

Introduction to Galaxy

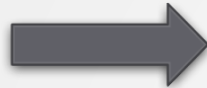
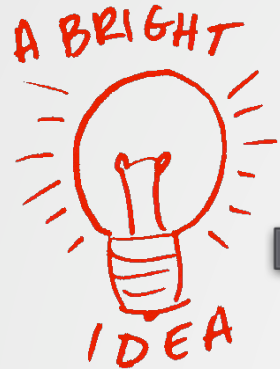
Stéphanie Le Gras
(slegras@igbmc.fr)

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

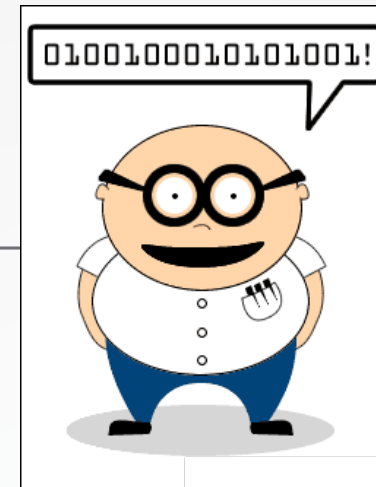
Bioinformatics analyses



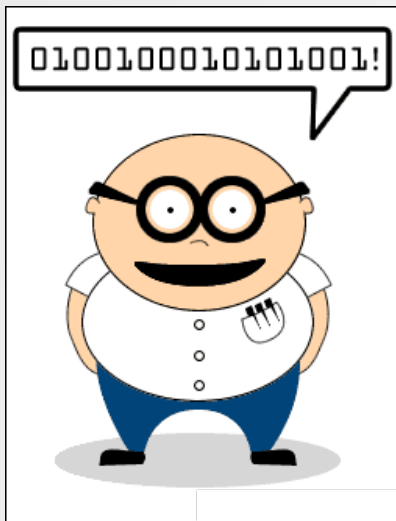
Informatics
data



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 );

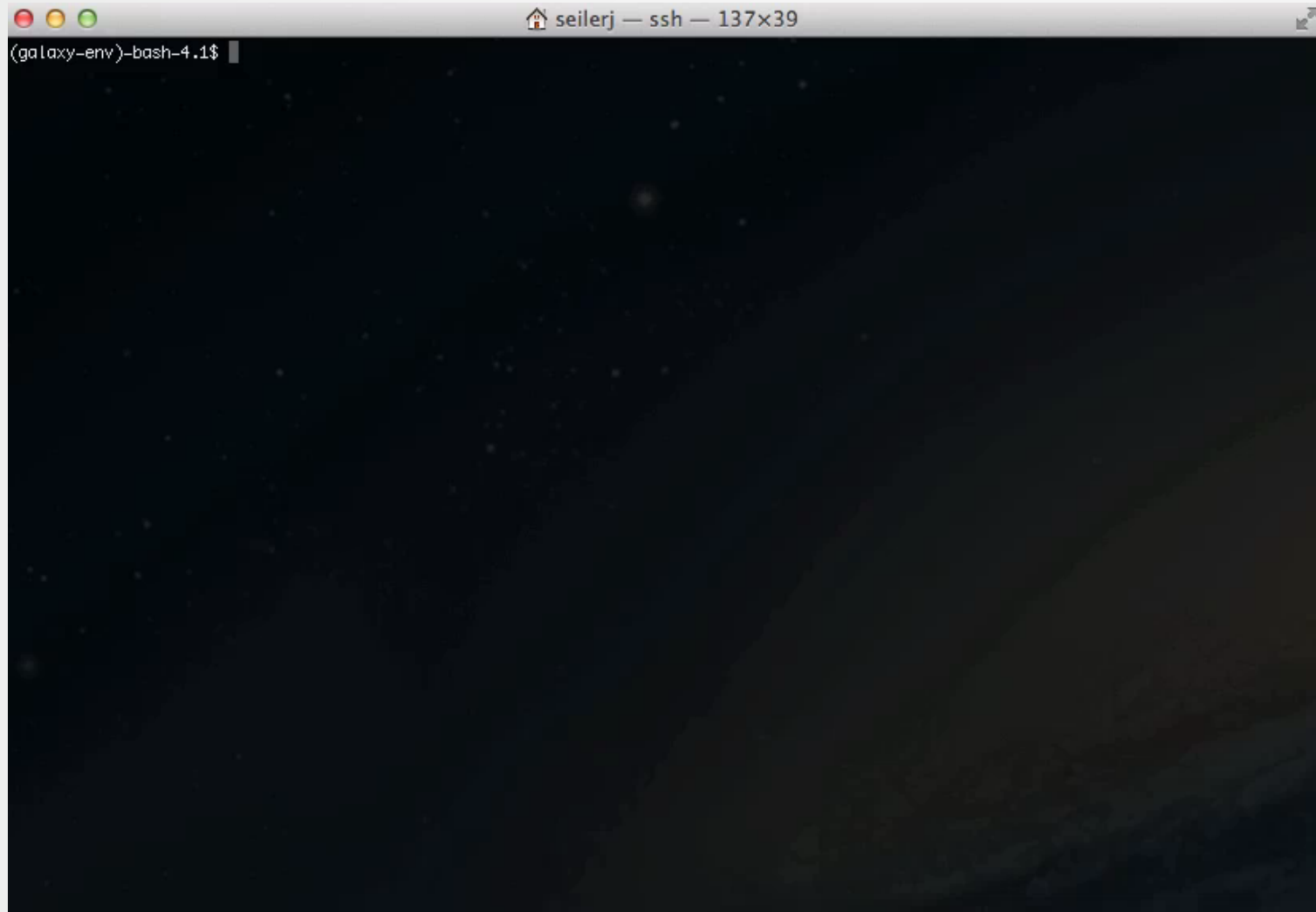
my %ids;
$out = ( defined $out ) ? $out : "results.txt";

## 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.
```

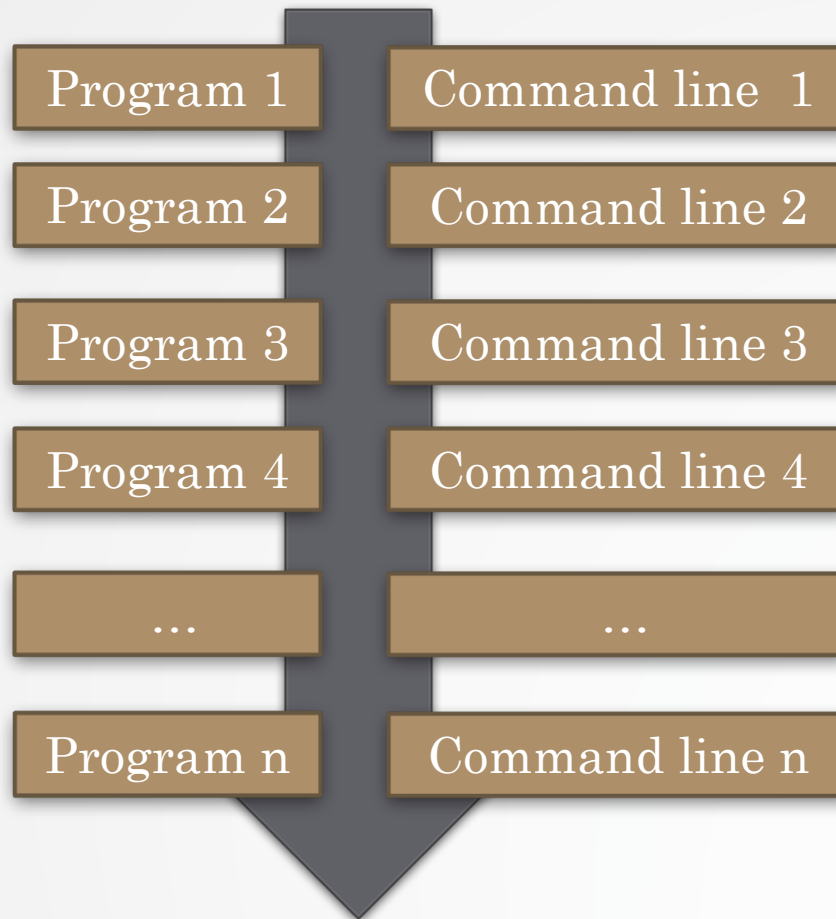
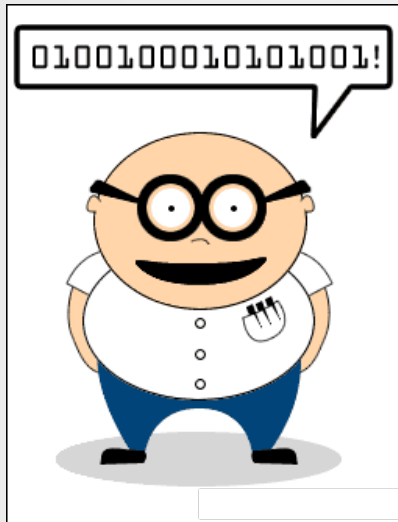
Command line

```
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**

Workflow tools

- CIPRES phylo.org
- eHive ensembl.org/info/docs/eHive/index.html
- Galaxy main.g2.bx.psu.edu
- Knime knime.org
- MobyLe mobyLe.pasteur.fr
- Taverna taverna.org.uk

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**.

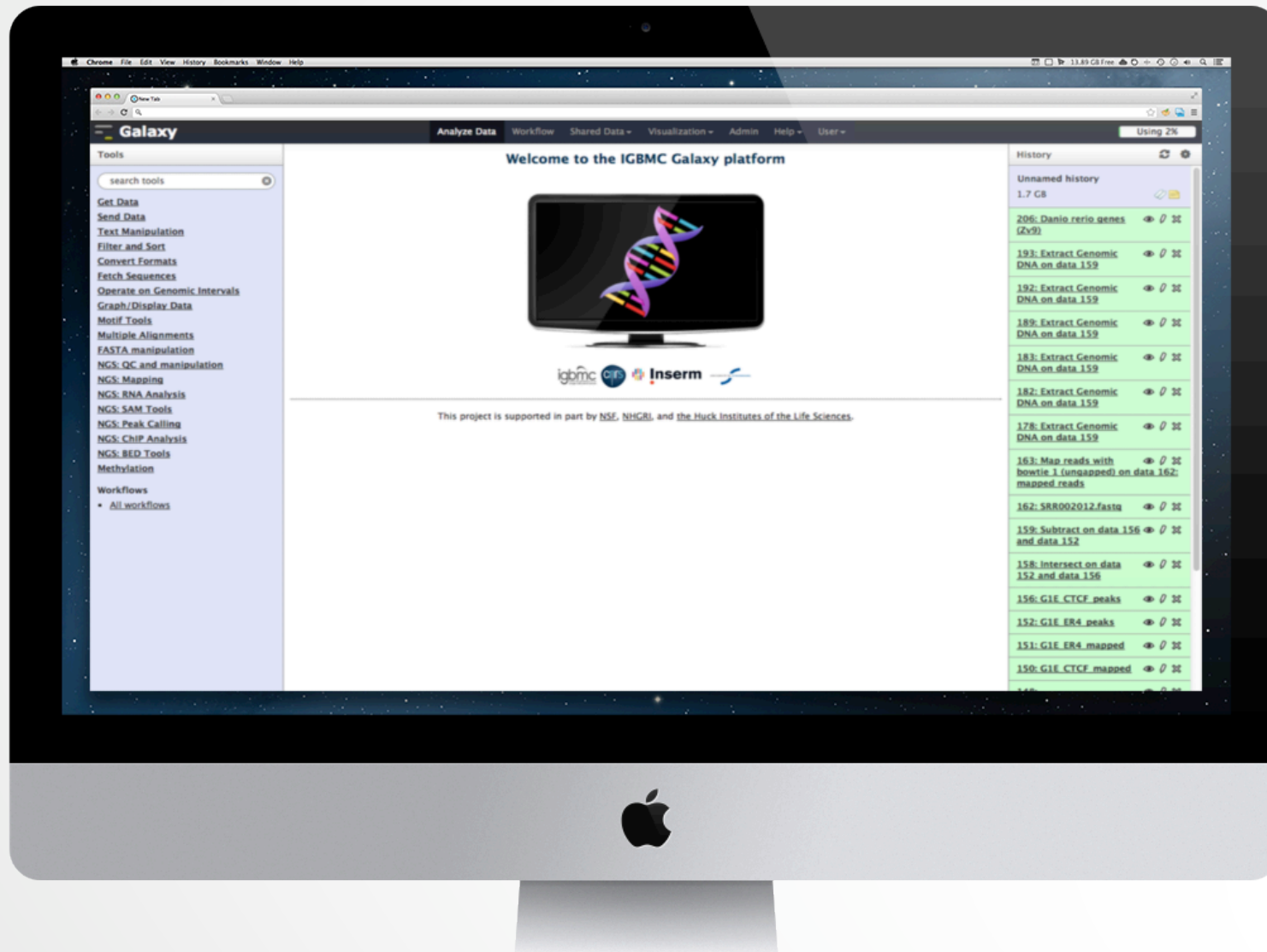


EMORY
UNIVERSITY

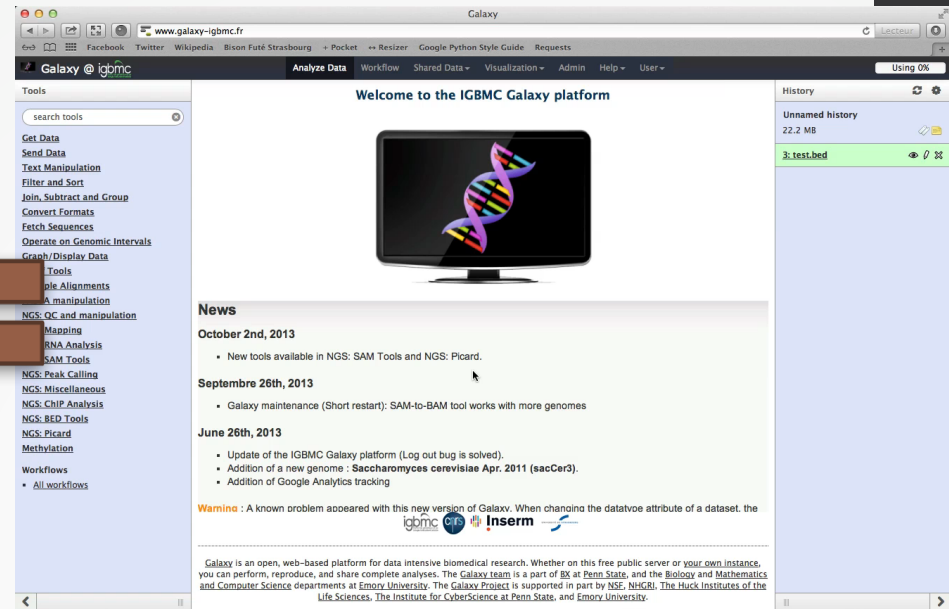
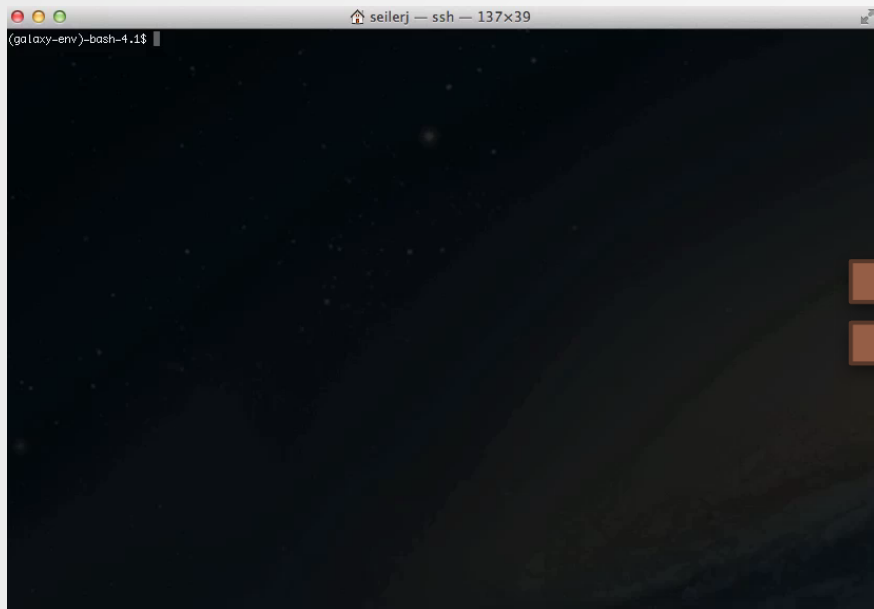


National
Human Genome
Research Institute

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

- **Public servers**
- **Local servers**
- Cloud
- Docker

Galaxy public servers

- Galaxy Project's public server (UseGalaxy.org)
- There are several public remote Galaxy instances worldwide
 - General Purpose servers (15)
 - Domain Servers (35)
 - Tool Publishing Servers (41)

Public Galaxy Servers list :

<https://wiki.galaxyproject.org/PublicGalaxyServers>

Last Update on: 2017, January 12nd

Galaxy public servers

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



Galaxy public servers

Your research institute



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

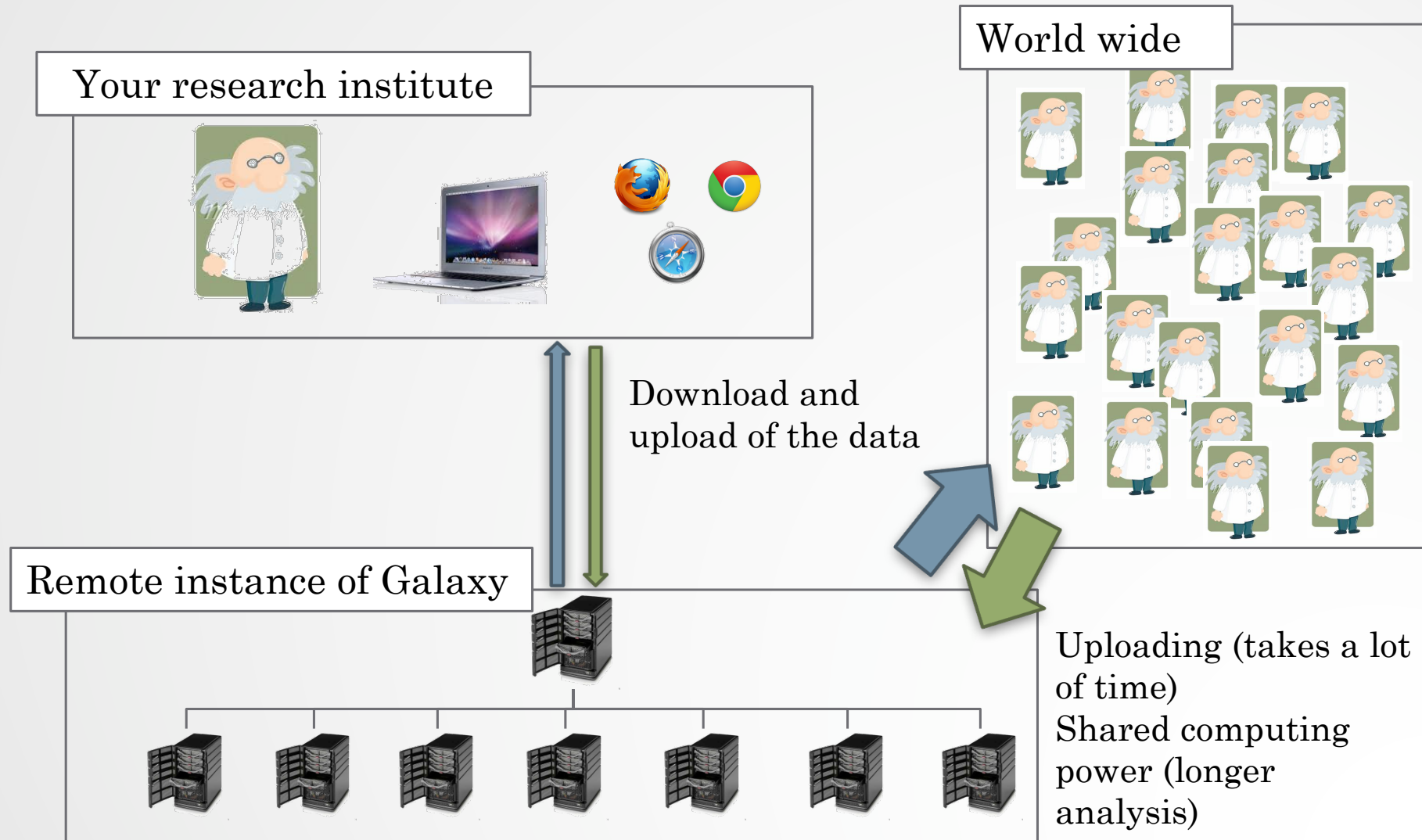
Download and
upload of the data

Remote instance of Galaxy



Run analyses

Galaxy public servers



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
 - plug-in new datasources
 - Develop Galaxy tools
 - Develop Galaxy itself



Description of the main features of Galaxy

Galaxy web interface

Top menu

The screenshot shows the Galaxy web interface with a dark blue top navigation bar containing the Galaxy logo and menu items: Analyze Data, Workflow, Shared Data, Visualization, Help, and User. The main content area is divided into three vertical panels. The left panel is the Tool panel, containing a search bar and a list of tool categories such as Get Data, Text Manipulation, and NGS tools. The center panel is the Data display and tools dialog window, featuring a large '080+' logo and the text 'Public Galaxy Servers and still counting'. The right panel is the History panel, showing a search bar and a message indicating that the history is empty. Three callout boxes with arrows point to these panels: 'Tool panel' on the left, 'Data display and tools dialog window' in the center, and 'History panel' on the right.

Galaxy

Analyze Data Workflow Shared Data Visualization Help User Using 0 bytes

Tools

search tools

Get Data
Lift-Over
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
Operate on Genomic Intervals
St
Gr
Cl
Ph
BE
Genome Dive
EMBOSS

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
Did we menti
registration E

History

search datasets

Unnamed history
0 b

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

Tool panel

Data display and tools dialog window

History panel

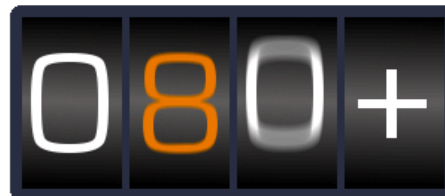
Top menu

Run workflows

Visualize your data

Get Help

The screenshot shows the Galaxy web interface. The top navigation bar includes 'Analyze Data', 'Workflow', 'Shared Data', 'Visualization', 'Help', and 'User'. The left sidebar lists various tool categories such as 'Get Data', 'Text Manipulation', and 'NGS: QC and manipulation'. The main content area displays a header with the text 'Galaxy is an open source, web-based platform for data intensive biomedical research' and a 'Log in/out, manage your account' button. The right sidebar shows 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'. A callout box labeled 'Run analyses' points to the 'Analyze Data' menu item. A callout box labeled 'Access public data' points to the 'Shared Data' menu item. A callout box labeled 'Log in/out, manage your account' points to the 'User' menu item.



Public Galaxy Servers
and *still* counting

Exercise 1 : Log in

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

Tool Panel / Run analyses

The screenshot displays the Galaxy web interface. At the top, the navigation bar includes 'Galaxy', 'Analyze Data', 'Workflow', 'Shared Data', 'Visualization', 'Help', 'User', and a 'Using 0%' indicator. On the left, the 'Tools' panel is visible, containing a search bar and a list of tool categories such as 'Get Data', 'Send Data', 'Text Manipulation', and various NGS tools. A callout box labeled 'Tool panel' with a downward arrow points to this list. The main content area features a banner for 'Public Galaxy Servers and still counting' with a '080+' logo and a tweet from @galaxyproject. On the right, the 'History' panel shows an empty state with a message: 'This history is empty. You can load your own data or get data from an external source'.

Tool Panel / Run analyses

Upload data

Search a tool

Tool category

Tool

The image displays two screenshots of a 'Tools' panel. The left screenshot shows a search bar labeled 'search tools' and a list of tool categories: Get Data, Lift-Over, 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, and NGS: BamTools. An 'Upload data' icon is in the top right. The right screenshot shows the same panel with the 'Convert Formats' category selected, displaying a list of tools: Convert BAM to Scldx, Tabular-to-FASTA converts tabular file to FASTA format, FASTA-to-Tabular converter, Tabular to FASTQ converter, FASTQ to Tabular converter, FASTQ to FASTA converter, and FASTQ to FASTA converter. A green arrow points from the 'Convert Formats' category in the left panel to the right panel. Labels with arrows point to the 'Upload data' icon, the search bar, the category list, and a specific tool.

Tools dialog window

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

Tools search tools

Convert BAM to Scidx (Galaxy Version 1.0.0) Options

BAM file
No bam dataset available.

Require proper mate-pairing?
Yes
Required if filtering by insert size for single-end Reqd 1 (below).

Read to output
Read1

Minimum insert size to output
Will not filter out single-end Read 1 unless proper mate-pairing is required (above).

Maximum insert size to output
Will not filter out single-end Read 1 unless proper mate-pairing is required (above).

Execute

What it does
Converts BAM data to Scidx, the Strand-specific coordinate count format, which is used by tools within the Chip-exo Galaxy toolshed. Some tools, like the strand-specific tools, require the position of the strand and the read number to create the Scidx file. For single-end reads, Read 1 will be used.

Options

- Require proper mate-pairing
- Minimum insert size
- Maximum insert size

Data display and tools dialog window

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

History search datasets

Unnamed history

0 b

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

Filter and Sort

History

The screenshot displays the Galaxy web interface. At the top, the navigation bar includes 'Galaxy', 'Analyze Data', 'Workflow', 'Shared Data', 'Visualization', 'Help', 'User', and a 'Using 0%' indicator. On the left, a 'Tools' sidebar lists various categories such as 'Get Data', 'Send Data', 'Text Manipulation', and 'NGS: QC and manipulation'. The main content area features a header for Galaxy, a central graphic with the number '080+' and the text 'Public Galaxy Servers and still counting', and a 'Tweets' section by @galaxyproject. On the right, the 'History' panel is highlighted with a green border. It contains a search bar for datasets, a message stating 'This history is empty. You can load your own data or get data from an external source', and a text box with the text 'History panel' and 'Keep track of each job run'.

History panel

Keep track of
each job run

History

View all histories

History options

Refresh History

Search datasets

History name

- HISTORY LISTS
 - Saved Histories
 - Histories Shared with Me
- HISTORY ACTIONS
 - Create New
 - Copy History
 - Share or Publish
 - Show Structure
 - Extract Workflow
 - Delete
 - Delete Permanently
- 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 four history panels. The top navigation bar includes 'Galaxy', 'Analyze Data', 'Workflow', 'Shared Data', 'Visualization', 'Help', 'User', and 'Using 0%'. Below the navigation bar are search bars for 'search histories' and 'search all datasets', along with a 'Create new' button. The main area is divided into four panels, each with a 'Switch to' dropdown menu.

- Current History:** Unnamed history, 0 b. A message states 'This history is empty'.
- History 1:** Unnamed history, 127.21 KB, 6 shown. Contains:
 - 6: L1spa_ORF1_1.fastq
 - 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_notRBPI +- 150_random80_adjacentSeq.fasta
- History 2:** Unnamed history, 97.7 KB, 2 shown, 3 deleted. Contains:
 - 5: Correspondance_JASPAR_CORE.txt
 - 4: fimo.txt
- History 3:** Unnamed history, 1.09 GB, 1 shown, 1 deleted. Contains:
 - 2: Brn2_Day2_rtta_rep2.sort.bed

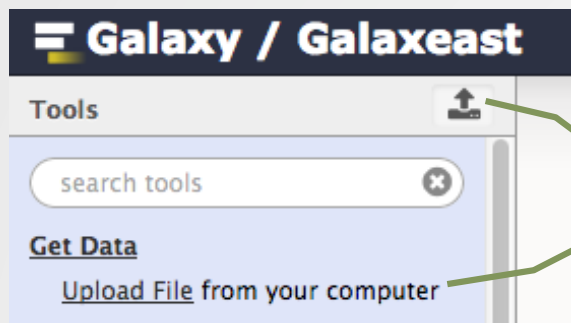
Exercise 2 : History

- 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 ▾ Visualiza (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	
Chip seq test dataset (chr10)	MITF test dataset for chipSeq - chr10	MITF test dataset for chipSeq - chr10	⊗
Chip seq test dataset (chr2)	Chip seq test dataset (chr2)	Chip seq test dataset (chr2)	⊗
Chromosome length	*.len files for several assemblies		⊗
Data demo AVIESAN	Data demo AVIESAN	Data demo AVIESAN	⊗
Data Rustenholz	Vitis vinifera data	Vitis vinifera data	⊗
Data_megadebug_jung	Data_megadebug_jung	Data_megadebug_jung	⊗
Data_megadebug_velt			⊗
EBA2016			⊗
Fimo database (updated Mar. 16 2015)			⊗
Fimo databases	Downloaded from MEME website	updated Jan 31 2013	⊗
Galaxy training - Chip seq test datasets	Galaxy training - Chip seq test datasets	Galaxy training - Chip seq test datasets	⊗
Galaxy training - RNA seq test datasets	Galaxy training - RNA seq test datasets	Galaxy training - RNA seq test datasets	⊗
Genome fasta files	Genome fasta files (GATK - IGBMC)	Fasta files used by the IGBMC microarray and sequencing platform (to be used with GATK)	⊗
GTF	Annotation files in GTF format	Annotation files in GTF format	⊗
Introduction 2 Galaxy (datasets)	Introduction 2 Galaxy (datasets)		⊗
Jaspar motifs ID <-> name	table of correspondences (for FIMO results...)		⊗
MITF test dataset for chipSeq	MITF test dataset for chipSeq	MITF (bam)	⊗
MITF test dataset for RNAseq	MITF test dataset for RNAseq		⊗
NGS course (Sep)	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 ↕	description	data type	size	time updated (UTC)	
..					
<input type="checkbox"/> ce10.len		len	98 bytes	2015-01-08 01:25	⌵
<input type="checkbox"/> dm3.len		len	227 bytes	2015-01-08 01:25	⌵
<input type="checkbox"/> hg19.len		len	376 bytes	2015-01-08 01:25	⌵
<input type="checkbox"/> mm10.len		len	1.4 KB	2015-01-08 01:25	⌵
<input type="checkbox"/> mm9.len		len	330 bytes	2015-01-08 01:25	⌵
<input type="checkbox"/> tair10.len		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: Mammal **genome:** Mouse **assembly:** Dec. 2011 (GRCm38/mm10)

group: Genes and Gene Predictions **track:** UCSC Genes

table: knownGene

region: genome position chr1:121427557-121432936

identifiers (names/accessions): paste list upload list

filter: create

intersection: create

correlation: create

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

get output summary/statistics

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](#).

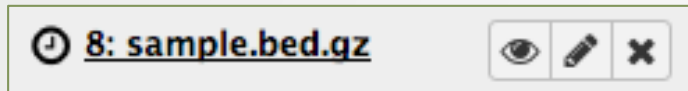
- clade:** Specifies which clade the organism is in.
- genome:** Specifies which organism data to use.

Exercise 3 : Import data to Galaxy

- 1. Import to Galaxy the file siLuc3_S12040.fastq from the “Shared data > Data Libraries > CNRS 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.
 - The genome is : Mouse (mm9)
 - The format is : bed

Datasets/Jobs in the History

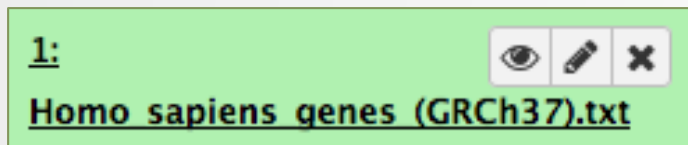
Grey: the job is waiting to run



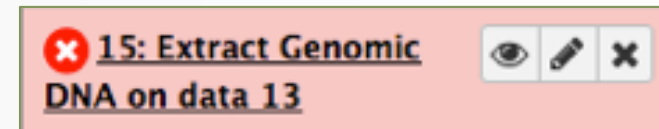
Yellow: the job is running



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

2: sample.bed   

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

uploaded bed file

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	chr1-193
chr1	64972363	64972563	chr1-648
chr1	134238383	134238583	chr1-134
chr1	51991430	51991630	chr1-518
chr1	53880739	53880939	chr1-537

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

The screenshot shows a dataset viewer for '2: sample.bed'. At the top right, there are three icons: an eye (view), a pencil (edit), and an 'X' (delete). Below these, the text reads '32,561 regions' and 'format: bed, database: mm9'. A text input field contains 'uploaded bed file'. Below this is a row of icons: a download icon, an information icon, a bar chart icon, a trash icon, and a speech bubble icon. The text below the icons lists various display options: 'display in IGB View', 'display at Ensembl Current', 'display with IGV local Mouse mm9', and 'display at UCSC main test'. At the bottom, there is a table with four columns: '1. Chrom', '2. Start', '3. End', and '4. Name'. The table contains five rows of data.

1. Chrom	2. Start	3. End	4. Name
chr1	193580486	193580686	chr1-193
chr1	64972363	64972563	chr1-648
chr1	134238383	134238583	chr1-134
chr1	51991430	51991630	chr1-518
chr1	53880739	53880939	chr1-537

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 blue header bar at the top containing a green progress indicator and the text "Using 20%". Below this is a "History" section with a search bar labeled "search datasets" and a refresh icon. The history list shows "RNA-seq data analysis" with "2 shown" items. The first item is "7.23 GB" with a checkmark, a shield icon, and a speech bubble icon. The second item is "2: sample.bed" with an eye icon, a pencil icon, and an 'x' icon. The third item is "1: siLuc3 S12040.fastq" with an eye icon, a pencil icon, and an 'x' icon. A green line points from the "Using 20%" indicator to the word "Quota" on the right. Another green line points from the "7.23 GB" text to the words "Size of history" on the left.

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

The screenshot shows the Galaxy web interface. The top navigation bar includes 'Galaxy', 'Analyze Data', 'Workflow', 'Shared Data', 'Visualization', 'Help', 'User', and 'Using 0%'. The left sidebar lists various tools and categories, with 'Workflows' and 'All workflows' highlighted at the bottom. The main content area features a text block about Galaxy, a '080+' logo for Public Galaxy Servers, and a 'Tweets' section with a message from the Galaxy Project. A green arrow points from the 'Workflow' menu item to the text 'Create, run, edit (...) workflows'. Another green arrow points from the 'All workflows' link in the sidebar to the text 'Run 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

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

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

Using 0%

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 for a Chip-seq workflow. The interface is divided into three main sections: Tools, Workflow Canvas, and Details.

Tools: A sidebar on the left contains 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, Database query, and Workflow control.

Workflow Canvas: The central workspace shows a grid with several workflow steps connected by lines. The steps are:

- Input dataset** (output)
- FastQC: Read QC** (Short read data from your current history, Contaminant list; output: html_file (html))
- Convert from BAM to BED** (Convert the following BAM file to BED; output: output (bed))
- Map reads with bowtie 1 (ungapped)** (FASTQ file; outputs: output (sam), output_suppressed_reads_l (fastq), output_suppressed_reads_r (fastq), output_unmapped_reads_l (fastq), output_unmapped_reads_r (fastq))
- SAM-to-BAM** (SAM File to Convert; output: output1 (bam))

The **Map reads with bowtie 1 (ungapped)** step is highlighted with a blue border.

Details: The right-hand panel provides configuration options for the selected tool, **Map reads with bowtie 1 (ungapped)**. The details include:

- Tool:** Map reads with bowtie 1 (ungapped)
- Version:** 0.12.8
- Will you select a reference genome from your history or use a built-in index?:** Use a built-in index
- Select a reference genome:** To be set at runtime
- Is this library mate-paired?:** Single-end
- FASTQ file:** Data input 'sinput1' (fastqsanger or fastqillumina or fastqsolexa)
- Bowtie settings to use:** Full parameter list
- Skip the first n reads (-s):** 0
- Only align the first n reads (-u):** -1
- Trim n bases from high-quality (left) end of each read before alignment (-5):** 0
- Trim n bases from low-quality (right) end of each read before alignment (-3):** 0

Workflows

Set input file(s)

The screenshot displays the Galaxy web interface for running a workflow. The main panel shows the workflow steps:

- 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: [edit icon]
 - Action: Hide output 'out_log'.

At the bottom of the workflow configuration, there is a checkbox for "Send results to a new history" and a "Run workflow" button.

The History panel on the right shows the dataset "4: chr10_ctr2_1.fastq" with format "fastqsanger" and 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