

# Introduction to Galaxy

Stéphanie Le Gras  
([slegras@igbmc.fr](mailto:slegras@igbmc.fr))

# Guidelines

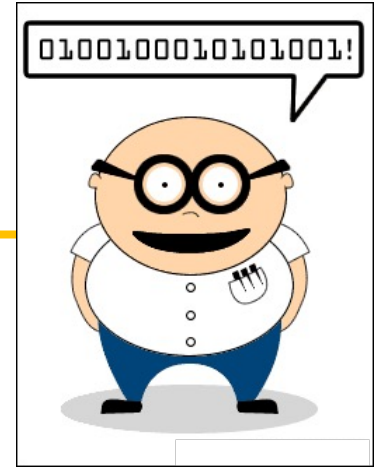
- Analyzing biological data with informatics tools
- Presentation of the Galaxy project
- Description of the main features of the Galaxy platform

# Analyzing biological data with informatics tools

# Bioinformatics analyses

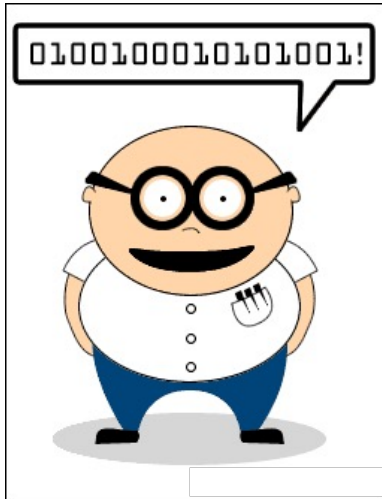


Informatics data



nature

# Bioinformatics analyses



## Scripts, softwares

```
#!/usr/bin/perl

use strict;
use warnings;
use Getopt::Long;

## Date : 22 fev 2011
## Author : Stephanie Le Gras

## Objectives :

my $num_arg = scalar @ARGV;
my $programe = "ExtractID.pl";
my $input;
my $out;
my $id;

my $result = GetOptions(
    "id=s" => \$id,
    "out=s" => \$out,
    "input=s" => \$input,
);

my $usage = <<END;

Usage: $programe --id=FILENAME --out=FILENAME --input=FILENAME

END

die $usage unless ($result);
my @files = @ARGV;
die "Enter at least two files\n$usage" if ( $num_arg < 2 );
die $usage if ( $num_arg == 0 );

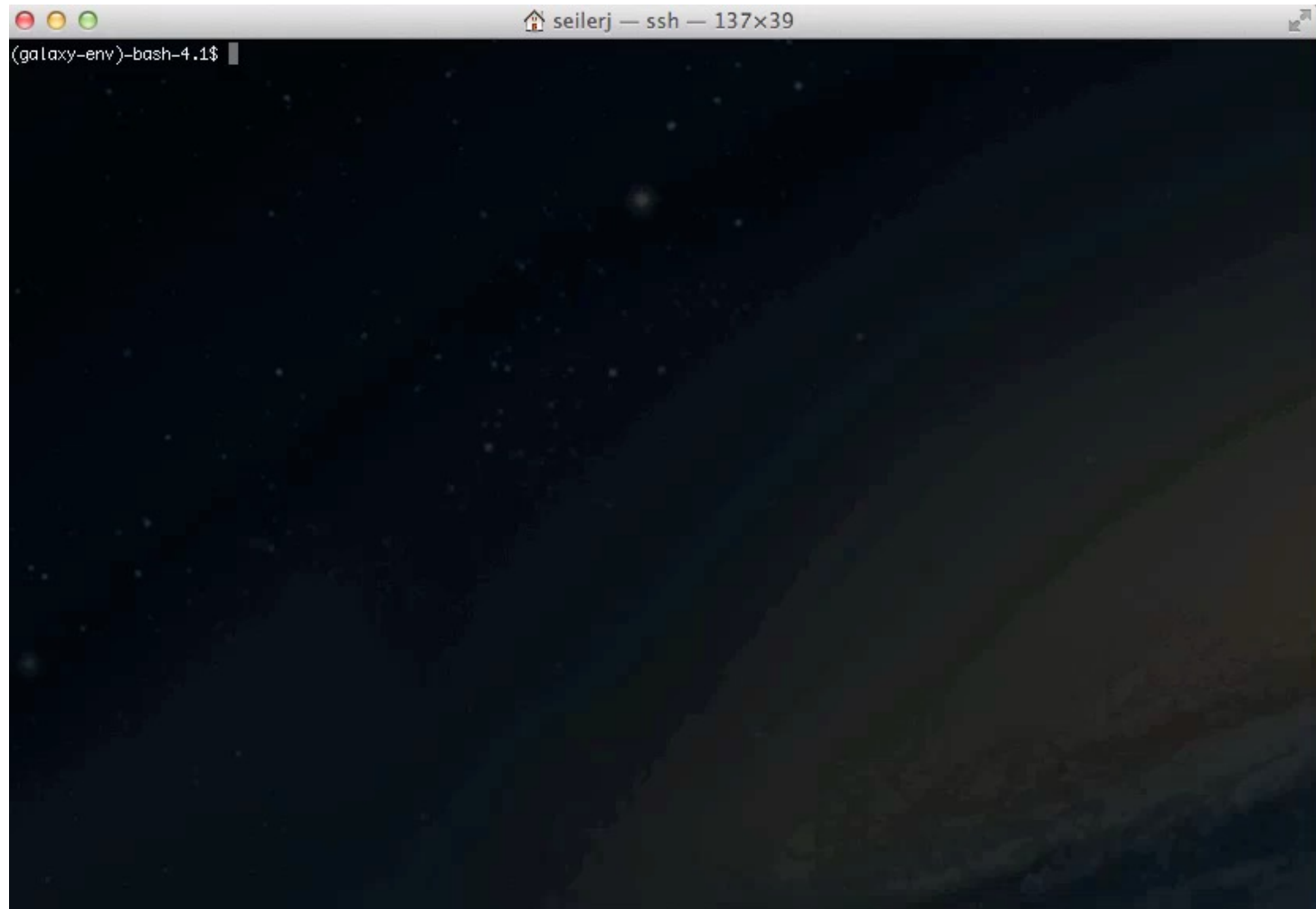
my %ids;
$out = ( defined $out ) ? $out : "results.txt";

## first, every lines of each files are put in the hash table ids. Variant ids are used as keys of the
## hash table and it contains a table.
```

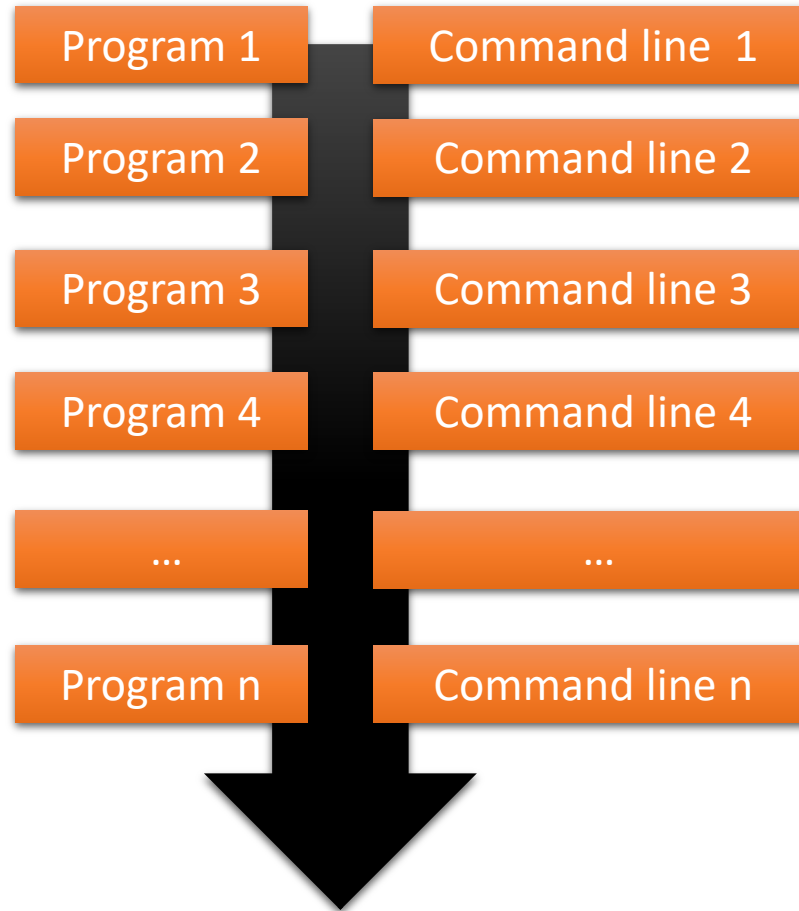
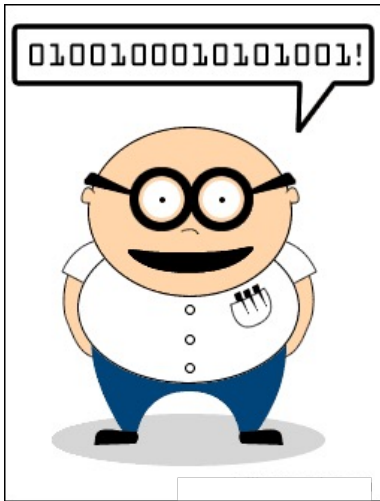
## Command line

```
macs14 -t treatment.sort.bed -c control.bed -f BED -g mm --name=name1 --llocal=50000
--slocal=5000 > macs1.nohup 2>&1 &
```

# Bioinformatics analyses



# Bioinformatics analyses



**PIPELINE /  
WORKFLOW**

Galaxy ?







# Galaxy

PROJECT

# Galaxy project

# What is Galaxy ?

Galaxy is a **computing platform** that enables people to **run complex bioinformatics tools** on a **compute cluster** through a **simple web interface**.



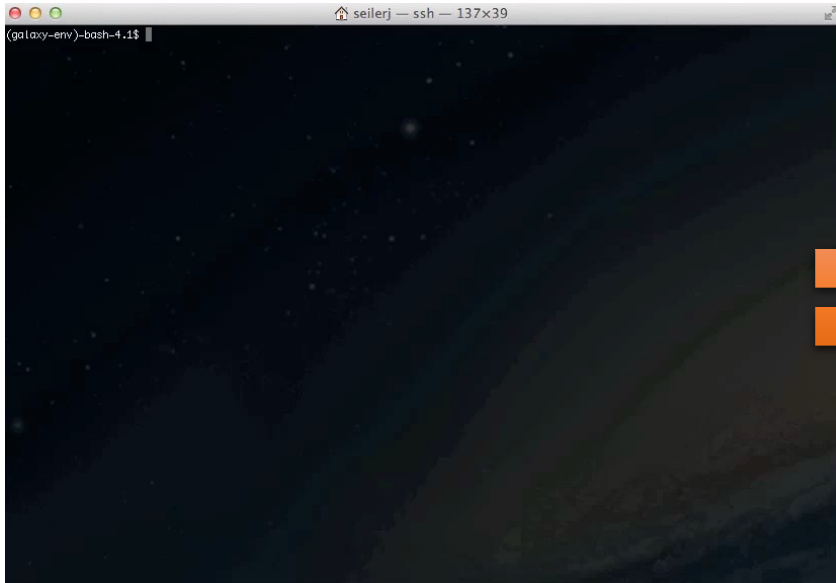
EMORY  
UNIVERSITY



# This is Galaxy



# Running analyses with tools



# Galaxy philosophy

- **Perform, and share** complete analysis
- **No programming skills** required
- **Open source** and **free** solution
- **Very large** and **active** community
- **Reproducibility/Usability/Transparency**

# How to use Galaxy

# Use Galaxy

- All analyses are run on remote computing infrastructures
- No need to have a Supercomputer to use Galaxy
- Web browser





# Use Galaxy

	UseGalaxy Servers	Public Servers	TlaaS	Academic Clouds	Commercial Clouds	Containers	VMs	Local
Free to use	Yes	Yes	Yes	Yes <sup>1</sup>	No	Yes	Yes	Yes
Uses your local compute infrastructure	No	No	No	No	No	Yes <sup>2</sup>	Yes <sup>2</sup>	Yes
Datasets (including intermediate) total > 250GB	No	? <sup>5</sup>	Yes	Yes	Yes	Yes <sup>3</sup>	Yes <sup>3</sup>	Yes
Computational requirements are similarly large	No	? <sup>5</sup>	Yes	Yes	Yes	Yes <sup>3</sup>	Yes <sup>3</sup>	Yes
Share Galaxy objects outside your organization	Yes	Yes	Yes	Yes	Yes	Yes <sup>4</sup>	Yes <sup>4</sup>	Yes <sup>5</sup>
Install custom tools and reference genomes	No	No	No	Yes <sup>5</sup>	Yes	Yes	Yes	Yes
Have absolute data security requirements	No	No	No	? <sup>5</sup>	? <sup>5</sup>	? <sup>5</sup>	? <sup>5</sup>	Yes

\* TlaaS: Training Infrastructure as a Service

- <https://galaxyproject.org/use/>
  - 171 resources for using Galaxy (Last Update on: 2021, June 7<sup>th</sup>)

# Galaxy public servers

Your research institute



Connect to Galaxy web site through  
a web browser  
(<https://usegalaxy.org/>)



Download and  
upload of the data

Remote instance of Galaxy



Run analyses

# Galaxy public servers

Your research institute



World wide



Download and upload of the data



Remote instance of Galaxy



Uploading (takes a lot of time)  
Shared computing power  
(longer analysis)

# Galaxy local server

- Run a local production Galaxy because you want to
  - install and use tools unavailable on public Galaxies
  - use sensitive data (e.g. clinical)
  - process large datasets that are too big for public Galaxies
  - Develop Galaxy tools
  - Develop Galaxy itself



# Description of the main features of Galaxy

# Galaxy web interface

Top menu

The screenshot displays the Galaxy web interface with several key components highlighted by orange callouts:

- Top menu:** Located at the top right, it includes navigation options like "Analyze Data", "Workflow", "Visualize", "Shared Data", "Help", and "Login or Register", along with a "Using 0%" indicator.
- Tool panel:** Located on the left side, it features a search bar and a list of tool categories such as "Get Data", "Collection Operations", "GENERAL TEXT TOOLS", "GENOMIC FILE MANIPULATION", and "COMMON GENOMICS TOOLS".
- Data display and tools dialog window:** The central area shows a text-based introduction to Galaxy, a "Galaxy Help" banner with the text "Got Questions? Get Answers." and the URL "help.galaxyproject.org", and a "Tweets" section featuring a tweet from the Galaxy Project.
- History panel:** Located on the right side, it includes a search bar for datasets and a message stating "This history is empty. You can load your own data or get data from an external source".

# Top menu

The image shows a screenshot of the Galaxy web interface with several orange callout boxes and arrows pointing to specific menu items. The top navigation bar includes 'Analyze Data', 'Workflow', 'Visualize', 'Shared Data', 'Help', 'Login or Register', and a grid icon. The left sidebar contains a search bar and tool categories like 'Get Data', 'Collection Operations', 'GENERAL TEXT TOOLS', 'GENOMIC FILE MANIPULATION', and 'FASTA/FASTQ'. The main content area features a 'Try Galaxy on the Cloud' banner and a 'Tweets by @galaxyproject' section. The right sidebar shows 'Histor' and 'Unnan' sections.

Run workflows

Get Help

Run analyses

Access public data

Log in/out, manage your account

Galaxy is an open source, web-based platform for data intensive research. If you can install you can install you can install you can install thousands of

Try Galaxy on the Cloud

Now you can have a personal Galaxy within the infinite Universe

Tweets by @galaxyproject

# Exercise 1 : Log in

- Go to <http://use.galaxeast.fr>
- Log in with your training account (look to the post-it).



# Tool Panel / Run analyses

The screenshot displays the Galaxy web interface. On the left is the **Tools** panel, which is highlighted with an orange border and an orange arrow pointing to it with the text "Tool panel". The panel lists various tool categories such as "Get Data", "Collection Operations", "GENERAL TEXT TOOLS", "GENOMIC FILE MANIPULATION", and "COMMON GENOMICS TOOLS". The main content area features a header with navigation links like "Analyze Data", "Workflow", "Visualize", "Shared Data", "Help", "Login or Register", and a "Using 0%" indicator. Below the header is a text block about Galaxy being an open-source platform for data-intensive biomedical research. A central "Galaxy Help" banner with the text "Got Questions? Get Answers." and the URL "help.galaxyproject.org" is displayed. Below the banner is a "Tweets" section by @galaxyproject, featuring a tweet from Galaxy Project about group tags for complex experimental designs. On the right is the "History" panel, which shows "Unnamed history (empty)" and a message: "This history is empty. You can load your own data or get data from an external source".

# Tool Panel / Run analyses

The diagram illustrates the workflow of selecting a tool from a panel. On the left, a 'Tools' panel is shown with a search bar and a list of categories. An arrow labeled 'Upload data' points to an upload icon in the top right of the panel. An arrow labeled 'Search a tool' points to the search bar. An arrow labeled 'Tool category' points to the 'Text Manipulation' category. A large orange arrow points from the 'Text Manipulation' category to a detailed view of the 'Text Manipulation' tool on the right. This detailed view shows the tool's name, description, and various options.

Upload data

Search a tool

Tool category

Tools

search tools

Get Data

Collection Operations

GENERAL TEXT TOOLS

Text Manipulation

Filter and Sort

Join, Subtract and Group

Datamash

GENOMIC FILE MANIPULATION

FASTA/FASTQ

FASTQ Quality Control

SAM/BAM

BED

VCF/BCF

Nanopore

Tools

search tools

Get Data

Collection Operations

GENERAL TEXT TOOLS

Text Manipulation

Compute an expression on every row

annotateMyIDs annotate a generic set of identifiers

UniProt ID mapping and retrieval

Add input name as column to an existing tabular file

Replace column by values which are defined in a convert file

Replace parts of text

Tool

# Tools dialog window

**Galaxy** Analyze Data Workflow Visualize Shared Data Help Login or Register Using 0%

**Tools** search tools

**Get Data**

**Collection Operations**

**GENERAL TEXT TOOLS**

**Text Manipulation**

[Compute](#) an expression on every row

[annotateMyIDs](#) annotate a generic set of identifiers

[UniProt](#) ID mapping and retrieval

[Add input name as column](#) to an existing tabular file

[Replace column](#) by values which are defined in a convert file

[Replace](#) parts of text

[Text transformation](#) with sed

[Unfold](#) columns from a table

[Unique lines](#) assuming sorted input file

[Replace Text](#) in entire line

[Replace Text](#) in a specific column

[Multi-Join](#) (combine multiple files)

[Select last](#) lines from a dataset (tail)

[Cut](#) columns from a table (cut)

[Create text file with recurring lines](#)

**Compute an expression on every row (Galaxy Version 1.2.0)** Versions Options

**Add expression**

c3-c2

**as a new column to**

No tabular dataset available. Dataset missing? See TIP below

**Round result?**

NO

**Skip a header line**

no

# characters are already considered as comments and kept

Execute

**TIP:** If you have a dataset that is not tabular, you can use the [Convert](#) tool to convert it to a tabular file.

**What it does**

This tool computes an expression on every row of a dataset. The expression is evaluated on each row of the input dataset and the result is stored in a new column. The expression is evaluated on each row of the input dataset and the result is stored in a new column. The expression is evaluated on each row of the input dataset and the result is stored in a new column.

- Column name (field).
- c3-c2

**Example**

If this is your input dataset:

```
chr1 151000000 151000000 151000000 151000000
```

**History** search datasets

**Unnamed history** (empty)

This history is empty. You can load your own data or get data from an external source

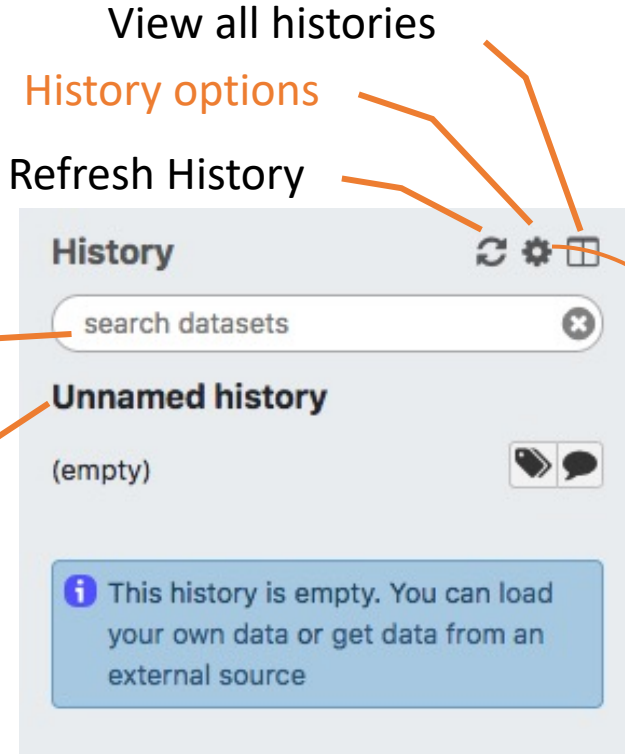
**Data display and tools dialog window**

- Set parameters
- Run tools
- Get help on tools
- Display content of dataset

# History

The screenshot displays the Galaxy web interface. At the top, the navigation bar includes the Galaxy logo, 'Analyze Data', 'Workflow', 'Visualize', 'Shared Data', 'Help', and 'Login or Register'. The left sidebar contains a 'Tools' section with a search bar and a list of tools under 'Text Manipulation', including 'Compute an expression on every row'. The main content area shows the configuration for the 'Compute' tool, with the expression 'c3-c2' entered. Below the configuration is an 'Execute' button and a tip: 'TIP: If your data is not TAB delimited, use Text Manipulation->Convert'. The right sidebar features a 'History' panel with a search bar and a message: 'This history is empty. You can load your own data or get data from an external source'. An orange callout box on the right side of the image contains the text: 'History panel Keep track of each job run'.

# History



Search datasets

History name

View all histories

History options

Refresh History

## HISTORY LISTS

Saved Histories

Histories Shared with Me

## CURRENT HISTORY

Create New

Copy History

Share or Publish

Show Structure

Extract Workflow

Delete

Delete Permanently

Make Data Private

## DATASET ACTIONS

Copy Datasets

Dataset Security

Resume Paused Jobs

Collapse Expanded Datasets

Unhide Hidden Datasets

Delete Hidden Datasets

Purge Deleted Datasets

## DOWNLOADS

Export Tool Citations

Export History to File

## OTHER ACTIONS

Import from File

# View all histories

The screenshot displays the Galaxy web interface with a dark header bar. The header includes the Galaxy logo, navigation menus for 'Analyze Data', 'Workflow', 'Visualize', 'Shared Data', 'Help', and 'User', and a 'Using 0%' indicator. Below the header, there are search bars for 'search histories' and 'search all datasets', along with a 'Create new' button. The main content area is divided into several vertical panels, each representing a 'Current History'. The first panel on the left is titled 'Unnamed history' and is empty, with a message 'This history is empty'. The other panels also show 'Unnamed history' and contain lists of datasets with their names, sizes, and actions (view, edit, delete). The datasets listed include: '6: L1spa\_ORF1.1.fastq' (127.21 KB), '5: TALs.fasta', '4: Count on data 3', '3: Galaxy14-(Intersect on data 13 and data 1).bed', '2: shuffleseq on data 1', '1: lkpeaks\_notRBPJ +-150\_ran\_dom80\_adjacentSeq.fasta', '5: Correspondance\_JASPAR\_C\_ORE.txt', '4: fimo.txt', and '2: Brn2\_Day2\_rtta\_rep2.sort.bed'. The interface is clean and organized, with a focus on data management and history tracking.

## Exercise 2 : History

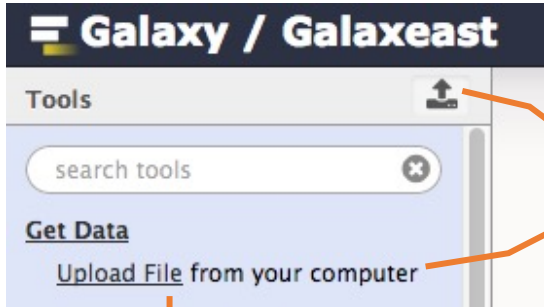
- Create a new history
- Change the name of the history to “RNA-seq data analysis”

# Import data into Galaxy

- Your own data (from your computer)
- Shared data
- Data from external sources



# Import your own data to Galaxy






Display the drag and drop utility used to upload local files








## Download from web or upload from disk

Regular Composite

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

Name	Size	Type	Genome	Settings	Status
 Mouse_ChIP-Seq_example_Control_Data_chr19_mm9.fastq	84.1 MB	Auto-det...	unspecified (?)		 0%

Type (set all): Auto-detect Genome (set all): unspecified (?)

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

Name of the dataset

Size of the dataset

File format

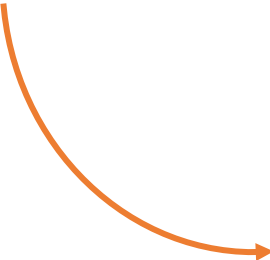
Genome

# Import shared data (data libraries)

Shared Data ▾ Visualizations (Top menu)

- Data Libraries
- Histories
- Workflows
- Visualizations
- Pages

Data Libraries



Galaxy / Galaxeast Analyze Data Workflow Shared Data ▾ Visualization ▾ Help ▾ User ▾ Using 0%

DATA LIBRARIES < 0 1 2 > showing 20 of 23 libraries Search Help

name	description	synopsis
<a href="#">Chip seq test dataset (chr10)</a>	MITF test dataset for chipSeq - chr10	MITF test dataset for chipSeq - chr10
<a href="#">Chip seq test dataset (chr2)</a>	Chip seq test dataset (chr2)	Chip seq test dataset (chr2)
<a href="#">Chromosome length</a>	*.len files for several assemblies	
<a href="#">Data demo AVIESAN</a>	Data demo AVIESAN	Data demo AVIESAN
<a href="#">Data Rustenholz</a>	Vitis vinifera data	Vitis vinifera data
<a href="#">Data_megadebug_jung</a>	Data_megadebug_jung	Data_megadebug_jung
<a href="#">Data_megadebug_velt</a>		
<a href="#">EBA2016</a>		
<a href="#">Fimo database (updated Mar. 16 2015)</a>		
<a href="#">Fimo databases</a>	Downloaded from MEME website	updated Jan 31 2013
<a href="#">Galaxy training - Chip seq test datasets</a>	Galaxy training - Chip seq test datasets	Galaxy training - Chip seq test datasets
<a href="#">Galaxy training - RNA seq test datasets</a>	Galaxy training - RNA seq test datasets	Galaxy training - RNA seq test datasets
<a href="#">Genome fasta files</a>	Genome fasta files (GATK - IGBMC)	Fasta files used by the IGBMC microarray and sequencing platform (to be used with GATK)
<a href="#">GTF</a>	Annotation files in GTF format	Annotation files in GTF format
<a href="#">Introduction 2 Galaxy (datasets)</a>	Introduction 2 Galaxy (datasets)	
<a href="#">Jaspar motifs ID &lt;-&gt; name</a>	table of correspondences (for FIMO results...)	
<a href="#">MITF test dataset for chipSeq</a>	MITF test dataset for chipSeq	MITF (bam)
<a href="#">MITF test dataset for RNAseq</a>	MITF test dataset for RNAseq	
<a href="#">NGS course (Sep)</a>	NGS course (Sep)	NGS course (Sep)

# Import shared data (data libraries)

2. Import selected dataset to history

Galaxy / Galaxeast Analyze Data Workflow Shared Data Visualization Help User Using 0%

DATA LIBRARIES << 0 1 2 >> showing 6 of 6 items  include deleted to History Download Delete Details Help

Libraries / Chromosome length

<input type="checkbox"/> name <sup>↓</sup>	description	data type	size	time updated (UTC)	
<input type="checkbox"/> ..					
<input type="checkbox"/> <a href="#">ce10.len</a>		len	98 bytes	2015-01-08 01:25	
<input type="checkbox"/> <a href="#">dm3.len</a>		len	227 bytes	2015-01-08 01:25	
<input type="checkbox"/> <a href="#">hg19.len</a>		len	376 bytes	2015-01-08 01:25	
<input type="checkbox"/> <a href="#">mm10.len</a>		len	1.4 KB	2015-01-08 01:25	
<input type="checkbox"/> <a href="#">mm9.len</a>		len	330 bytes	2015-01-08 01:25	
<input type="checkbox"/> <a href="#">tair10.len</a>		len	75 bytes	2015-01-08 01:25	

<< 0 1 2 >> showing 6 of 6 items

1. Select dataset

# Import public data

Tools

search tools

**Get Data**

- Upload File from your computer
- UCSC Main table browser
- UCSC Archaea table browser
- EBI SRA ENA SRA
- BioMart Ensembl 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

Browse and import external data from public databases

Genomes Genome Browser Tools Mirrors Downloads My Data Help About Us

### 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.

clade:  genome:  assembly:

group:  track:

table:

region:  genome  position

identifiers (names/accessions):

filter:

intersection:

correlation:

output format:  Send output to  Galaxy  GREAT  GenomeSpace

output file:  (leave blank to keep output in browser)

file type returned:  plain text  gzip compressed

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

### Using the Table Browser

This section provides brief line-by-line descriptions of the Table Browser controls. For more information on using this program, see the [Table Browser User's Guide](#).

- **clade:** Specifies which clade the organism is in.
- **genome:** Specifies which organism data to use.

# Exercise 3 : Import data to Galaxy

- 1. Import to Galaxy the file siLuc3\_S12040.fastq from the “Shared data > Data Libraries > NGS data analysis training > RNAseq > rawdata” to the history called “RNA-seq data analysis”
- 2. You should be in the history “RNA-seq data analysis” (Switch to it if needed)
- 3. Import to Galaxy the file sample.bed.gz located in the directory galaxy.
  - Type : bed
  - Genome : Mouse July 2007 (NCBI37/mm9) (mm9)

# Datasets/Jobs in the History

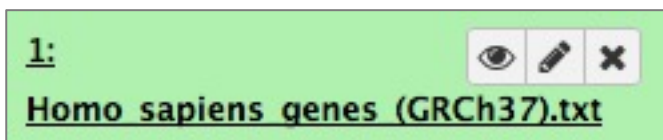
Grey: the job is waiting to run



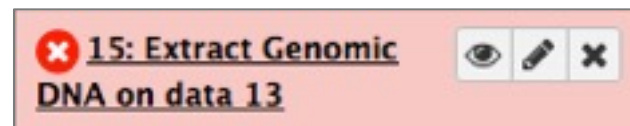
Yellow: the job is running



Green: the job is successfully done



Red: the job encountered a problem



# Datasets/Jobs in the History

Number of lines  
in the file or size  
of the file

Format

Genome

The screenshot shows a web interface for a dataset named '2: sample.bed'. At the top right, there are icons for view, edit, and close. Below the title, it displays '32,561 regions' and 'format: bed, database: mm9'. A text box contains 'uploaded bed file'. Below this are icons for save, info, chart, and help, along with share and comment icons. A list of links for viewing the data in various tools (IGB View, Ensembl, IGV, UCSC) is provided. At the bottom, a table shows the first five lines of the BED file.

1.Chrom	2.Start	3.End	4.Name
chr1	193580486	193580686	chr1-19345
chr1	64972363	64972563	chr1-64860
chr1	134238383	134238583	chr1-13416
chr1	51991430	51991630	chr1-51879
chr1	53880739	53880939	chr1-53768

If the dataset is a text  
file, the first lines of  
the file are displayed

# Datasets/Jobs in the History

View dataset (if possible) in the middle panel of Galaxy

Download dataset

**2: sample.bed** [eye icon] [pencil icon] [x icon]

32,561 regions  
format: **bed**, database: **mm9**

uploaded bed file

[download icon] [info icon] [chart icon] [help icon] [share icon] [comment icon]

display in IGB [View](#)  
display at Ensembl [Current](#)  
display with IGV [local](#) [Mouse](#) [mm9](#)  
display at UCSC [main](#) [test](#)

1. Chrom	2. Start	3. End	4. Name
chr1	193580486	193580686	chr1-19345
chr1	64972363	64972563	chr1-64860
chr1	134238383	134238583	chr1-13416
chr1	51991430	51991630	chr1-51879
chr1	53880739	53880939	chr1-53768

Delete dataset

Edit attributes of the dataset (change name, format, genome, permission)



# Size of histories and quota

The screenshot displays a software interface with a dark header bar at the top right showing a green progress indicator and the text "Using 20%". Below this is a "History" panel with a search bar labeled "search datasets" and a refresh icon. The panel lists "RNA-seq data analysis" with "2 shown" items. The first item is "7.23 GB" with a checkmark, a folder icon, and a speech bubble icon. The second item is "2: sample.bed" with an eye icon, a pencil icon, and a close icon. The third item is "1: siLuc3 S12040.fastq" with an eye icon, a pencil icon, and a close icon. An orange arrow points from the text "Size of history" to the "7.23 GB" entry. Another orange arrow points from the text "Quota" to the "Using 20%" indicator.

Using 20%

History

search datasets

RNA-seq data analysis  
2 shown

7.23 GB


2: sample.bed

1: siLuc3 S12040.fastq

Size of history

Quota

# Exercise 4 : remove dataset

- 1. Remove the dataset sample.bed from your history by clicking on the button 
- 2.
  - A. Click on “deleted” in the top of the history panel (below the history name). Remove definitely the file from the disk by clicking on ” Permanently remove it from disk”.
  - B. Click on “hide deleted”

# Exercise 5 : Running a tool

- 1. Use the tool “FastQC Read Quality reports” to compute quality analysis on the dataset “siLuc3\_S12040.fastq”
  - Use default parameters.

# Workflows

Galaxy

Analyze Data Workflow Shared Data Visualization Help User Using 0%

Tools

- NGS: SAMtools
- NGS: BamTools
- NGS: Picard
- NGS: VCF Manipulation
- NGS: Peak Calling
- NGS: Variant Analysis
- NGS: RNA Structure
- NGS: Du Novo
- NGS: Gemini
- Operate on Genomic Intervals
- Statistics
- Graph/Display Data
- CloudMap
- Phenotype Association
- BEDTools
- Genome Diversity
- EMBOSS
- Regional Variation
- FASTA manipulation
- Multiple Alignments
- Metagenomic Analysis
- Multiple regression
- Multivariate Analysis
- Motif Tools
- STR-FM: Microsatellite Analysis
- NCBI SRA Tools
- DEPRECATED
- NGS: GATK Tools (beta)
- Workflows
  - All workflows

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](#).

080+

Public Galaxy Servers  
and still counting

Tweets by @galaxyproject

Galaxy Project @galaxyproject  
Did we mention: Galaxy Admin Training early registration ENDS IN 12 HOURS. [bit.ly/gat2016](http://bit.ly/gat2016)

History

search datasets

Unnamed history

0 b

This history is empty. You can [load your own data](#) or [get data from an external source](#)

Run workflows

Create, run, edit (...) workflows

# Workflows

Create new workflow

## Your workflows

Create new workflow Upload or import workflow

Name	# of Steps
imported: CloudMap Variant Discovery Mapping (and Variant Calling) workflow_2-7-2014	33
imported: CloudMap Variant Discovery Mapping (and Variant Calling) workflow_2-7-2014	33
imported: metagenomic analysis	16
imported: imported: MACS (mm8)	10
imported: metagenomic analysis	16
imported: CHIP-Seq analysis on BAM files	11
imported: Sort BAM for Peak Calling MACS tool	5
Bergen workshop	0

## Workflows shared with you by others

No workflows have been shared with you.

## Other options

Configure your workflow menu

Edit, run, share (...) workflows

# Workflows

Save, run (...) workflows

The screenshot displays a workflow editor interface with the following components:

- Tools Panel (Left):** A search bar and a list of tool categories including Get Data, Send Data, Text Manipulation, Filter and Sort, Join, Subtract and Group, Convert Formats, Fetch Sequences, Operate on Genomic Intervals, Graph/Display Data, Motif Tools, Multiple Alignments, FASTA manipulation, NGS: QC and manipulation, NGS: Mapping, NGS: RNA Analysis, NGS: SAM Tools, NGS: GATK Tools (beta), NGS: Peak Calling, NGS: Miscellaneous, NGS: ChIP Analysis, NGS: BED Tools, NGS: Picard, NGS: Variant Annotation, NGS: S-mart, NGS: DeepTools, and Database query.
- Workflow Canvas (Center):** A grid-based workspace containing several tool nodes:
  - Input dataset:** Provides an 'output' port.
  - FastQC: Read QC:** Takes 'Short read data from your current history' and a 'Contaminant list' (html\_file) as input.
  - Convert from BAM to BED:** Takes 'output (sam)' as input and produces 'output (bed)'. It is highlighted with a blue border.
  - Map reads with bowtie 1 (ungapped):** Takes 'FASTQ file' as input and produces multiple outputs: 'output (sam)', 'output\_suppressed\_reads\_l (fastq)', 'output\_suppressed\_reads\_r (fastq)', 'output\_unmapped\_reads\_l (fastq)', and 'output\_unmapped\_reads\_r (fastq)'. It is highlighted with a blue border.
  - SAM-to-BAM:** Takes 'output1 (bam)' as input and produces 'output1 (bam)'.
- Details Panel (Right):** A configuration panel for the selected tool, 'Map reads with bowtie 1 (ungapped)', version 0.12.8. It includes options for selecting a reference genome, library type (Single-end), FASTQ file source, Bowtie settings (Full parameter list), and read filtering options (Skip the first n reads, Only align the first n reads, Trim n bases from high-quality, Trim n bases from low-quality).

# Workflows

Set input file(s)

**Galaxy / Galaxeast** Analyze Data Workflow Shared Data Visualization Admin Help User Using 34%

## Running workflow "chip workflow"

Expand All Collapse

**Step 1: Input dataset**

**Input Dataset**

4: chr10\_ctr2\_1.fastq.gz

type to filter

**Step 2: Map with Bowtie for Illumina (version 1.1.3)**

**Step 3: MACS (version 1.4.2)**

**Step 4: homer annotatePeaks (version 0.0.5)**

**Homer peaks OR BED format**

Output dataset 'output\_bed\_file' from step 3

**Genome version**

tair10

**Extra options**

Send results to a new history

**Action:**

Hide output 'out\_log'.

**Run workflow**

**History**

search datasets

**test**

1 shown, 3 deleted

120.7 MB

**4: chr10\_ctr2\_1.fastq**

format: fastqsanger, database: hg19

Set parameters

Run workflow

# Privacy

- By default datasets, workflows, histories are private to the user that generated/uploaded them.
- They can be shared across Galaxy users (of the same Galaxy instance) or via links