar includes 'Galaxy', 'Analyze Data', 'Workflow', 'Shared Data', 'Visualization', 'Admin', 'Help', and 'User'. The main content area is divided into three tabs: 'Attributes', 'Conversion', and 'Datatype' (which is selected and circled with a '2'). Below the 'Datatype' tab, a 'Change data type' dialog is open, showing a search bar and a list of data types. The 'New Type:' field is circled with a '3', and 'fastqsanger' is selected in the list. The history panel on the right shows a list of datasets. The first dataset, '1: mm10_SRR1261611.fastq', is circled with a '1' and has its edit icon highlighted. The dataset's format is shown as 'fastq' and its database as '?'. The second dataset, '4: FastQC on data 1: Webpage', is also circled with a '1'.

Check Your HPRC SUs Balance

Galaxy Analyze Data Workflow Shared Data Visualization Admin Help User Using 31.0 GB

Tools search tools

- My HPRC SU balance** (Galaxy Tool Version 1.0.0) Options
- History divider
- Get Data
- Text Manipulation
- Datamash
- Statistics
- Filter and Sort
- Join, Subtract and Group
- Convert Formats
- Extract Features
- Fetch Alignments/Sequences
- Operate on Genomic Intervals
- Graph/Display Data
- NCBI SRA Tools
- Protein tools
- Sequence Alignment
- FASTA Tools
- deepTools
- BLAST+
- EMBOSS
- NGSEP
- TAMU HPRC NGS TOOLBOX
- NGS: QC and manipulation
- NGS: Mapping
- NGS: SAMtools
- NGS: Picard Tools

My HPRC SU balance (Galaxy Tool Version 1.0.0)

I want to

Show my current SU balance

Set or change my 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 not set as both accounts have N in the Default column:

=====

Account	Default	Allocation	Used & Pending SUs	Balance
0000000000001	N	5000.00	-4990.00	10.00
0000000000002	N	100000.00	-24555.76	75444.24

Then you can set your default account by running this tool using the option 'Set or change my default account' In this example the account 0000000000002 was entered in the field 'Account number to set as default account' Then you will see that the account you specified will be set as the default account to use:

=====

Account	Default	Allocation	Used & Pending SUs	Balance
0000000000001	N	5000.00	-4990.00	10.00
0000000000002	Y	100000.00	-24555.76	75444.24

History search datasets

- hisat2 picard 4 shown 632.4 MB
- 4: EstimateLibraryComplexity on data 1: Library complexity report
- 3: EstimateLibraryComplexity on data 1: Library complexity report
- 2: EstimateLibraryComplexity on data 1: Library complexity report
- 1: hisat2.bam



BLAST+ Multiple Waltime Tools

===== 1 DAY JOBS =====

NCBI BLAST+ blastn 480 SUs.
Search nucleotide database with nucleotide query sequence(s) (max runtime 1 day, 480 SUs required)

NCBI BLAST+ blastp 480 SUs.
Search protein database with protein query sequence(s) (max runtime 1 day, 480 SUs required)

NCBI BLAST+ blastx 480 SUs.
Search protein database with translated nucleotide query sequence(s) (max runtime 1 days, 480 SUs required)

NCBI BLAST+ tblastn 480 SUs
Search translated nucleotide database with protein query sequence(s) (max runtime 1 days, 480 SUs required)

NCBI BLAST+ tblastx 480 SUs.
Search translated nucleotide database with translated nucleotide query sequence(s) (max runtime 1 day, 480 SUs required)

BLAST Reciprocal Search from two FASTA files (max runtime 1 day, 480 SUs required)

===== 3 DAY JOBS =====

NCBI BLAST+ blastn 1440 SUs.
Search nucleotide database with nucleotide query sequence(s) (max runtime 3 days, 1440 SUs required)

NCBI BLAST+ blastp 1440 SUs.
Search protein database with protein query sequence(s) (max runtime 3 days, 1440 SUs required)

NCBI BLAST+ blastx 1440 SUs.
Search protein database with translated nucleotide query sequence(s) (max runtime 3 days, 1440 SUs required)

NCBI BLAST+ tblastn 1440 SUs.
Search translated nucleotide database with protein query sequence(s) (max runtime 3 days, 1440 SUs required)

NCBI BLAST+ tblastx 1440 SUs.
Search translated nucleotide database with translated nucleotide query sequence(s) (max runtime 3 days, 1440 SUs required)

===== 7 DAY JOBS =====

NCBI BLAST+ blastn 3360 SUs.
Search nucleotide database with nucleotide query sequence(s) (max runtime 7 days, 3360 SUs required)

NCBI BLAST+ blastp 3360 SUs.
Search protein database with protein query sequence(s) (max runtime 7 days, 3360 SUs required)

NCBI BLAST+ blastx 3360 SUs.
Search protein database with translated nucleotide query sequence(s) (max runtime 7 days, 3360 SUs required)

NCBI BLAST+ tblastn 3360 SUs.
Search translated nucleotide database with protein query sequence(s) (max runtime 7 days, 3360 SUs required)

NCBI BLAST+ tblastx 3360 SUs.
Search translated nucleotide database with translated nucleotide query sequence(s) (max runtime 7 days, 3360 SUs required)

== 7 DAY 5 NODE JOBS ==

NCBI BLAST+ blastp tamulauncher 16,800 SUs. Search protein database with protein query sequence(s) (max runtime 7 days on 5 nodes, 16,800 SUs required)

NCBI BLAST+ blastn tamulauncher 16,800 SUs. Search nucleotide database with nucleotide query sequence(s) (max runtime 7 days, 5 nodes, 16,800 SUs required)

NCBI BLAST+ blastx 16,800 SUs. Search protein database with translated nucleotide query sequence(s) (max runtime 7 days on 5 nodes, 16,800 SUs required)

NCBI BLAST+ tblastn 16,800 SUs. Search translated nucleotide database with protein query sequence(s) (max runtime 7 days, 5 nodes, 16,800 SUs required)

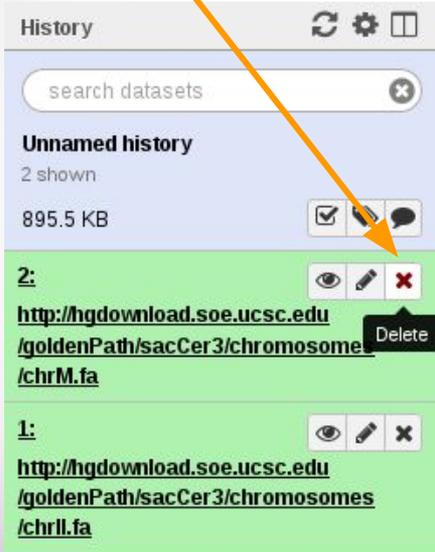
NCBI BLAST+ tblastx 16,800 SUs. Search translated nucleotide database with translated nucleotide query sequence(s) (max runtime 7 days on 5 nodes, 16,800 SUs required)

Galaxy jobs cannot be restarted and checkpoints are not supported



Permanently Delete Unwanted Files

delete file



History

search datasets

Unnamed history
2 shown

895.5 KB

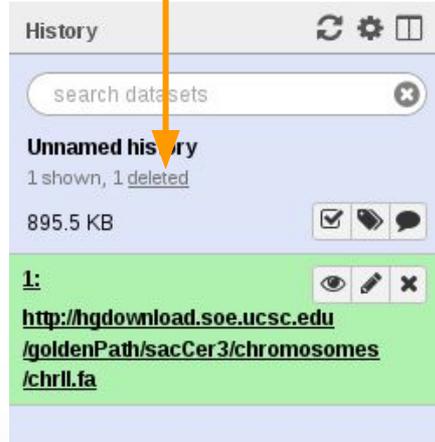
2:
<http://hgdownload.soe.ucsc.edu/goldenPath/sacCer3/chromosomes/chrM.fa>

1:
<http://hgdownload.soe.ucsc.edu/goldenPath/sacCer3/chromosomes/chrll.fa>

Delete

An orange arrow points from the text 'delete file' to the 'Delete' button (an 'x' icon) next to the file entry '2:'.

show deleted files



History

search datasets

Unnamed history
1 shown, 1 deleted

895.5 KB

1:
<http://hgdownload.soe.ucsc.edu/goldenPath/sacCer3/chromosomes/chrll.fa>

An orange arrow points from the text 'show deleted files' to the 'deleted' link in the 'Unnamed history' section.

permanently delete file



History

search datasets

Unnamed history
2 shown, [hide deleted](#)

895.5 KB

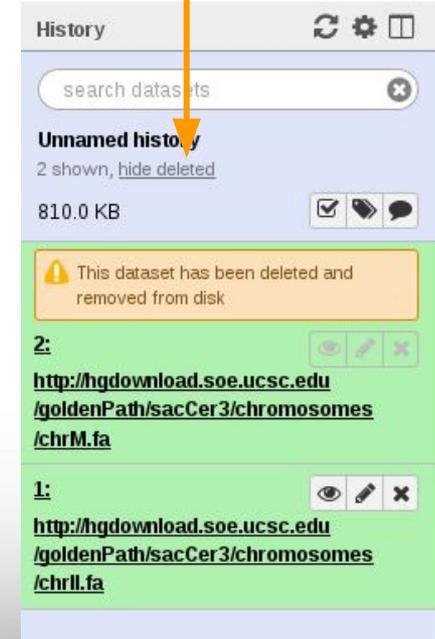
This dataset has been deleted
[Undo it](#)
[Permanently remove it from disk](#)

2:
<http://hgdownload.soe.ucsc.edu/goldenPath/sacCer3/chromosomes/chrM.fa>

1:
<http://hgdownload.soe.ucsc.edu/goldenPath/sacCer3/chromosomes/chrll.fa>

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

hide deleted files



History

search datasets

Unnamed history
2 shown, [hide deleted](#)

810.0 KB

This dataset has been deleted and removed from disk

2:
<http://hgdownload.soe.ucsc.edu/goldenPath/sacCer3/chromosomes/chrM.fa>

1:
<http://hgdownload.soe.ucsc.edu/goldenPath/sacCer3/chromosomes/chrll.fa>

An orange arrow points from the text 'hide deleted files' to the 'hide deleted' link in the 'Unnamed history' section.

History Divider

The screenshot displays the Galaxy web interface. On the left is a 'Tools' sidebar with a search bar and a list of tool categories. An orange arrow points to the 'History divider' tool in the 'Text Manipulation' section. The main panel shows the 'History divider (Galaxy Tool Version 1.0.0)' configuration window. The 'spacer character' is set to a space character, and the 'Execute' button is visible. Below the configuration, a text box explains: 'What it does: This tool just adds a spacer between distinct jobs in your Galaxy history panel.' To the right is the 'History' panel, which lists various jobs. Orange arrows point to specific job entries: 'job 4 files' (job 567), 'job 3 files' (jobs 566, 565, 564), 'job 2 files' (jobs 563, 562), and 'job 1 files' (jobs 561, 560). A central text box explains the tool's purpose: 'Used to add a spacer between distinct job files so you can see which files were created with each job' and 'Run the History Divider tool between jobs'.

Tools

search tools

My HPRC SU balance

History divider

Get Data

Text Manipulation

Datamash

Statistics

Filter and Sort

Join, Subtract and Group

Convert Formats

Extract Features

Fetch Alignments/Sequences

Operate on Genomic Intervals

Graph/Display Data

NCBI SRA Tools

Protein tools

Sequence Alignment

FASTA Tools

deepTools

BLAST+

EMBOSS

NCSEP

TAMU HPRC NGS TOOLBOX

NGS: QC and manipulation

NGS: Mapping

NGS: SAMtools

NGS: Picard Tools

NGS: BAMtools

NGS: BEDTools

NGS: Variant Analysis

NGS: de novo assembly

NGS: RNA-seq

NGS: CHIP-seq

History divider (Galaxy Tool Version 1.0.0)

Options

spacer character

will be 25 characters long

Execute

What it does

This tool just adds a spacer between distinct jobs in your Galaxy history panel.

Used to add a spacer between distinct job files so you can see which files were created with each job

Run the History Divider tool between jobs

History

search datasets

test stacks

240 shown, 288 deleted, 39 hidden

680.2 MB

567: job 4 files

566: velvety on data 561: job 3 files

565: velvety on data 561: job 3 files

564: job 3 files

563: velvety on data 561: job 2 files

562: velvety on data 561: job 2 files

561: velvet on data 1 job 1 files

560: job 1 files

559: velvety on data 557: job 1 files

558: velvety on data 557: job 1 files

557: velvet on data 1 job 1 files

556: job 1 files

555: all_files.zip with STACKS: Process radtags on data 2 and data 1; demultiplexed and cleaned reads job 1 files

cluttered history items

The screenshot shows the Galaxy interface with the 'History divider' tool selected. The 'History' panel on the right lists 235 items, including 'test stacks' (680.2 MB) and various tool runs like '566: velvetq on data 561: Configs'. The items are densely packed, with many lines of text for each entry, making it difficult to read. The 'What it does' section for the tool is also visible, explaining that it adds a spacer between jobs.

divided history items

The screenshot shows the Galaxy interface with the 'History divider' tool selected. The 'History' panel on the right lists 240 items, including 'test stacks' (680.2 MB) and various tool runs like '567:', '566: velvetq on data 561: Configs', and '565: velvetq on data 561: Stats'. The items are more clearly separated and easier to read compared to the cluttered view. The 'What it does' section for the tool is also visible, explaining that it adds a spacer between jobs.

Queued Jobs

- Make sure you have enough SUs to run the job
 - My HPRC SU Balance
- Make sure your account is renewed for current fiscal year
 - My HPRC SU Balance
- Check to see if there is an Ada maintenance scheduled
- Check hprc.tamu.edu to see if Ada node usage is high >95%

The screenshot displays the Galaxy web interface with a job titled "NCBI BLAST+ blastp" in a "Queued" state. The job parameters are as follows:

- Protein query sequence(s):** 1: fasta, Select fasta sequences on data 2 and data 1
- Subject database/sequences:** Locally installed BLAST database
- Protein BLAST database:** nr (09 May 2017)
- Type of BLAST:** blastp - Traditional BLASTP to compare a protein query to a protein database
- Set expectation value cutoff:** 0.001
- Output format:** Tabular (extended 25 columns)
- Advanced Options:** Hide Advanced Options

The job status is "Queued" with a message: "Note. Database searches may take a substantial amount of time. For large input datasets it is advisable to allow overnight processing." The right-hand panel shows the job history, including a search for datasets and a list of jobs, with the current job highlighted in green.

Failed Jobs

- Check the stderr file link
- Check error log
- Read the email you received to see if job requires more memory or more time

TERM_MEMLIMIT: job killed after reaching LSF memory usage limit.

Tools

search tools

[My HPRC SU balance](#)

[History divider](#)

[Get Data](#)

[Text Manipulation](#)

[Datamash](#)

[Convert Formats](#)

[Filter and Sort](#)

[Join, Subtract and Group](#)

[Extract Features](#)

[Statistics](#)

[Graph/Display Data](#)

[Fetch Alignments/Sequences](#)

[Operate on Genomic Intervals](#)

[NCBI SRA Tools](#)

[Protein tools](#)

[FASTA Tools](#)

[BLAST+](#)

[EMBOSS](#)

[Sequence Alignment](#)

[NGSEP](#)

[deepTools](#)

TAMU HPRC NGS TOOLBOX

[NGS: QC and manipulation](#)

[NGS: Mapping](#)

[NGS: SAMtools](#)

[NGS: BAMtools](#)

[NGS: Bismark Tools](#)

Tool: GATK

Name: GATK - AnalyzeCovariates on (log)

Created: Tue Aug 8 18:58:06 2017 (UTC)

Filesize: 1.8 KB

Dbkey: ?

Format: txt

Galaxy Tool ID: gatk

Galaxy Tool Version: 3.6.d9

Tool Standard Error: [stdout](#) [stderr](#)

Tool Exit Code: 0

API ID: ac52ef819012accb

History ID: ed50a...

UUID: 602b65e2-b303-465d-a3dd-11b9806ac7af

Job Runtime (Wall Clock): 28 seconds

Cores Allocated: 10

Job Start Time: 2017-08-08 13:58:25

Job End Time: 2017-08-08 13:58:53

Input Parameter	Value	Note for rerun
Using reference genome	mm9	
Select interval subset to operate on?	false	
Select covariates for on-the-fly recalibration?	false	
Number of data threads to allocate to this analysis	1	
Number of CPU threads to allocate per data thread	20	
Overwrite Memory in MB (0 = don't overwrite)	0	
Analysis Type	AnalyzeCovariates	
...

History

search datasets

gatk

44 shown, 38 deleted

148.0 MB

82: GATK - AnalyzeCovariates on (log)

error

An error occurred with this dataset:

Fatal error: Matched on ##### ERROR
ERROR
ERROR A USER ERROR has c
ERROR
ERROR This means that one o

80:

79: GATK - CombineGVCFs on data 68 and data 29 (log)

78: GATK - CombineGVCFs on data 68 and data 29 (VCF)

75:

Galaxy Notes on Ada

<https://hprc.tamu.edu/wiki/index.php/Ada:Galaxy>

Galaxy

Contents [hide]

- 1 Galaxy
 - 1.1 Account Security
 - 1.2 FishCamp Galaxy Accounts
 - 1.3 Maroon Galaxy Accounts
 - 1.4 Uploading Files > 2GB via FTP to Maroon Galaxy
 - 1.4.1 From a UNIX Computer (Mac or Linux)
 - 1.4.2 Reveille Galaxy
 - 1.5 Requesting New Galaxy Tools
 - 1.5.1 When a Galaxy tool is Available
 - 1.5.2 When a tool has no Galaxy interface
 - 1.6 FAQ
 - 1.7 Tool specific notes
 - 1.7.1 Trinity
 - 1.7.1.1 Before you run a Trinity job
 - 1.7.1.2 If your Trinity job Fails
 - 1.7.2 RSEM
 - 1.7.3 BLAST
 - 1.7.4 bwa, bowtie, bowtie2 hisat2

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.

FishCamp Galaxy Accounts

The FishCamp Galaxy instance is reserved for training purposes such as Galaxy workshops.

When requesting access to FishCamp for a training workshop, please include your ada NetID in your request.

- The FishCamp Galaxy is configured for training purposes.
 - Most jobs will run a maximum of 1 hour.
 - This is to enable jobs to be scheduled faster in the cluster queue.
 - Keep your input datasets small so that they will complete within one hour.

FishCamp Galaxy is not intended for research projects and data on FishCamp Galaxy should be considered to have short term accessibility.

Request a Maroon Galaxy account only if you have data to analyze.

The tools available on FishCamp and Maroon are the same.



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.





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

Thank you.

Any question?

