

Very short Galaxy Introduction

Sebastian Schmeier

s.schmeier@gmail.com

<http://sschmeier.github.io/bioinf-workshop/>

17.08.2015



Galaxy

- Accessible and transparent scientific workflow management software, which aims at facilitating reproducible research

Galaxy

localhost:8080/galaxy/ Google

Analyze Data Workflow Shared Data Visualization Help User

Using 0 bytes

History Unnamed history 0 bytes Your history is empty. Click 'Get Data' on the left pane to start

Tools search tools

Get Data 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 Graph/Display Data Regional Variation Multiple regression Multivariate Analysis Evolution Motif Tools Multiple Alignments FASTA manipulation NGS: QC and manipulation NGS: Mapping NGS: RNA Analysis NGS: SAM Tools NGS: GATK Tools (beta) NGS: Simulation Phenotype Association NGS: Picard

This is the Galaxy server, packaged for Bio-Linux.

To customize this page edit /etc/galaxy/static/welcome.html then sudo restart galaxy.

To find out more about Galaxy, [start here](#).

WWFSMD?
grow noodly appendages...

usegalaxy.org

This project is supported in part by [NSF](#), [NHGRI](#), and [the Huck Institutes of the Life Sciences](#).

The core Galaxy server and dependencies were packaged for Bio-Linux by the [NEBC Team](#) as part of a project under the [NERC Biomolecular Analysis Facility](#) in collaboration with the [Bergman Lab at the University of Manchester](#) and with [Simon Wagstaff at the Liverpool School of Tropical Medicine Bioinformatics Unit](#).

Galaxy

localhost:8080/galaxy/

Analyze Data Workflow Shared Data Visualization Help User

Using 0 bytes

Tools

search tools

Get Data

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

Graph/Display Data

Regional Variation

Multiple regression

Multivariate Analysis

Evolution

Motif Tools

Multiple Alignments

FASTA manipulation

NGS: QC and manipulation

NGS: Mapping

NGS: RNA Analysis

NGS: SAM Tools

NGS: GATK Tools (beta)

NGS: Simulation

Phenotype Association

NGS: Picard

This is the Galaxy server, packaged for Bio-Linux.

To customize this page edit /etc/galaxy/static/welcome.html then sudo restart galaxy.

To find out more about Galaxy, [start here](#).

WWFSMD?
grow noodly appendages...

usegalaxy.org

This project is supported in part by [NSF](#), [NHGRI](#), and the Huck Institutes of the Life Sciences.

The core Galaxy server and dependencies were packaged for Bio-Linux by the [NEBC Team](#) as part of a project under the [NERC Biomolecular Analysis Facility](#) in collaboration with the Bergman Lab at the [University of Manchester](#) and with [Simon Wagstaff](#) at the Liverpool School of Tropical Medicine.

History

Unnamed history
0 bytes

Q X

1

2

3

I. Tools panel

2. The working area

3. The history panel

1. THE LINKS TO THE TOOLS THAT THE GALAXY INSTALLATION CONTAINS.
2. THE WORKING AREA, WHERE WE CAN CHANGE PARAMETERS OF THE TOOLS.
3. THE HISTORY PANEL THAT CONTAINS ALL THE DATA AND STEPS WE PERFORMED.

The screenshot shows the Galaxy web interface with several panels:

- Left Panel (Tools):** A sidebar titled "Tools" containing a search bar and a list of tool categories. Red numbers "1" and "2" are overlaid on this panel.
- Search Bar:** A central search bar with the placeholder "search tools".
- Middle Panel (Tool Categories):** A list of tool categories:
 - Get Data
 - Lift-Over
 - Text Manipulation
 - Convert Formats
 - Filter and Sort
 - Join, Subtract and Group
 - Operate on Genomic Intervals
 - Statistics
 - Graph/Display Data
 - Regional Variation
 - Multiple regression
 - Multivariate Analysis
 - Evolution
 - Motif Tools
 - Multiple Alignments
 - FASTA manipulation
 - NGS: QC and manipulation
 - NGS: Mapping
 - NGS: RNA Analysis
 - NGS: SAM Tools
 - NGS: GATK Tools (beta)
 - NGS: Simulation
 - Phenotype Association
 - NGS: Picard
- Right Panel (History):** A panel titled "History" showing an "Unnamed history" section with 0 bytes. It includes a message: "Your history is empty. Click 'Get Data' on the left pane to start". Green number "3" is overlaid on this panel.
- Bottom Panel (Status):** A status bar at the bottom of the interface.

Red numbers "1" and "2" are overlaid on the left panel, and green number "3" is overlaid on the right panel.

3. THE HISTORY PANEL THAT CONTAINS ALL THE DATA AND STEPS WE PERFORMED.

The screenshot shows the Galaxy web interface at localhost:8080/galaxy/. The left sidebar, labeled '1', contains a list of tools under the 'Tools' heading, with 'Get Data' highlighted by a red box. The main content area, labeled '2', displays the 'Tools' page with various links like 'Get Data', 'Lift-Over', 'Text Manipulation', etc. A modal window titled 'Get Data' is open, listing external data sources. The right sidebar, labeled '3', shows the 'History' panel with an 'Unnamed history' entry. Red and green boxes highlight specific areas of interest: the 'Get Data' link in the sidebar, the 'Get Data' link in the modal, and the 'History' panel.

Galaxy

localhost:8080/galaxy/ Google

Analyze Data Workflow Shared Data Visualization Help User Using 0 bytes

Tools

search tools

[Get Data](#)

[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](#)

[Graph/Display Data](#)

[Regional Variation](#)

[Multiple regression](#)

[Multivariate Analysis](#)

[Evolution](#)

[Motif Tools](#)

[Multiple Alignments](#)

[FASTA manipulation](#)

[NGS: QC and manipulation](#)

[NGS: Mapping](#)

[NGS: RNA Analysis](#)

[NGS: SAM Tools](#)

[NGS: GATK Tools \(beta\)](#)

[NGS: Simulation](#)

[Phenotype Association](#)

[NGS: Picard](#)

1

Tools

search tools

Get Data

Lift-Over

Text Manipulation

Convert Formats

Filter and Sort

Join, Subtract and Group

NGS: QC and manipulation

NGS: Mapping

NGS: RNA-seq

NGS: SAMtools

NGS: BAM Tools

NGS: Picard

NGS: VCF Manipulation

Extract Features

2

Tools

search tools

Get Data

Upload File from your computer

[UCSC Main table browser](#)

[UCSC Archaea table browser](#)

[EBI SRA ENA SRA](#)

[BioMart Central server](#)

[GrameneMart Central server](#)

[Flymine server](#)

[modENCODE fly server](#)

[modENCODE modMine server](#)

[MouseMine server](#)

[Ratmine server](#)

[YeastMine server](#)

[modENCODE worm server](#)

[WormBase server](#)

3

History

Unnamed history 0 bytes

Your history is empty. Click 'Get Data' on the left pane to start

Galaxy

Analyze Data Workflow Shared Data Visualization Cloud Help User

Using 0% Google Analytics Twitter Button

Tools

search tools

Get Data 1

Upload File from your computer 2

[UCSC Main table browser](#)
[UCSC Archaea table browser](#)
[EBI SRA ENA SRA](#)
[BioMart Central server](#)
[GrameneMart Central server](#)
[Flymine server](#)
[modENCODE fly server](#)
[modENCODE modMine server](#)
[MouseMine server](#)
[Ratmine server](#)
[YeastMine server](#)
[modENCODE worm server](#)
[WormBase server](#)
[ZebrafishMine server](#)
[EuPathDB server](#)
[GenomeSpace import from file browser](#)

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](#).

Galaxy 101

Start small

The very first tutorial you need

Tweets by @galaxyproject

Galaxy

Analyze Data Workflow Shared Data Visualization Cloud Help User Using 0%

Tools

Get Data

Upload

UCSC

UCSC

EBI SRA

BioMart

Gramene

Flymine

modENe

modE

Mouse

Ratmine

Yeastmine

modE

Wormbase

Zebrafish

EuPathDB

Genome browser

Download data directly from web or upload files from your disk

You can Drag & Drop files into this box.

Type (set all): Auto-detect

2 Genome (set all): Additional Species ...

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

This screenshot shows the 'Get Data' interface in Galaxy. It includes a sidebar with various data sources like UCSC, EBI SRA, and BioMart. The main area has a large dashed box for dragging files. Below it are dropdowns for 'Type' and 'Genome'. The 'Choose local file' button is highlighted with a red box and labeled '1'. The 'Genome' dropdown is also highlighted with a red box and labeled '2'.

Galaxy

localhost:8080/galaxy/ Google

Analyze Data Workflow Shared Data Visualization Help User

Using 0 bytes

Tools

search tools

Get Data 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

Graph/Display Data Regional Variation

Multiple regression Multivariate Analysis

Evolution Motif Tools

Multiple Alignments FASTA manipulation

NGS: QC and manipulation NGS: Mapping

NGS: RNA Analysis NGS: SAM Tools

NGS: GATK Tools (beta) NGS: Simulation

Phenotype Association NGS: Picard

This is the Galaxy server, packaged for Bio-Linux.

To customize this page edit /etc/galaxy/static/welcome.html then sudo restart galaxy.

To find out more about Galaxy, [start here](#).

WWFSMD?
grow noodly appendages...

usegalaxy.org

1 19: mm9_chrX_SNP12
8_set.bed

2 19: mm9_chrX_SNP12
8_set.bed

3 19: mm9_chrX_SNP128_s
et.bed

This project is supported in part by NSF, NHGRI, and the Huck Institutes of the Life Sciences.

The core Galaxy server and dependencies were packaged for Bio-Linux by the NEBC Team as part of a project under the NE University of Manchester and with Simon Wagstaff at the Liverpool School of Tropical Medicine Bioinformatics Unit.

History

Unnamed history 0 bytes

1. Task is waiting for execution

2. Task is currently running

3. Task is finished

Galaxy

Analyze Data Workflow Shared Data Visualization Cloud Help User Using 0%

Tools

search tools

Get Data

Lift-Over

Text Manipulation

Convert Formats

Filter and Sort

Join, Subtract and Group

NGS: QC and manipulation

NGS: Mapping

NGS: BAM Tools

NGS: Picard

Extract Features

Fetch Sequences

Fetch Alignments

Get Genomic Scores

Operate on Genomic Intervals

Statistics

Graph/Display Data

Phenotype Association

snpEff

BEDTools

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](#).

Galaxy 101

Start small

The very first tutorial you need

History

search datasets

Unnamed history
1 shown, 18 deleted, 1 hidden

29.1 MB 1

19: mm9_chrX SNP128_s et.bed 2

20,000 regions
format: bed, database: mm9 2

uploaded bed file

display in IGB View 3

display at Ensembl Current

display at UCSC main

1.Chrom	2.Start	3.End	4.Name	5.
chrX	3242568	3242569	rs51257154	0
chrX	3242572	3242573	rs49693543	0
chrX	3242573	3242574	rs45795462	0
chrX	3749157	3749158	rs45795462	0
chrX	3749158	3749159	rs49693543	0
chrX	3749162	3749163	rs51257154	0

- I. Clicking the name gives more information
2. Here more general information is available of the dataset
3. Here you can download the dataset, get more infoarmtion, and run the task again.

The screenshot shows the Galaxy web interface. The top navigation bar includes 'Galaxy', 'Analyze Data', 'Workflow', 'Shared Data', 'Visualization', 'Cloud', 'Help', 'User', and status indicators ('Using 0%').

The left sidebar has a 'Tools' section with a search bar ('search tools') and a list of 'Get Data' sources: 'Upload File from your computer', 'UCSC Main table browser', 'UCSC Archaea table browser', 'EBI SRA ENA SRA', 'BioMart Central server', and 'GrameneMart Central server'. There is also a green upload icon.

The main area displays a table with columns labeled 1 through 7. The data consists of genomic coordinates and identifiers:

1	2	3	4	5	6	7
chrX	3242568	3242569	rs51257154	0	-	1
chrX	3242572	3242573	rs49693543	0	-	1
chrX	3242573	3242574	rs45795462	0	-	1
chrX	3749157	3749158	rs45795462	0	+	1
chrX	3749158	3749159	rs49693543	0	+	1
chrX	3749162	3749163	rs51257154	0	+	1
chrX	3907318	3907319	rs48647149	0	+	1
chrX	3907321	3907322	rs48584752	0	+	1
chrX	3907739	3907740	rs45858970	0	+	1
chrX	3907803	3907804	rs48529475	0	+	1
chrX	3907824	3907825	rs46088235	0	+	1

The right sidebar shows the 'History' panel with an 'Unnamed history' entry containing one dataset: '19: mm9_chrX_SNP12 8_set.bed'. It includes an eye icon, edit/pencil icon, and delete/x icon.

The eye button lets you visualise the data in the working area

Galaxy

Analyze Data Workflow Shared Data Visualization Cloud Help User Using 0%

Tools

search tools

Get Data

- Upload File from your computer
- UCSC Main table browser
- UCSC Archaea table browser
- EBI SRA ENA SRA
- BioMart Central server
- GrameneMart Central server
- Flymine server
- modENCODE fly server
- modENCODE modMine server
- MouseMine server
- Ratmine server
- YeastMine server
- modENCODE worm server
- WormBase server
- ZebrafishMine server
- EuPathDB server
- GenomeSpace import from file browser

Lift-Over

Text Manipulation

Convert Formats

Filter and Sort

Join, Subtract and Group

NGS: QC and manipulation

NGS: Mapping

NGS: BAM Tools

NGS: Picard

Attributes Convert Format Datatype Permissions

Edit Attributes

Name: mm9_chrX_SNP128_set.bed 2

Info: uploaded bed file

Database/Build: Mouse July 2007 (NCBI37/mm9) (mm9)

Number of comment lines:

Chrom column: 1 3

Start column: 2

End column: 3

Strand column (click box & select): 6

Name/Identifier column (click box & select): 4

Score column for visualization: 1
2
3
4

Save 4

Auto-detect

This will inspect the dataset and attempt to correct the above column values if they are not accurate.

History

search datasets

Unnamed history

1 shown, 18 deleted, 1 hidden

29.1 MB

19: mm9_chrX_SNP128_set.bed 1

- I. The pencil lets you edit the data
2. Filename
3. Specify which column contain which information
4. Save

The screenshot shows the Galaxy web interface. The top navigation bar includes 'Galaxy', 'Analyze Data', 'Workflow', 'Shared Data', 'Visualization', 'Cloud', 'Help', 'User', and resource icons. On the left, there's a 'Tools' section with a search bar and links to 'Get Data' (including 'Upload File', 'UCSC Main table browser', 'UCSC Archaea table browser', and 'EBI SRA ENA SRA'). The main content area is titled 'Edit Attributes' for a dataset named 'mm9_chrX_SNP128_set.bed'. The dataset is described as an 'uploaded bed file' and is 29.1 MB. In the 'History' panel, the dataset appears under 'Unnamed history' with the ID '19: mm9_chrX_SNP128_set.bed'. A red box highlights the delete icon (a red X) next to the dataset name.

Delete the data

Galaxy

Analyze Data Workflow Shared Data ▾ Visualization ▾ Cloud ▾ Help ▾ User ▾ Using 0%

Tools Genomes Genome Browser Tools Mirrors Downloads My Data Help

search tools

Get Data 1
[Upload File from your computer](#)
UCSC Main table browser 2
[UCSC Archaea table browser](#)
[EBI SRA ENA SRA](#)
[BioMart Central server](#)
[GrameneMart Central server](#)
[Flymine server](#)
[modENCODE fly server](#)
[modENCODE modMine server](#)
[MouseMine server](#)
[Ratmine server](#)
[YeastMine server](#)
[modENCODE worm server](#)
[WormBase server](#)
[ZebrafishMine server](#)
[EuPathDB server](#)
[GenomeSpace import from file browser](#)
[Send Data](#)
[Lift-Over](#)
[Text Manipulation](#)

Table Browser

Use this program to retrieve the data associated with a track in text format, to calculate intersections between tracks, and to retrieve DNA sequence covered by a track. For help in using this application see [Using the Table Browser](#) for a description of the controls in this form, the [User's Guide](#) for general information and sample queries, and the OpenHelix Table Browser [tutorial](#) for a narrated presentation of the software features and usage. For more complex queries, you may want to use [Galaxy](#) or our [public MySQL server](#). To examine the biological function of your set through annotation enrichments, send the data to [GREAT](#). Send data to [GenomeSpace](#) for use with diverse computational tools. Refer to the [Credits](#) page for the list of contributors and usage restrictions associated with these data. All tables can be downloaded in their entirety from the [Sequence and Annotation Downloads](#) page.

3 **clade:** Mammal **genome:** Mouse **assembly:** July 2007 (NCBI37/mm9)
group: Genes and Gene Predictions **track:** UCSC Genes [manage custom tracks](#) [track hubs](#)

4 **table:** knownGene [describe table schema](#)

5 **region:** genome position chrX:1-166650296 [lookup](#) [define regions](#)
identifiers (names/accessions): [paste list](#) [upload list](#)

filter: [create](#)

intersection: [create](#)

correlation: [create](#)

6 **output format:** BED – browser extensible data **Send output to** [Galaxy](#) [GREAT](#)
[GenomeSpace](#)
output file: * (leave blank to keep output in browser)
file type returned: plain text gzip compressed

7 **get output** [summary/statistics](#)

To reset **all** user cart settings (including custom tracks), [click here](#).

We can also load data from the web directly into Galaxy

Galaxy strengths

- **Accessibility:** You can access Galaxy in many different ways
- Once your data is in Galaxy you can make the use of the full collection of available programs to analyse your data
- Galaxy facilitates the use of command-line tools in a web interface
- The history lets you keep track of your analyses steps
- You can save the history -> **reproducible research**
- You can create workflows from your history for re-use of common tasks
- **Transparency:** You can share any of your data and workflows
- **Extensible:** You can add your own tools to Galaxy

Galaxy weaknesses

- Not trivial to setup
- Not as flexible as the command-line, as you are restricted to the tools and options that the authors initially intended
- However you can try to mitigate this by installing your own tool or version of the tool => Also not trivial
- Not as fast as a purely command-line tool

References

Blankenberg D, Von Kuster G, Coraor N, Ananda G, Lazarus R, Mangan M, Nekrutenko A, Taylor J. Galaxy: a web-based genome analysis tool for experimentalists. Current Protocols in Molecular Biology. 2010 Jan; Chapter 19:Unit 19.10.1-21.

Giardine B, Riemer C, Hardison RC, Burhans R, Elnitski L, Shah P, Zhang Y, Blankenberg D, Albert I, Taylor J, Miller W, Kent WJ, Nekrutenko A. Galaxy: a platform for interactive large-scale genome analysis. Genome Research. 2005 Oct; 15(10):1451-5.

Goecks J, Nekrutenko A, Taylor J and The Galaxy Team. Galaxy: a comprehensive approach for supporting accessible, reproducible, and transparent computational research in the life sciences. Genome Biol. 2010 Aug 25; 11(8):R86.

Sebastian Schmeier
s.schmeier@gmail.com

THE ENGINE
OF THE NEW
NEW ZEALAND
