

# What is Galaxy?



From Galaxy training

<https://galaxyproject.github.io/training-material/>



## Data Intensive *analysis* for everyone

- Versatile and reproducible workflows
- Web platform
- Open source under [Academic Free License](#)
- Developed at Penn State, Johns Hopkins, OHSU and Cleveland Clinic with substantial outside contributions



# Core values

- Accessibility**

- Users without programming experience can easily upload/retrieve data, run complex tools and workflows, and visualize data

- Reproducibility**

- Galaxy captures information so that any user can understand and repeat a complete computational analysis

- Transparency**

- Users can share or publish their analyses (histories, workflows, visualizations)
- Pages: online Methods for your paper

# Galaxy growth

- More than 7,000 ready to use tools for users
- More than 7,500 [citations](#)
- More than 350 [public Galaxy resources](#)
  - 120+ public servers, many more non-public
  - Both general-purpose and domain-specific

User Interface

# Main Galaxy interface

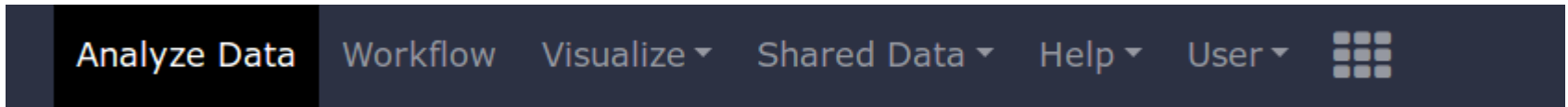
Home page divided into 3 panels

The screenshot displays the Galaxy web interface in a browser window. The interface is divided into three main panels:

- Tools Panel (Left):** A sidebar with a search bar and a list of tool categories. The categories include: Get Data, Send Data, Lift-Over, Collection Operations, Text Manipulation, Datamash, Convert Formats, Filter and Sort, Join, Subtract and Group, Fetch Alignments/Sequences, NGS: QC and manipulation, NGS: DeepTools, 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: Mothur, Operate on Genomic Intervals, Statistics, Graph/Display Data, Phenotype Association, BEDTools, Genome Diversity, EMBOS, Regional Variation, FASTA manipulation, Multiple Alignments, and Metagenomic Analysis. The word "Tools" is written in large blue letters over the list.
- Main Panel (Center):** The central area contains a welcome message: "Galaxy is an open source, web-based platform for data intensive biomedical research. If you are new to Galaxy start here or consult our help resources. You can install your own Galaxy by following the [tutorial](#) and choose from thousands of tools from the [Tool Shed](#)." Below this is a section titled "Running Your Own Understanding how Galaxy works" with the subtitle "An in-depth tutorial". To the right of this section is a "Tweets by @galaxyproject" feed showing two tweets. At the bottom of the main panel, there are logos for Penn State, Johns Hopkins University, Oregon Health & Science University, TACC, and CyVerse. Below the logos, text states: "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." and "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."
- History Panel (Right):** A sidebar titled "History" with a search bar. It contains a "Galaxy introduction (empty)" section and a message: "This history is empty. You can load your own data or get data from an external source". The word "History" is written in large green letters over the panel.

The browser window shows the URL "https://usegalaxy.org" and the page title "Galaxy". The top navigation bar includes links for "Analyze Data", "Workflow", "Shared Data", "Visualization", "Help", and "User". The top right corner shows "Using 0%".

# Top menu



Link	Usage
<i>Analyze Data</i>	go back to the homepage
<i>Workflow</i>	access existing workflows or create new one using the editable diagrammatic pipeline
<i>Visualize</i>	create new visualisations and launch Interactive Environments
<i>Shared data</i>	access data libraries, histories, workflows, visualizations and pages shared with you
<i>Help</i>	links to Galaxy Help Forum (Q&A), Galaxy Community Hub (Wiki), and Interactive Tours
<i>User</i>	your preferences and saved histories, datasets, pages and visualizations

# Tools

- The tool search helps in finding a tool in a crowded toolbox

The screenshot displays the Galaxy web interface. The top navigation bar includes 'Galaxy', 'Analyze Data', 'Workflow', 'Shared Data', 'Visualization', 'Help', 'User', and a 'Using 0%' status indicator. The left sidebar contains a 'Tools' section with a search bar and a list of tool categories: NGS: Peak Calling, NGS: Variant Analysis, NGS: Du Novo, NGS: Mothur, Operate on Genomic Intervals, Graph/Display Data, and Genome Diversity. The 'Join the intervals of two datasets side-by-side' tool is highlighted in the search results. The main panel shows the tool's configuration page, titled 'Join the intervals of two datasets side-by-side (Galaxy Version 1.0.0)'. The configuration includes two dataset inputs: '1: Exons' and '2: SNPs'. The 'Return' dropdown is set to 'Only records that are joined (INNER JOIN)'. A 'Execute' button is at the bottom. A tip below the button states: 'TIP: If your dataset does not appear in the pulldown menu, it means that it is not in interval format. Use "edit attributes" to set chromosome, start, end, and strand columns.' The right sidebar shows the 'History' section with a search bar and a list of datasets: 'Galaxy 101', '2: SNPs', and '1: Exons'. The '2: SNPs' dataset is highlighted in green.

**Galaxy** Analyze Data Workflow Shared Data Visualization Help User Using 0%

**Tools**

Join the intervals of two datasets side-by-side (Galaxy Version 1.0.0) Options

**Join**

First dataset: 1: Exons

with

Second dataset: 2: SNPs

with min overlap

1 (bp)

**Return**

Only records that are joined (INNER JOIN)

Execute

**TIP:** If your dataset does not appear in the pulldown menu, it means that it is not in interval format. Use "edit attributes" to set chromosome, start, end, and strand columns.

**Screencasts!**

See Galaxy Interval Operation [Screencasts](#) (right click to open this link in another window).

**Syntax**

- **Where overlap** specifies the minimum overlap between intervals that allows them to be joined.
- **Return only records that are joined** returns only the records of the first dataset that join to a record in the second dataset. This is analogous to an INNER JOIN.
- **Return all records of first dataset (fill null with ".")** returns all intervals of the first dataset, and any intervals that do not join an interval from the second dataset are filled in with a period(.). This is analogous to a LEFT JOIN.
- **Return all records of second dataset (fill null with ".")** returns all intervals of the second dataset, and any intervals that do not join an interval from the first dataset are filled in with a period(.). **Note that this may produce an invalid interval file, since a period(.) is not a valid chrom, start, end or strand.**
- **Return all records of both datasets (fill nulls with ".")** returns all records from both datasets, and fills on either the right or left with periods. **Note that this may produce an invalid interval file, since a period(.) is not a valid chrom, start, end or strand.**

**History**

search datasets

**Galaxy 101**  
2 shown, 5 deleted

9.06 MB

**2: SNPs**

**1: Exons**



# Tool interface

- A tool form contains:
  - input datasets and parameters
  - help, citations, metadata
  - an `Execute` button to start a job, which will add some output datasets to the history
- New tool versions can be installed without removing old ones to ensure reproducibility

The screenshot shows the NCBI BLAST+ tool interface in Galaxy. The title bar indicates the tool name and version: "NCBI BLAST+ blastp Search protein database with protein query sequence(s) (Galaxy Version 0.3.1)". There are links for "Versions" and "Options".


The main form contains several sections:

- Protein query sequence(s)**: A dropdown menu with a file upload icon and a folder icon. The text "No fasta or fasta.gz dataset available." is displayed. Below the dropdown is the label "(-query)".
- Subject database/sequences**: A dropdown menu showing "Locally installed BLAST database".
- Protein BLAST database**: A section with a checkbox labeled "Select/Unselect all" and an empty text input field below it.
- Type of BLAST**: A section with three radio button options:
  - ☒ blastp - Traditional BLASTP to compare a protein query to a protein database
  - ☐ blastp-short - BLASTP optimized for queries shorter than 30 residues
  - ☐ blastp-fast - Use longer words for seeding, faster but less accurateBelow the options is a link: "See help text for default parameter values for each BLAST type. (-task)".
- Set expectation value cutoff**: A text input field containing "0.001". Below the field is the label "(-evalue)".
- Output format**: A dropdown menu showing "Tabular (extended 25 columns)". Below the dropdown is the label "(-outfmt)".
- Advanced Options**: A dropdown menu showing "Hide Advanced Options".

At the bottom of the form is a blue button with a checkmark icon and the text "Execute".

# Tool Shed

- Free "app" store: [Galaxy Tool Shed](#)
  - Thousands of tools already available
  - Most software can be integrated
    - If a tool is not available, ask the Galaxy community for help!
  - Only a Galaxy admin can install tools

 **Galaxy Tool Shed**

Repositories Groups Help ▾ User ▾

6532 valid tools on Dec 04, 2018

**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
<a href="#">Assembly</a>	Tools for working with assemblies	128
<a href="#">ChIP-seq</a>	Tools for analyzing and manipulating ChIP-seq data.	65
<a href="#">Combinatorial Selections</a>	Tools for combinatorial selection	10
<a href="#">Computational chemistry</a>	Tools for use in computational chemistry	76
<a href="#">Constructive Solid Geometry</a>	Tools for constructing and analyzing 3-dimensional shapes and their properties	12
<a href="#">Convert Formats</a>	Tools for converting data formats	114
	Tools for exporting data to various	~

# History

- Location of all analyses
  - collects all datasets produced by tools
  - collects all operations performed on the data
- For each dataset (the heart of Galaxy's reproducibility), the history tracks
  - name, format, size, creation time, datatype-specific metadata
  - tool id, version, inputs, parameters
  - standard output (`stdout`) and error (`stderr`)
  - state (waiting, running, success, failed)
  - hidden, deleted, purged

The screenshot shows the 'History' panel in Galaxy 101. At the top, there's a search bar labeled 'search datasets'. Below it, the panel title is 'Galaxy 101' with '7 shown' and a size of '9.07 MB'. There are icons for checkmark, refresh, and chat. The main content area lists several datasets and their associated tools. The first dataset is '7: Compare two Datasets on data 6 and data 1', which has 5 regions and is in 'bed' format using the 'hg38' database. It shows the source code for 'join (GNU coreutils) 8.22'. Below this, there are links to 'display in IGB View', 'display with IGV local Human hg38', and 'display at UCSC main test'. A table of genomic data is shown with columns: 1.Chrom, 2.Start, 3.End, 4.Name. The table contains five rows of data. Below the table, there are more datasets listed: '6: Select first on data 5', '5: Sort on data 4', '4: Group on data 3', '3: Join on data 2 and data 1', and '2: SNPs'. Each dataset entry has icons for view, edit, and delete.

**History**

search datasets

**Galaxy 101**  
7 shown  
9.07 MB

**7: Compare two Datasets on data 6 and data 1**  
5 regions  
format: **bed**, database: **hg38**

join (GNU coreutils) 8.22  
Copyright (C) 2013 Free Software Foundation, Inc.  
License GPLv3+: GNU GPL version 3 or later <<http://gnu.org/licenses/gpl.html>>.  
This is free software: you are free to change and redistribute it.  
There is NO WARRANTY, to the ext

display in IGB [View](#)  
display with IGV [local](#) [Human hg38](#)  
display at UCSC [main](#) [test](#)

1.Chrom	2.Start	3.End	4.Name
chr22	46256560	46263322	uc003bhh.
chr22	15690077	15690709	uc010gqp.
chr22	15528158	15529139	uc011agd.
chr22	15690245	15690709	uc062bek.
chr22	22376182	22376505	uc062cbs.

**6: Select first on data 5**

**5: Sort on data 4**

**4: Group on data 3**

**3: Join on data 2 and data 1**

**2: SNPs**

# Multiple histories

- You can have as many histories as you want
- each history should correspond to a **different analysis**
- and should have a meaningful **name**

The screenshot displays the Galaxy Europe web interface, which is designed for managing and analyzing biological data. At the top, a dark navigation bar includes the 'Galaxy / Europe' logo, a search bar for 'search histories', and a dropdown menu for 'search all datasets'. The main content area is divided into four vertical panels, each representing a different 'history' of analyses.

**Current History:** This panel is titled 'Workflow extract error' and contains a list of datasets. It shows 6 shown, 16 deleted, and 3 hidden datasets, with a total size of 10.83 KB. The datasets listed include '24: data 7 (flattened)', '23: Venn on collection 1: svg', '22: Venn on collection 1: sharedotus', '5: Venn on collection 1: svg', '4: Venn on collection 1: sharedotus', and '1: Sub.sample on data 76: subsample.shared'.

**Unnamed history:** This panel is titled 'Unnamed history' and contains a list of datasets. It shows 86 shown, 3 deleted, and 44 hidden datasets, with a total size of 910.45 MB. The datasets listed include '127: Heatmap.sim on collection 86: heatmap.sim.svg', '119: Plotting tool on collection 83', '113: Classify.seqs on data 48, data 9, and others: tree.sum', '112: Classify.seqs on data 48, data 9, and others: tax.summary', '87: Rarefaction.single on data 79: rarefaction curves', '86: Dist.shared on data 76: dist files', '85: Summary.single on data 76: summary', '84: Summary.single on data 76: ave-std.summary', '83: Rarefaction.single on data 76: rarefaction curves', and '82: Sub.sample on data 76: subsample.shared'.

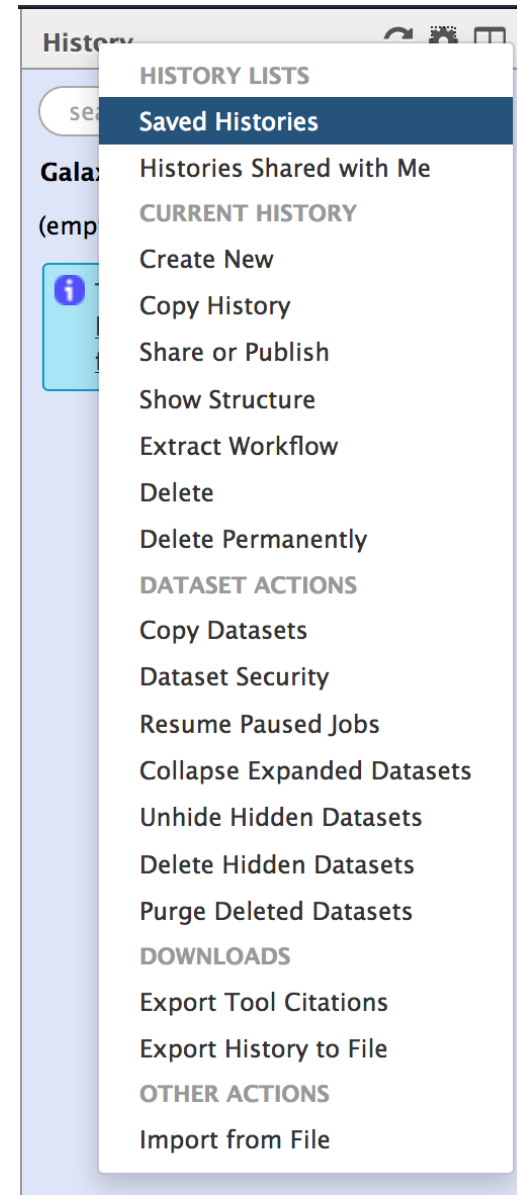
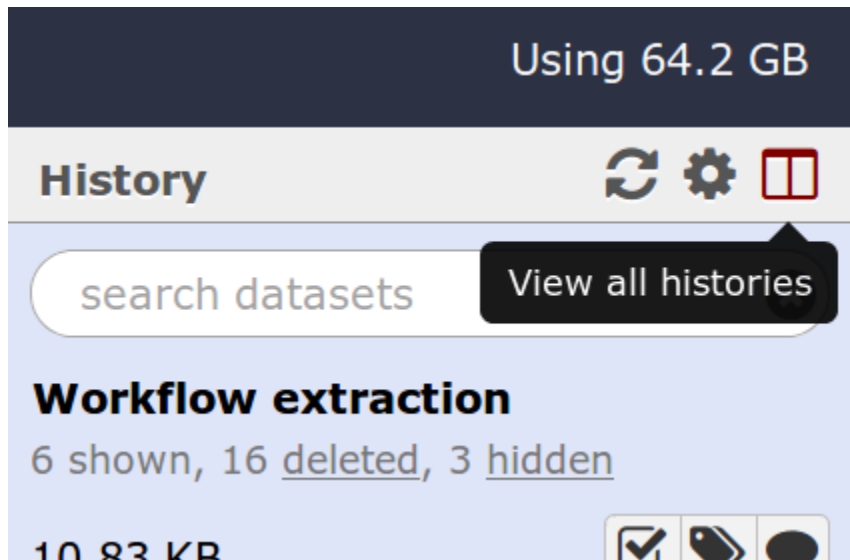
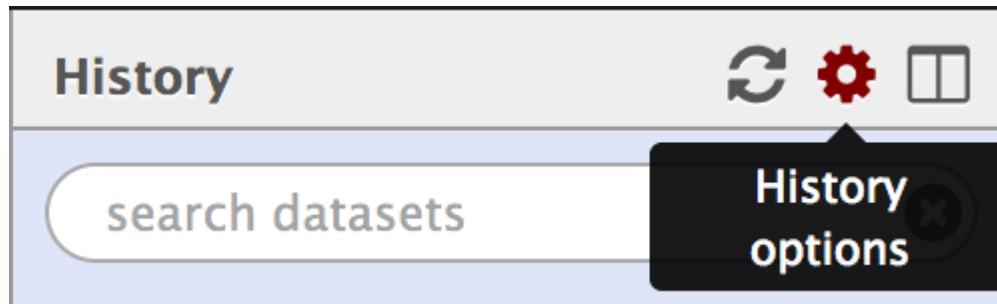
**Training: 16S rRNA sequencing with mothur:** This panel is titled 'Training: 16S rRNA sequencing with mothur' and contains a list of datasets. It shows 134 shown, 54 deleted, and 56 hidden datasets, with a total size of 1.05 GB. The datasets listed include '236: Krona pie chart on data', '235: HTML', '234: Taxonomy-to-Krona on collection 184: krona-formatted taxonomy file', '232: Make.biom on collection 189 and collection 184: biom files', '231: Newick Display on data', '218: Tree Graph', '217: Tree.shared on collection 199: tre', '214: Venn on collection 189: svg', '213: Venn on collection 189: sharedotus', and '206: Heatmap.sim on collection 199: heatmap.sim.svg'.

**Unnamed history:** This panel is titled 'Unnamed history' and contains a list of datasets. It shows 41 shown datasets, with a total size of 163.07 MB. The datasets listed include '41: samples', '40: https://zenodo.org/record/800651/files/Mock\_R2.fastq', '39: https://zenodo.org/record/800651/files/Mock\_R1.fastq', '38: https://zenodo.org/record/800651/files/F3D9\_R2.fastq', '37: https://zenodo.org/record/800651/files/F3D9\_R1.fastq', '36: https://zenodo.org/record/800651/files/F3D8\_R2.fastq', '35: https://zenodo.org/record/800651/files/F3D8\_R1.fastq', '34: https://zenodo.org/record/800651/files/F3D7\_R2.fastq', '33: https://zenodo.org/record/800651/files/F3D7\_R1.fastq', '32: https://zenodo.org/record/800651/files/F3D6\_R2.fastq', and '31: https://zenodo.org/record/800651/files/F3D6\_R1.fastq'.

# History options menu

History behavior is controlled by the *History options* (gear icon)

- *Create New* history will **not** make your current history disappear
- To see all of your histories, use the history switcher



- *Copy Datasets* from one history to another and save disk space for your quota

Loading data

## Importing data

- Copy/paste from a file
- Upload data from a local computer
- Upload data from internet using URL
- Upload data from online databases:  
UCSC, BioMart, ENCODE, modENCODE,  
Flymine etc.
- Import from Shared Data (libraries,  
histories, pages)
- Upload data from FTP

See [Getting data into Galaxy](#)

## Datatypes

- Tools only accept input datasets with the appropriate datatypes
- When uploading a dataset, its datatype can be either:
  - automatically detected
  - assigned by user
- Dataset produced by a tool: datatype assigned by the tool
- To change the datatype of a dataset:
  - galaxy-pencil *Edit Attributes and Datatype*
  - galaxy-pencil *Edit Attributes and Convert Formats*



# Reference datasets

## Example: reference Genome

- Genome build specifies which genome assembly a dataset is associated with
  - e.g. mm10, hg38...
- Can be automatically detected or assigned by user
  - Users can create custom genome builds
  - New builds can be added by the admin

### Database/Build

Mouse July 2007 (NCBI37/mm9) (mm9)

Burmese python Sep. 2013 (Python\_molurus\_bivittatus-5.0.2/pytBiv1) (pytBiv1)

Burton's mouthbreeder Oct 2011 (AstBur1.0/hapBur1) (hapBur1)

Bushbaby Mar. 2011 (Broad/otoGar3) (otoGar3)

Bushbaby Dec. 2006 (Broad/otoGar1) (otoGar1)

C. angaria Oct. 2010 (WS225/caeAng1) (caeAng1)

C. brenneri Nov. 2010 (C. brenneri 6.0.1b/caePb3) (caePb3)

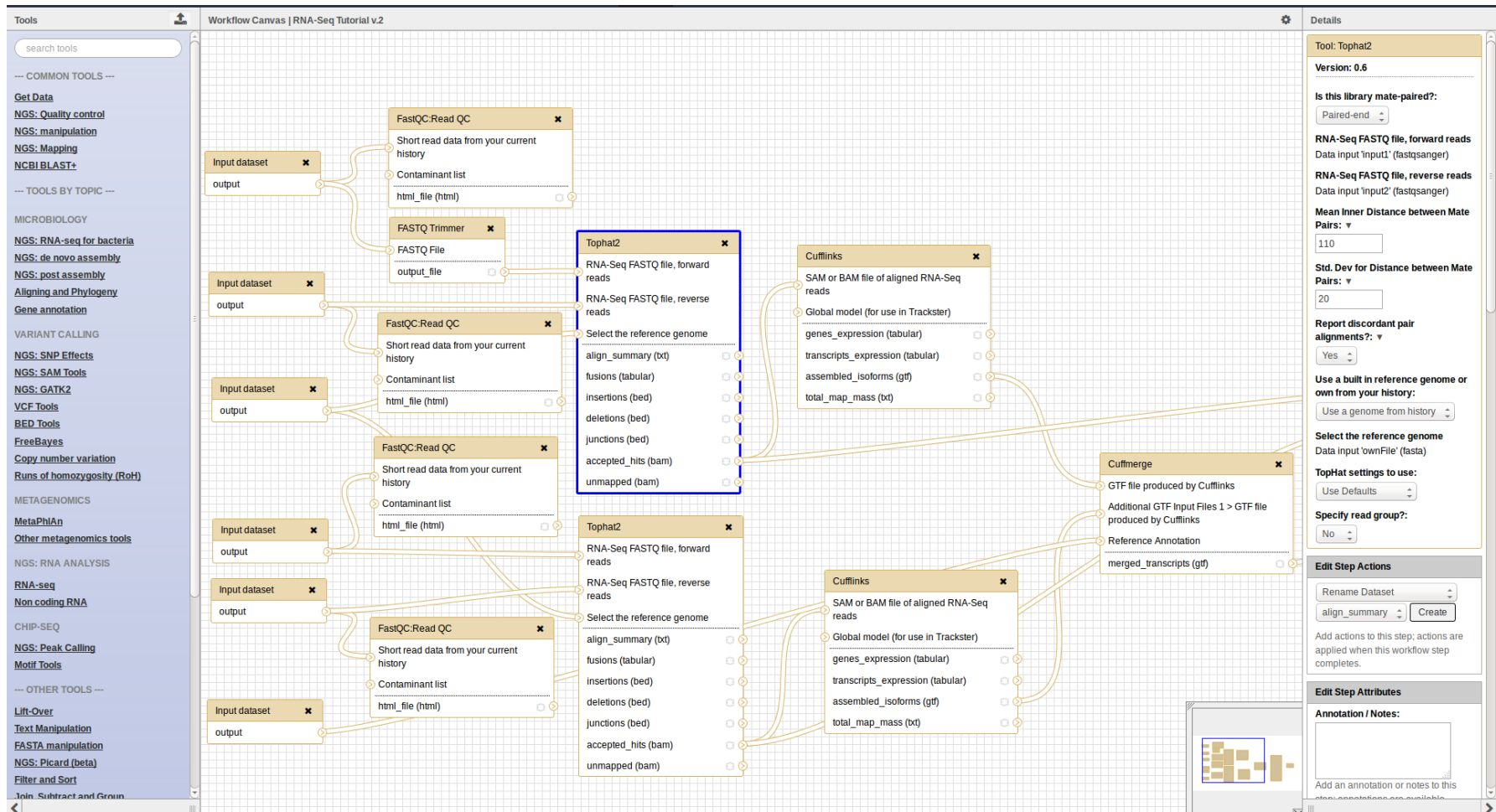
C. brenneri Feb. 2008 (WUGSC 6.0.1/caePb2) (caePb2)

C. brenneri Jan. 2007 (WUGSC 4.0/caePb1) (caePb1)

Workflows

# Workflow Editor

- **Extracted** from a history
- **Built manually** by adding and configuring tools using the canvas
- **Imported** using an existing shared workflow



Why would you want to create workflows?

- **Re-run** the same analysis on different input data sets
- **Change parameters** before re-running a similar analysis
- Make use of the workflow job **scheduling**
  - jobs are submitted as soon as their inputs are ready
- Create **sub**-workflows: a workflow inside another workflow
- **Share** workflows for publication and with the community

## Sharing data

- Share everything you do in Galaxy - histories, workflows, and visualizations
  - Directly using a Galaxy account's email addresses on the same instance
  - Using a web link, with anyone who knows the link
  - Using a web link and publishing it to make it accessible to everyone from the *Shared Data* menu

# Community

•Support forum: [Galaxy Help](#)

 **GalaxyHelp**

search topics, posts, users, or categories



Sign Up

Log In



all categories ▸

all tags ▸

Latest

Top

Categories

Topic



Category

Users

Replies

Views

Activity

  Troubleshooting resources for errors or unexpected results

Start by reviewing the troubleshooting FAQ. Common reasons and solutions for tool errors are explained. Most job errors can be resolved by correcting your input data's format/content. Others indicate a tool setting/param... [read more](#)



 usegalaxy.org support



1

85

7d

  Welcome to Galaxy Community Help

For assistance with a specific Galaxy server please post into appropriate category.



1

75

15d

- Community curated documentation: [Galaxy Community Hub](#)
- [Events](#) all around the world
- Galaxy Training for scientists, developers, admins, instructors: [Galaxy Training Community](#)
- Training questions? Chat with us on [Gitter](#)

# Transcriptomics

Training material for all kinds of transcriptomics analysis.

## Requirements

Before diving into this topic, we recommend you to have a look at:

- [Introduction to Galaxy Analyses](#)
- [Sequence analysis](#)
  - Quality Control: [slides](#) - [hands-on](#)
  - Mapping: [slides](#) - [hands-on](#)

## Material

Search

Lesson	Slides	Hands-on	Input dataset	Workflows	Galaxy tour	Galaxy instances
Introduction to Transcriptomics						
CLIP-Seq data analysis from pre-processing to motif detection						
De novo transcriptome reconstruction with RNA-Seq						
Differential abundance testing of small RNAs						
Downstream Single-cell RNA analysis with RaceID <a href="#">single-cell</a>						
GO Enrichment Analysis						
Network analysis with Heinz <a href="#">metatranscriptomics</a> <a href="#">network analysis</a>						
<a href="#">Plates, Batches, and Barcodes</a>						

-creating an account

<https://usegalaxy.eu/>