

Exercise Instructions for Galaxy Tutorial

1. Get data from EBI SRA exercise

- Click on Get data on the left side panel
- Click on EBI SRA
- Enter SRR1016067 on the search box and
- Click search
- Click on Fastq files (galaxy) File1
- repeat steps a-d and select c click on Fastq files (galaxy) File2

The screenshot shows the Galaxy web interface at maxwell.abdn.ac.uk/galaxy. The 'Tools' sidebar on the left has 'Get Data' highlighted with a red box 'a'. Under 'Get Data', 'EBI SRA ENA SRA' is highlighted with a red box 'b'. The main content area shows the EBI SRA search results for SRR1016067. The search box contains 'SRR1016067' and the 'Search' button is highlighted with a red box 'd'. Below the search results, the 'Fastq files (galaxy)' section is visible, with 'File 1' and 'File 2' highlighted with red boxes 'e' and 'f' respectively. The right sidebar shows the 'History' panel with 'Unnamed history' and '0 bytes'.

g. check if you were redirected back to maxwell.abdn.ac.uk/galaxy and not usegalaxy.org, or try again

The screenshot shows the Galaxy web interface with a success message for job 1: EBI SRA. The message states: "The following job has been successfully added to the queue: 1: EBI SRA. You can check the status of queued jobs and view the resulting data by refreshing the History pane. When the job has been run the status will change from 'running' to 'finished' if completed successfully or 'error' if problems were encountered." The 'History' panel on the right shows the job details for '1: EBI SRA: SRR1016067' with a file path: ftp://ftp.sra.ebi.ac.uk/vol1/fastq/SRR101/007/SRR1016067/SRR1016067.1.fastq.gz.

The screenshot shows the Galaxy web interface with a success message for job 2: EBI SRA. The message states: "The following job has been successfully added to the queue: 2: EBI SRA. You can check the status of queued jobs and view the resulting data by refreshing the History pane. When the job has been run the status will change from 'running' to 'finished' if completed successfully or 'error' if problems were encountered." The 'History' panel on the right shows the job details for '2: EBI SRA: SRR1016067' with a file path: ftp://ftp.sra.ebi.ac.uk/vol1/fastq/SRR101/007/SRR1016067/SRR1016067.2.fastq.gz.

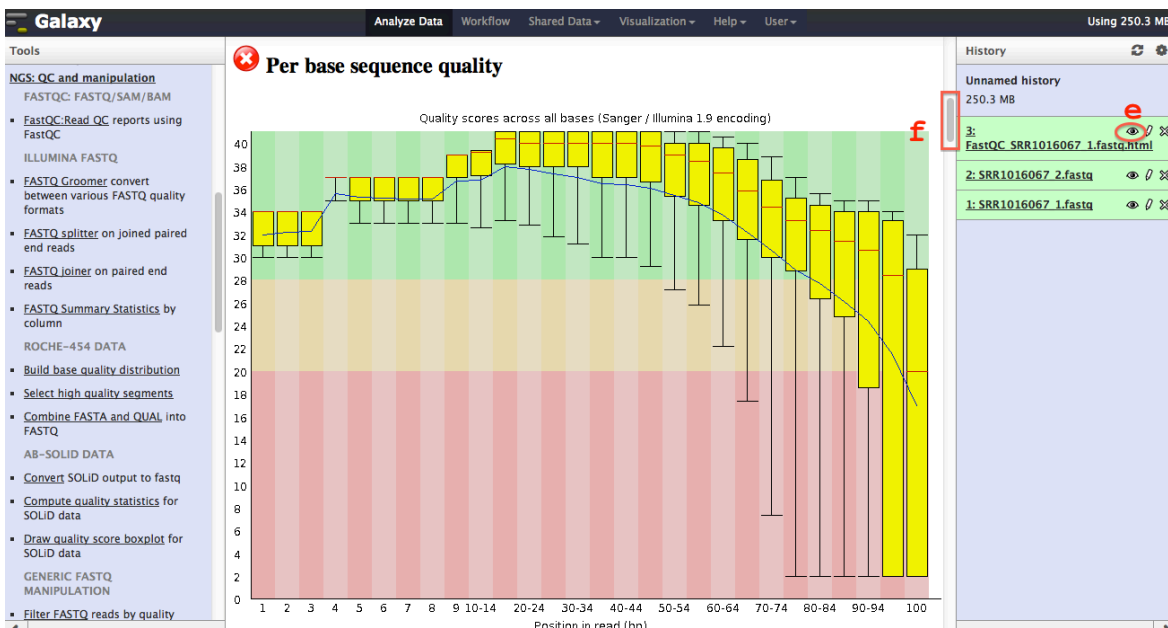
- h. click on the pencil icon for result 1 in history
- i. change the name to SRR1016067_1.fastq (or any name without gz extension)
- j. select "SacCer2" in the Database/Build" dropdown and save
- k. click Save

- l. click on the pencil icon for result 2 in history
- m. change the name to SRR1016067_2.fastq (or any name without gz extension)
- n. select "SacCer2" in the Database/Build" dropdown and save
- o. click Save

2. QA using FastQC exercise

- a. click on NGS:QC and manipulation
- b. click on FASTQC : Reads QC
- c. Select result 1
- d. Click execute

- e. click on the eye icon to visualize the results
- f. scroll down to per base sequence

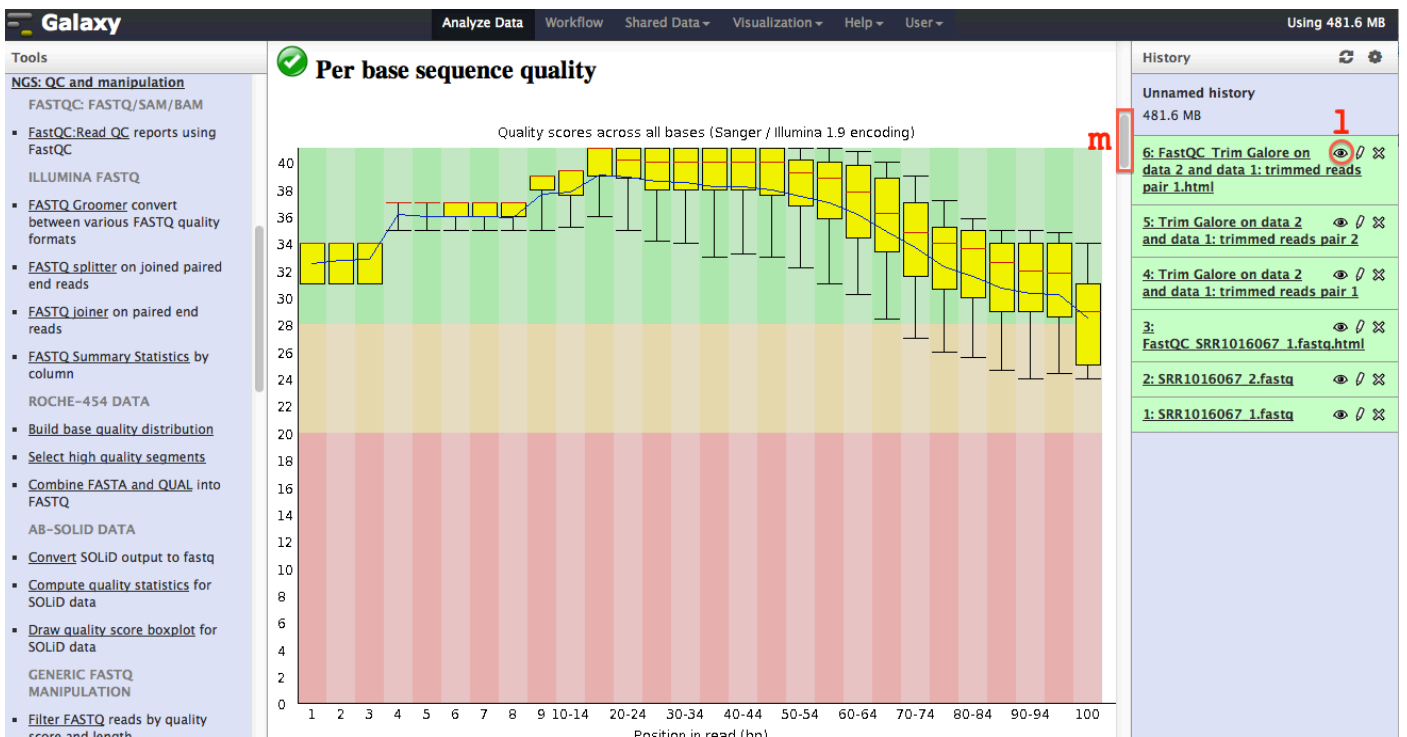


3. Filtering using trimgalore! exercise

- a. click on Trim Galore in the tools panel
- b click on trimgalore sub menu
- c. select paired-end
- d. select result 1
- e. select result 2
- f. check that all parameters are adequate
- g. click on execute

- h. click on NGS: Qc and manipulation
- i. click on FASTQC : Reads QC
- j: Select result 4
- k. Click execute

- l. click on the eye icon to visualize the results
- m. scroll down to per base sequence



4a. Upload data using ftp exercise

- click on Get Data
- click on Upload file
- select File Format fasta
- paste into URL/File `ftp://ftp.ensembl.org/pub/release-59/fasta/saccharomyces_cerevisiae/dna/Saccharomyces_cerevisiae.EF2.59.dna.toplevel.fa.gz`
- select Genome SacCer2
- click Execute

Galaxy Analyze Data Workflow Shared Data Visualization Help User Using 481.6 MB

Tools

search tools

Get Data **a**

- Upload File** **b** from your computer
- UCSC Main table browser
- UCSC Test table browser
- UCSC Archaea table browser
- BL table browser
- EBI SRA ENA SRA
- Get Microbial Data
- BioMart Central server
- BioMart Test server
- CBI Rice Mart rice mart
- GrameneMart Central server
- modENCODE fly server
- Flymine server
- Flymine test server
- modENCODE modMine server
- MouseMine server
- Ratmine server
- YeastMine server
- metabolicMine server
- modENCODE worm server

Upload File (version 1.1.3)

File Format:

fasta **c**

Which format? See help below

File:

Choose File no file selected

TIP: Due to browser limitations, uploading files larger than 2GB is guaranteed to fail. To upload large files, use the URL method (below) or FTP (if enabled by the site administrator).

URL/Text:

`ftp://ftp.ensembl.org/pub/release-59/fasta/saccharomyces_cerevisiae/dna/Saccharomyces_cerevisiae.EF2.59.dna.toplevel.fa.gz` **d**

Here you may specify a list of URLs (one per line) or paste the contents of a file.

Files uploaded via FTP:

File	Size	Date
------	------	------

Please create or log in to a Galaxy account to view files uploaded via FTP.

This Galaxy server allows you to upload files via FTP. To upload some files, log in to the FTP server at '~/.galaxyinput on the cluster or use Alces Portal.' using your Galaxy credentials (email address and password).

Convert spaces to tabs:

☐ Yes

Use this option if you are entering intervals by hand.

Genome:

S. cerevisiae June 2008 (SGD/sacCer2) (sacCer2) **e**

Execute **f**

History

Unnamed history

481.6 MB

- 6: FastQC Trim Galore on data 2 and data 1: trimmed reads pair 1.html
- 5: Trim Galore on data 2 and data 1: trimmed reads pair 2
- 4: Trim Galore on data 2 and data 1: trimmed reads pair 1
- 3: FastQC SRR1016067_1.fastq.html
- 11.5 KB
- format: html, database: sacCer2
- HTML file
- 2: SRR1016067_2.fastq
- 1: SRR1016067_1.fastq

4b. Modify datatype

- click on the pencil icon for Trim Galore trimmed reads pair 1
- select New Type: fastqsanger
- Click Save
- click on the pencil icon for Trim Galore trimmed reads pair 2
- select New Type: fastqsanger
- Click Save

The image displays two screenshots of the Galaxy web interface, specifically the 'Change data type' dialog. The top screenshot shows the 'New Type' dropdown menu set to 'fastqsanger' (labeled 'b') and the 'Save' button (labeled 'c'). The bottom screenshot shows the same dialog with the 'Save' button labeled 'f' and a red 'e' next to the dropdown. Both screenshots show a history panel on the right with entries for 'FastQC Trim Galore on data 2 and data 1: trimmed reads pair 1.html' and 'Trim Galore on data 2 and data 1: trimmed reads pair 2'. Red circles and letters 'a' and 'd' highlight specific elements in the history panel.

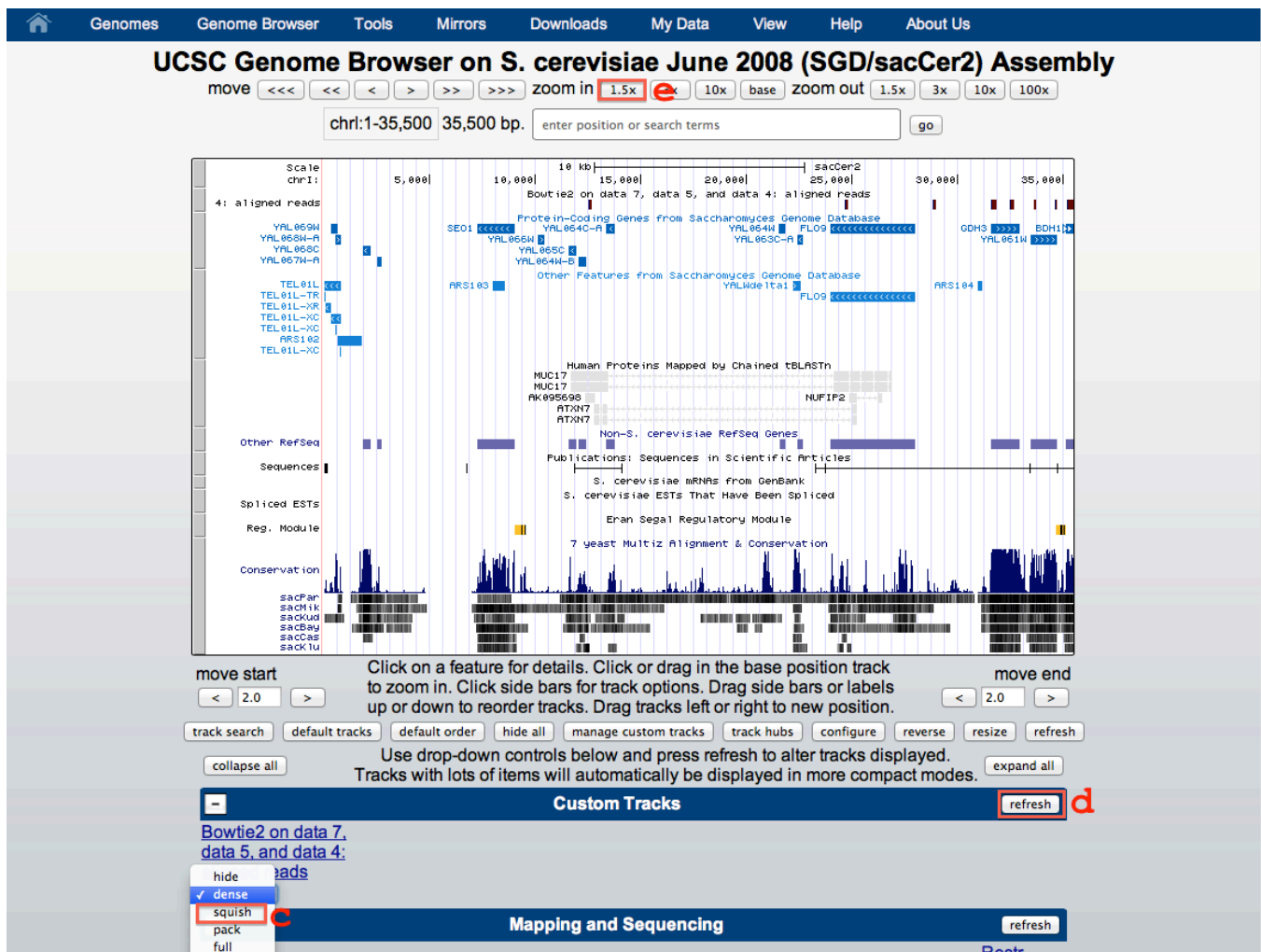
4c. Bowtie alignment

- Click on Bowtie2 category
- Click on Bowtie2
- Select paired-end
- Select Trim Galore trimmed reads pair 1
- Select Trim Galore trimmed reads pair 2
- Select "use one from history"
- Select result 7 (ftp://ftp.ensembl.org/pub/release-59/fasta/saccharomyces_cerevisiae/dna/Saccharomyces_cerevisiae.EF2.59.dna.toplevel.fa)
- Click Execute

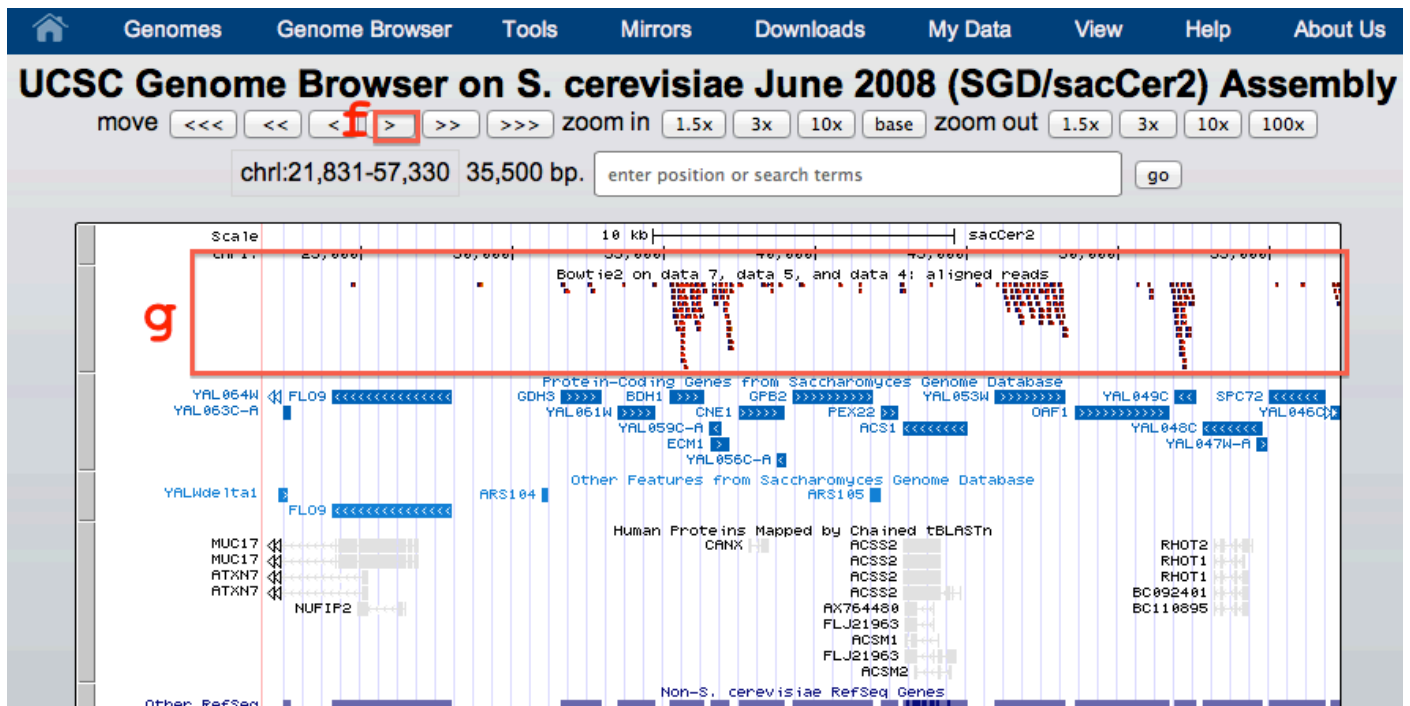
4d. Visualizing results in UCSC

- Click on result 8 link (on the underlined result name “Bowtie 2 on data...”)
- Click on display at UCSC Main

- Select “squished”
- Click “refresh”
- Click Zoom 1.5x

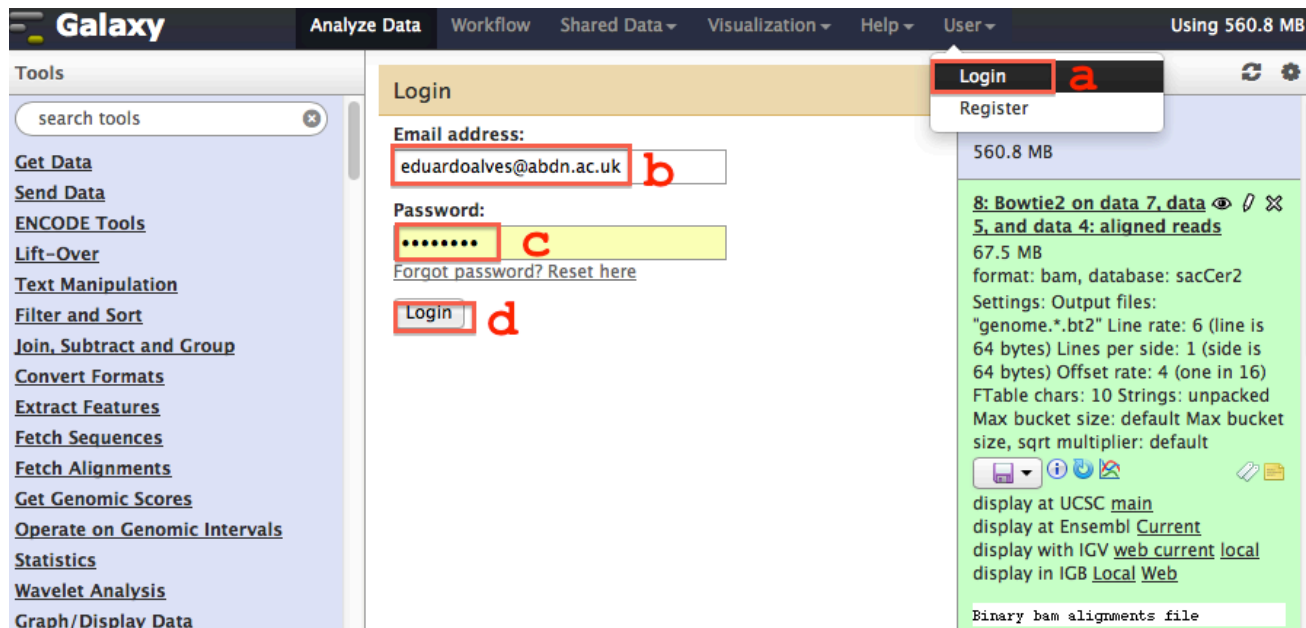


- f. use the buttons to navigate
- g. or right click on the browser to navigate

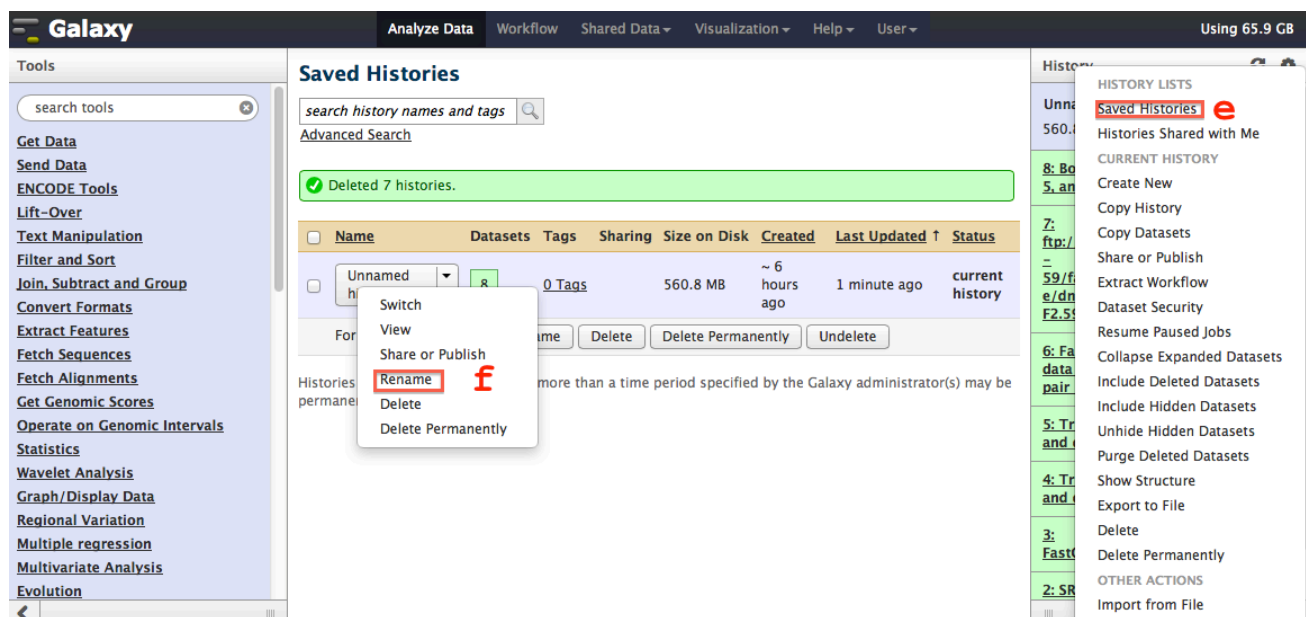


5a. Sharing history and creating workflow exercises (requires Galaxy Account)

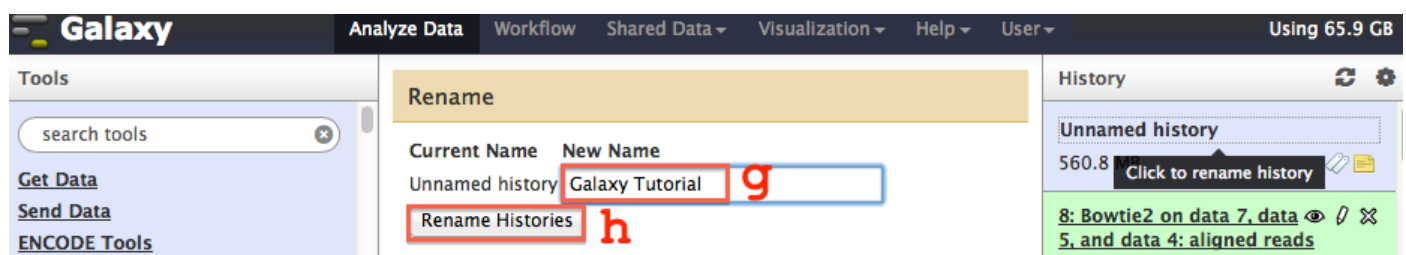
- click on User -> Login
- Enter your email address
- enter your password
- Click "Login"



- Click on saved histories
- Click on rename



- Enter a name for the History
- Click rename history



- click on share or publish

j. click Make accessible and publish

The screenshot shows the Galaxy web interface. On the left is a 'Tools' sidebar with a search bar and various tool categories. The main panel is titled 'Share or Publish History 'Galaxy Tutorial''. It contains sections for 'Make History Accessible via Link and Publish It' and 'Share History with Individual Users'. A red box highlights the 'Make History Accessible and Publish' button, with a red letter 'j' next to it. On the right, a 'History' sidebar is open, showing a list of history items. A red box highlights the 'Share or Publish' button in the history sidebar, with a red letter 'i' next to it.

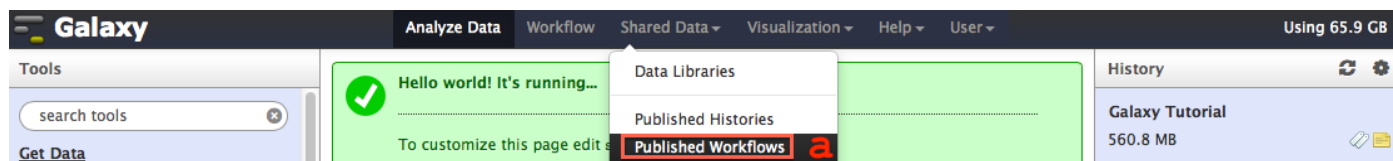
5b. Create workflow exercise (requires Galaxy Account)

- Select Extract Workflow
- Uncheck "treat as input dataset" 3x
- Click on "Create Workflow"

The screenshot shows the Galaxy web interface for creating a workflow. The main panel is titled '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.' It lists tools and their corresponding history items. A red box highlights the 'Create Workflow' button, with a red letter 'c' next to it. A red box highlights the 'Treat as input dataset' checkbox for item 1, with a red letter 'b' next to it. A red box highlights the 'Treat as input dataset' checkbox for item 2, with a red letter 'b' next to it. A red box highlights the 'Treat as input dataset' checkbox for item 7, with a red letter 'b' next to it. On the right, a 'History' sidebar is open, showing a list of history items. A red box highlights the 'Extract Workflow' button in the history sidebar, with a red letter 'a' next to it.

6. Converting BAM to UCSC bedGraph using a public workflow

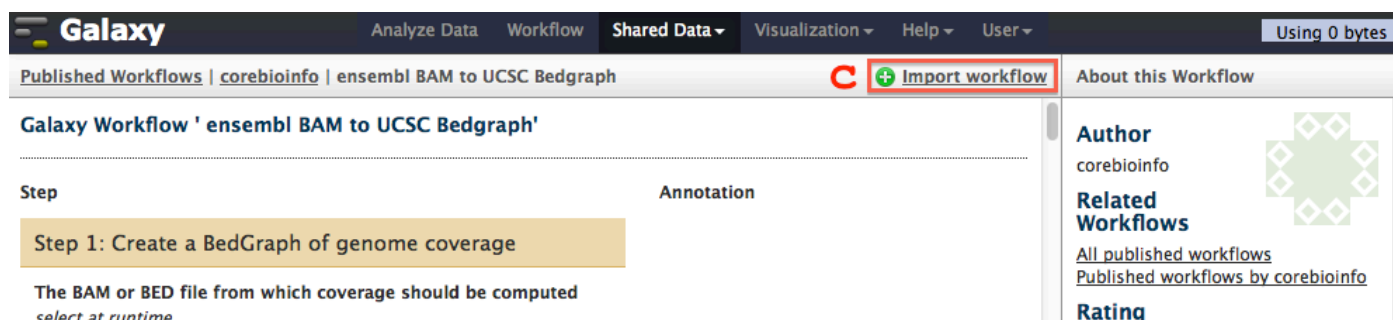
a. click on Shared Data -> Published Workflows



b. Click on Ensembl BAM to UCSC bedGraph



c. Click on Import Workflow



d. Click on imported: ensembl BAM to UCSC bedGraph



- e. Select result 8 (Bowtie 2 on data...")
- f. Click on run Workflow

Galaxy Analyze Data Workflow Shared Data Visualization Help User Using 65.9 GB

Tools

search tools

Get Data
[Send Data](#)
[ENCODE Tools](#)
[Lift-Over](#)
[Text Manipulation](#)
[Filter and Sort](#)
[Join, Subtract and Group](#)
[Convert Formats](#)
[Extract Features](#)
[Fetch Sequences](#)
[Fetch Alignments](#)
[Get Genomic Scores](#)
[Operate on Genomic Intervals](#)
[Statistics](#)
[Wavelet Analysis](#)
[Graph/Display Data](#)
[Regional Variation](#)
[Multiple regression](#)
[Multivariate Analysis](#)
[Evolution](#)
[Motif Tools](#)
[Multiple Alignments](#)
[Metagenomic analyses](#)
[FASTA manipulation](#)
[NGS: QC and manipulation](#)
[NGS: Mapping](#)
[NGS: Indel Analysis](#)
[NGS: RNA Analysis](#)
[NGS: SAM Tools](#)
[NGS: GATK Tools \(beta\)](#)
[NGS: Peak Calling](#)
[NGS: Simulation](#)

Running workflow Expand All Collapse

"imported: ensembl BAM to UCSC Bedgraph"

Step 1: Create a BedGraph of genome coverage (version 0.1.0)

The BAM or BED file from which coverage should be computed
 8: Bowtie2 on data 7..igned reads e

Report regions with zero coverage
 False

Treat split/spliced BAM or BED12 entries as distinct BED intervals when computing coverage.
 False

Calculate coverage based on
 both strands combined

Scale the coverage by a constant factor

Step 2: Create single interval (version 1.0.0)

Step 3: Filter (version 1.1.0)

Step 4: Cut (version 1.0.2)

Step 5: Add column (version 1.0.0)

Step 6: Merge Columns (version 1.0.1)

Step 7: Cut (version 1.0.2)

Step 8: Concatenate datasets (version 1.0.0)

☐ Send results to a new history

Run workflow f

History

Galaxy Tutorial
 560.8 MB

8: Bowtie2 on data 7, data 5, and data 4: aligned reads

7: ftp://ftp.ensembl.org/pub/release-59/fasta/saccharomyces_cerevisiae/dna/Saccharomyces_cerevisiae.EF2.5.9.dna.toplevel.fa

6: FastQC Trim Galore on data 2 and data 1: trimmed reads pair 1.html

5: Trim Galore on data 2 and data 1: trimmed reads pair 2

4: Trim Galore on data 2 and data 1: trimmed reads pair 1

3: FastQC SRR1016067 1.fastq.html

2: SRR1016067 2.fastq

1: SRR1016067 1.fastq

- g. click on the pencil icon
- h. Select Datatype tab
- i. Select New Type: bedGraph and press Save

Galaxy Analyze Data Workflow Shared Data Visualization Help User Using 66.0 GB

Tools

search tools

Attributes **Convert Format** **Datatype** h **Permissions**

Change data type

New Type:

tabular

axt
 bam
 bcf
 bed
 bedgraph i
 bgzip
 bigbed
 bigwig
 hmn

dataset but not modified by guessed the type of

History

Galaxy Tutorial
 655.5 MB

16: Concatenate datasets on data 12 and data 15 g

15: Cut on data 14

14: Merge Columns on data 13

13: Add column on data 11

12: Cut on data 10

11: Filter on data 9

10: Create single interval

9: Bowtie2 on data 7, data 5, and data 4: aligned reads (Genome Coverage BedGraph)

- j. Click on the result link
- k. Click on display at UCSC Main

The screenshot shows the Galaxy web interface. At the top, there's a navigation bar with 'Galaxy' logo and tabs for 'Analyze Data', 'Workflow', 'Shared Data', 'Visualization', 'Help', and 'User'. The 'Analyze Data' tab is active. On the left, there's a 'Tools' sidebar with a search bar and links for 'Get Data', 'Send Data', 'ENCODE Tools', 'Lift-Over', 'Text Manipulation', 'Filter and Sort', 'Join, Subtract and Group', and 'Convert Formats'. The main area shows a workflow step 'Concatenate datasets on data 12 and data 15' with a green status bar indicating it was changed to a bedgraph. Below this, there's an 'Edit Attributes' section with a 'Name' field containing 'Concatenate datasets on data 12 and' and an 'Info' field. On the right, there's a 'History' sidebar showing a 'Galaxy Tutorial' with a size of 655.5 MB. A red box highlights the step '16: Concatenate datasets on data 12 and data 15' in the history, with a red 'j' next to it. Below this, there's a red box around the text 'display at UCSC main' with a red 'k' next to it.

- l. click on full
- m. click refresh

The screenshot shows the UCSC Genome Browser interface for the S. cerevisiae June 2008 (SGD/sacCer2) Assembly. The top navigation bar includes links for 'Genomes', 'Genome Browser', 'Tools', 'Mirrors', 'Downloads', 'My Data', 'View', 'Help', and 'About Us'. The main title is 'UCSC Genome Browser on S. cerevisiae June 2008 (SGD/sacCer2) Assembly'. Below the title, there's a 'move' section with navigation buttons (move, zoom in, zoom out) and a 'zoom in' section with buttons (1.5x, 3x, 10x, base, 1.5x, 3x, 10x, 100x). A search bar contains 'chr1:12,510-35,633' and '23,124 bp'. The main display area shows a genomic track with various features: 'User Track', 'Protein-Coding Genes', 'Other Features from Saccharomyces Genome Database', 'Human Proteins Mapped by Chained tBLASTn', 'Non-S. cerevisiae RefSeq Genes', 'Publications: Sequences in Scientific Articles', 'S. cerevisiae mRNAs from GenBank', 'S. cerevisiae ESTs That Have Been Spliced', 'Eran Segal Regulatory Module', '7 yeast Multiz Alignment & Conservation', and 'Conservation'. The bottom section contains a 'Custom Tracks' area with a 'refresh' button and a 'collapse all' button. A red box highlights the 'full' option in the 'Custom Tracks' area, with a red '1' next to it.

n. click on the browser area to scroll along the genome

