

Introduction to Galaxy

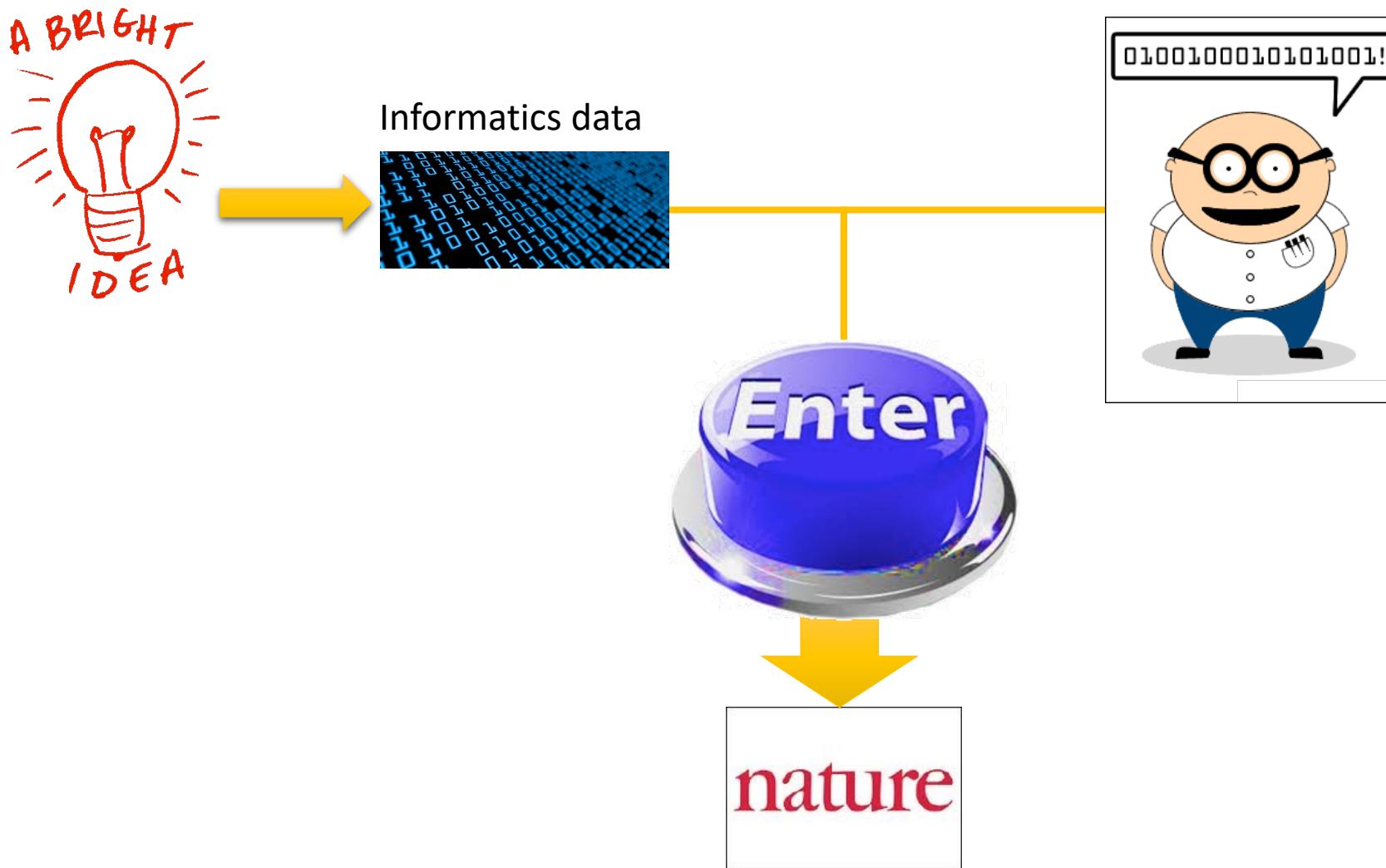
Stéphanie Le Gras
slegras@igbmc.fr

Guidelines

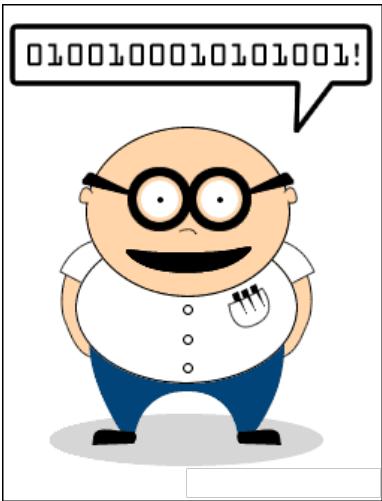
- Analyzing biological data with informatics tools
- Presentation of the Galaxy project
- Description of the main features of the Galaxy platform

Analyzing biological data with informatics tools

Bioinformatics analyses



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(
    "ids=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 );

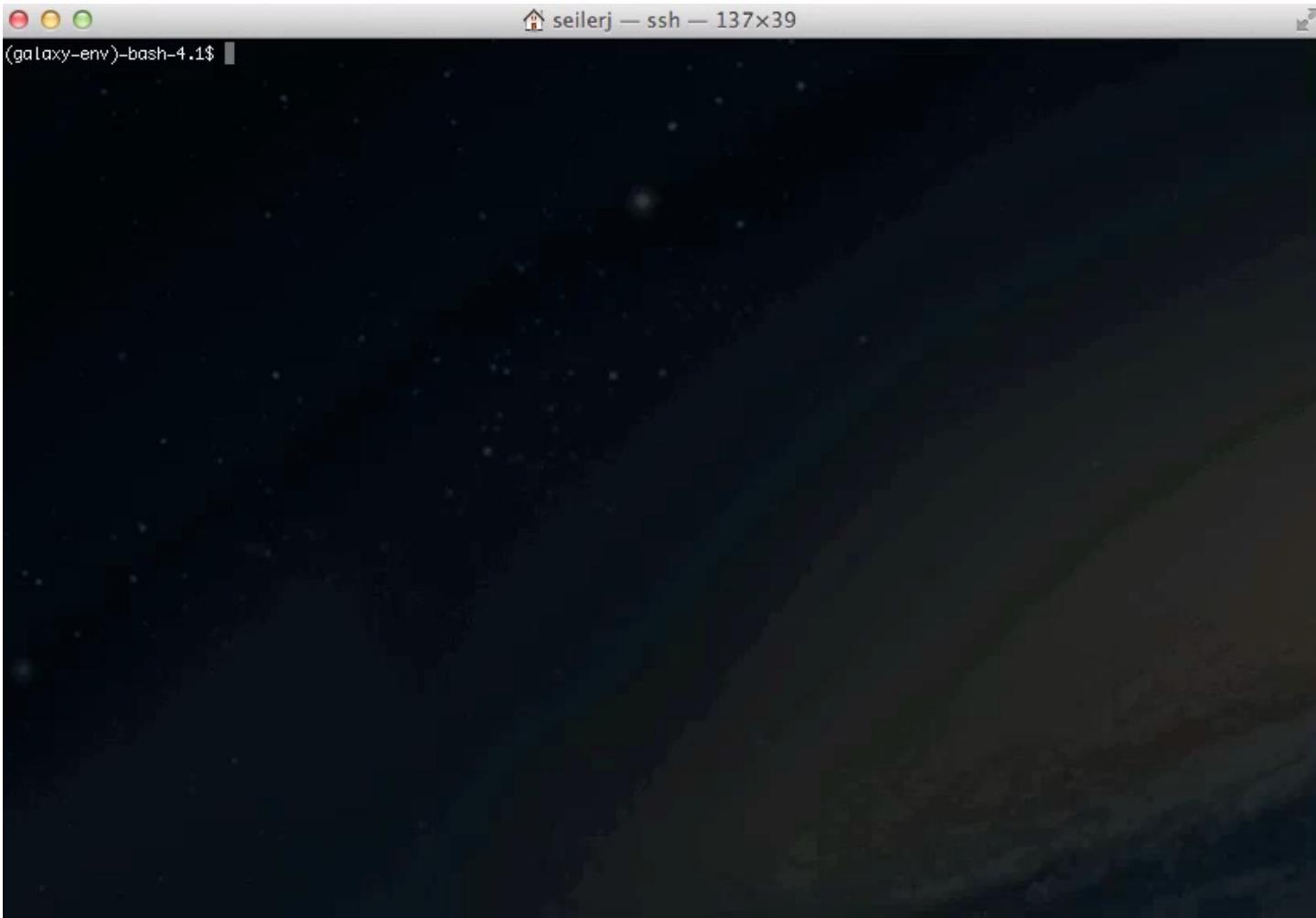
my %ids;
$out = ( defined $out ) ? $out : "results.txt";

## 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.
```

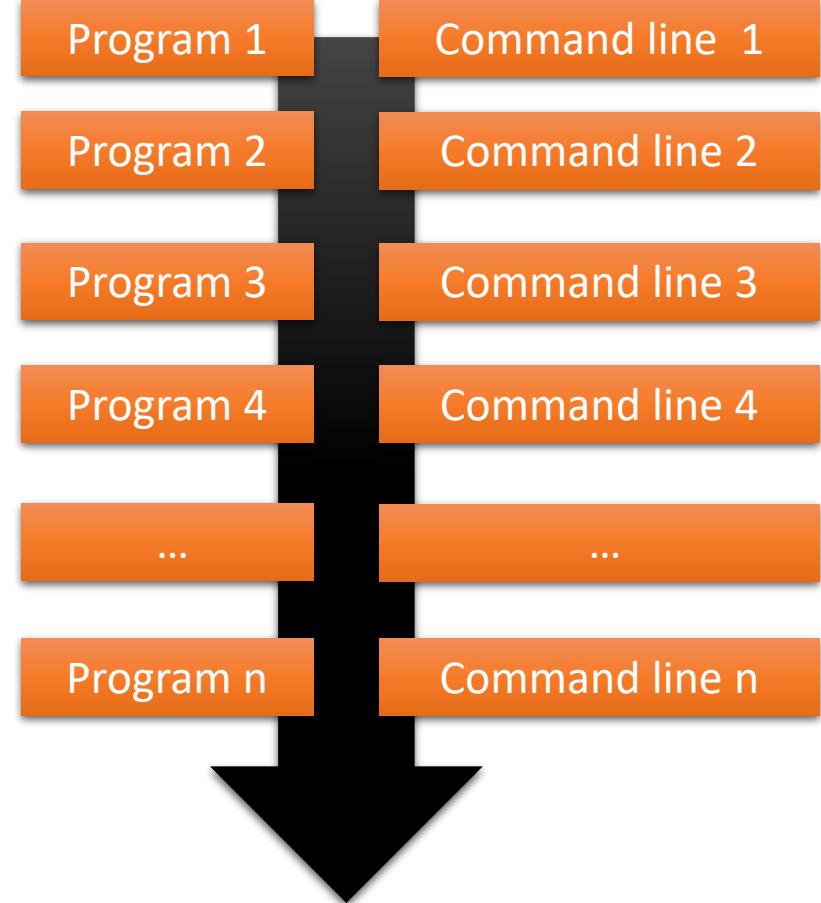
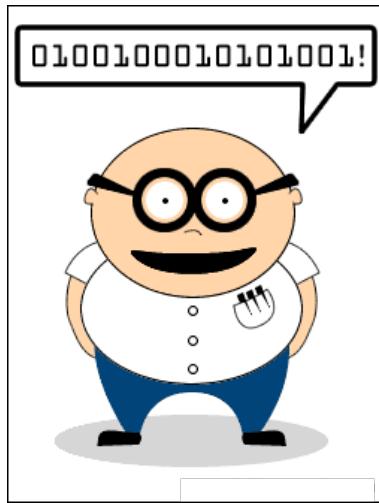
Command line

```
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 ?

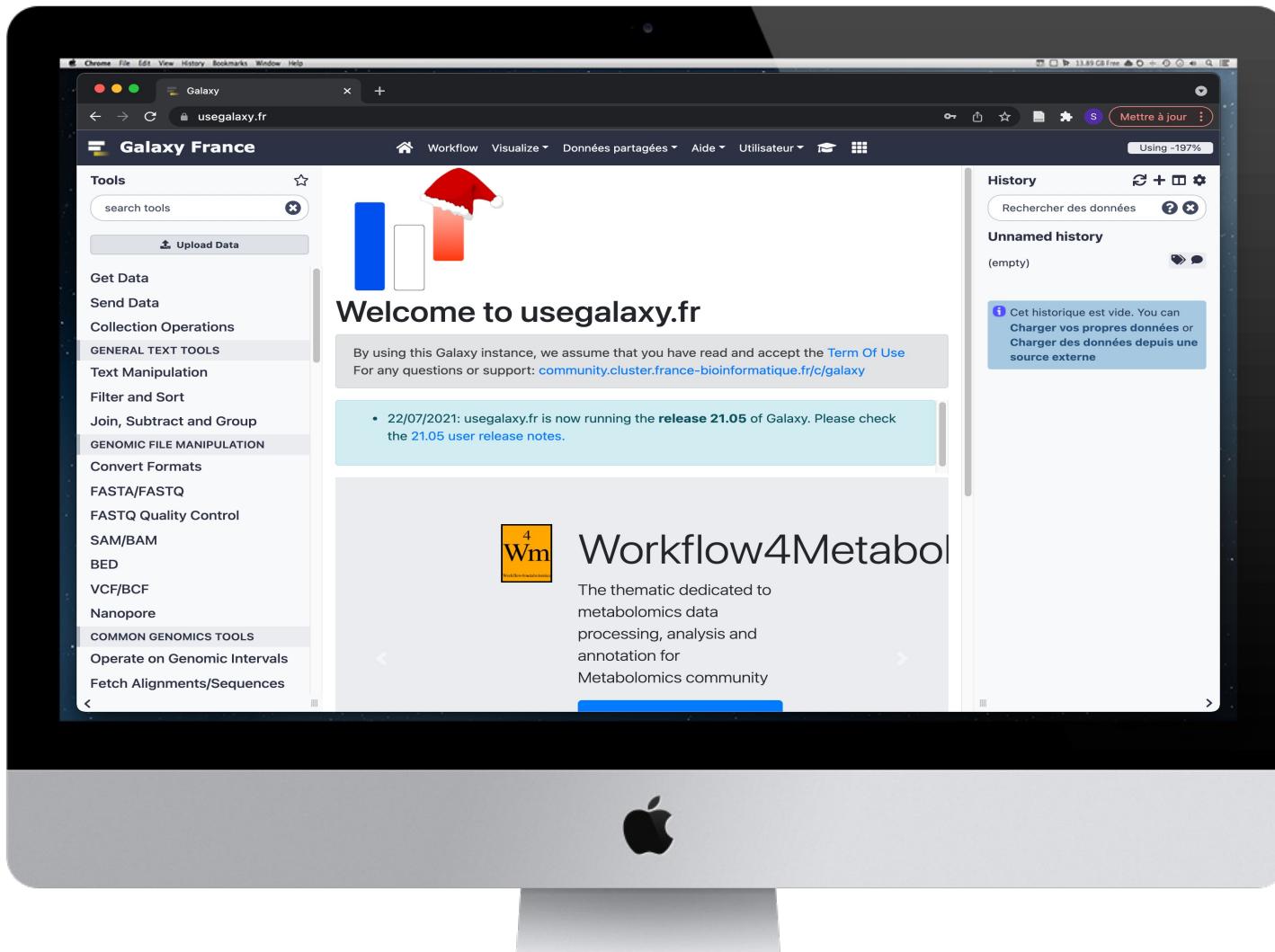
Galaxy is a **computing platform** that enables people to run **complex bioinformatics tools** on a **compute cluster** through a **simple web interface**.



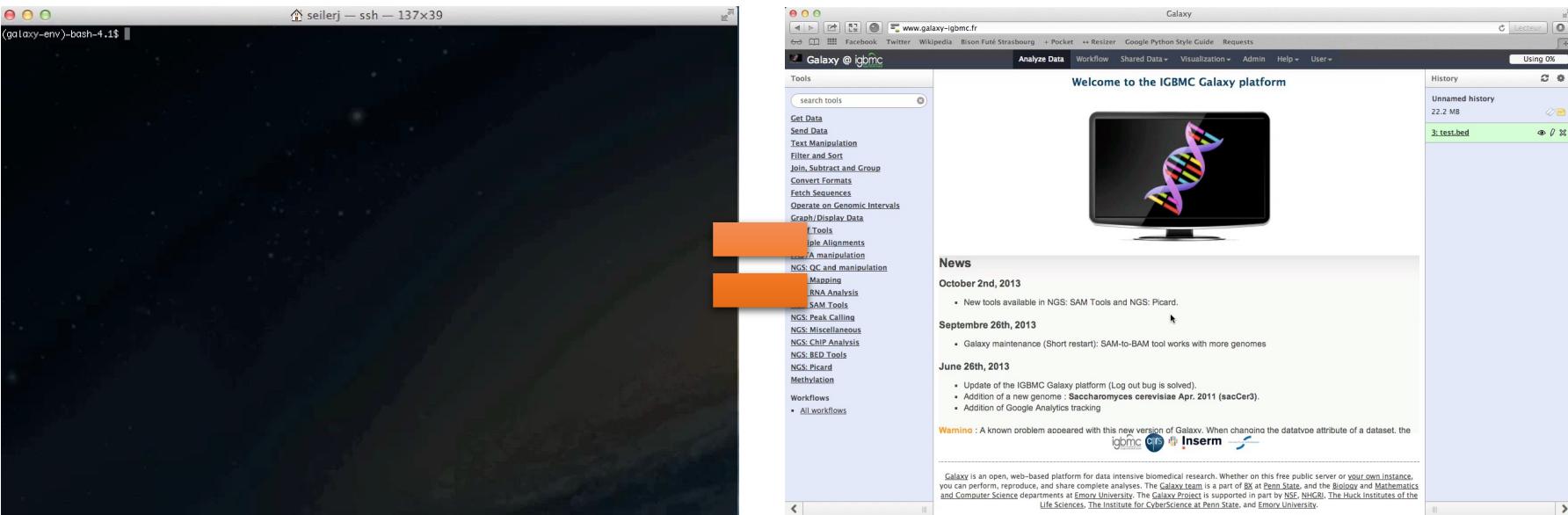
EMORY
UNIVERSITY



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

- **Public servers**
- **Local servers**
- Clouds (Public, Commercial or Academic)
- Docker
- Virtual Machines

Galaxy public servers

- Galaxy Project's public server (<https://usegalaxy.org/>) (3)
- There are several public remote Galaxy instances worldwide (136)
 - Genomics Servers
 - Domain Servers
 - Tool Publishing Servers

Public Galaxy Servers list :
<https://galaxyproject.org/use/>
Last Update on: 2023, December 13th

Galaxy public servers

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

¹ Depends on provider, and if you are eligible for the service

² These technologies can be deployed on clouds or locally.

³ Depends on the size of the system you are running it on.

⁴ With these technologies you can save the server and share the entire platform with them.

⁵ Depends on configuration.

* TIaaS: Training Infrastructure as a Service

- <https://galaxyproject.org/use/>

- 167 resources for using Galaxy (Last Update on: 2023, December 13th)

Galaxy public servers

Your research institute

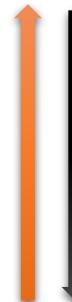


Connect to Galaxy web site through
a web browser
(<https://usegalaxy.org/>)

Remote instance of Galaxy

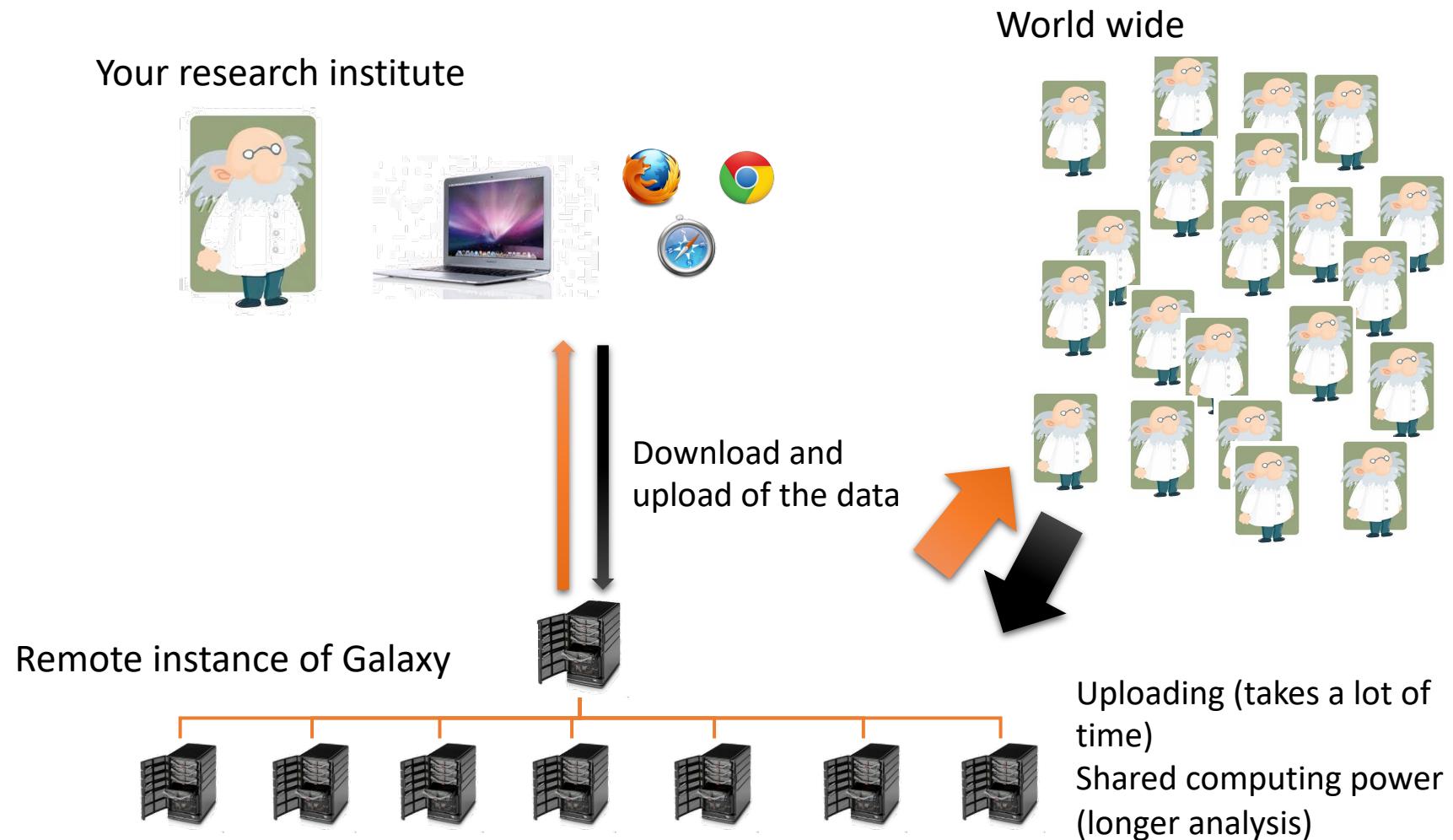


Run analyses



Download and
upload of the data

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
 - Develop Galaxy tools
 - Develop Galaxy itself



Description of the main features of Galaxy

Galaxy web interface

The screenshot illustrates the Galaxy web interface with several key components highlighted:

- Top menu:** Located at the top center, it includes links for Workflow, Visualize, Shared Data, Help, User, and a search bar.
- Tool panel:** Located on the left side, it contains sections for Tools, Get Data, Send Data, Collection Operations, GENERAL TEXT TOOLS (Text Manipulation, Filter and Sort, Join, Subtract and Group), GENOMIC FILE MANIPULATION (Convert Formats, FASTA/FASTQ, FASTQ Quality Control, SAM/BAM, BED, VCF/BCF), and a section for domain-specific subdomains (Workflow4Metabolomics).
- Data display and tools dialog window:** The central area displays a "Welcome to usegalaxy.fr" message, terms of use information, and a release note about the 22.05 release. It also features a "Domain specific subdomains" section for Workflow4Metabolomics, which includes a logo for "Workflow4Metabolomics" and a "Variant" visualization.
- History panel:** Located on the right side, it shows an empty history named "Unnamed history" with 0 B of data.

Annotations with orange arrows point to each of these four main areas: the Top menu, the Tool panel, the Data display and tools dialog window, and the History panel.

Top menu

The screenshot shows the Galaxy France web interface. At the top, there is a dark header bar with the title "Galaxy France". Below the header, the main content area has a "Welcome to usegalaxy.fr" message. On the left, there is a sidebar with a "Tools" section containing a search bar and an "Upload Data" button, followed by sections for "Get Data", "Send Data", "Collection Operations", "GENERAL TEXT TOOLS" (which is currently selected), "Text Manipulation", "Filter and Sort", "Join, Subtract and Group", "GENOMIC FILE MANIPULATION" (which is also selected), "Convert Formats", and "FASTA/FASTQ". In the center, there is a large orange button labeled "Run analyses" with an upward arrow pointing to it. Above the "Run analyses" button, there are two smaller orange boxes: one labeled "Run workflows" with a downward arrow pointing to the "Workflow" menu item in the top navigation bar, and another labeled "Access public data" with a downward arrow pointing to the "Shared Data" menu item. To the right of the "Run analyses" button, there are two more orange boxes: one labeled "Log in/out, manage your account" with an upward arrow pointing to the "User" icon in the top navigation bar, and another labeled "Get Help" with a curved arrow pointing to the "Help" icon. Further to the right, there is a box labeled "Get access to training materials" with a curved arrow pointing to the "Training Materials" icon in the top navigation bar. On the far right, there is a "History" panel showing an empty history named "Unnamed history".

Galaxy France

Using 79%

Tools

search tools

Upload Data

Get Data

Send Data

Collection Operations

GENERAL TEXT TOOLS

Text Manipulation

Filter and Sort

Join, Subtract and Group

GENOMIC FILE MANIPULATION

Convert Formats

FASTA/FASTQ

Workflow Visualize Shared Data Help User Training Materials

Run workflows

Access public data

Get Help

Get access to training materials

Welcome to usegalaxy.fr

By using this Galaxy instance, we assume that you have read and accepted the [Terms of Use](#). For any questions or support: community.cluster.france-bioinformatique.fr/c/galaxy

- 28/11/2022: usegalaxy.fr is now running the **22.05 release** of Galaxy. Please check the [22.05 user release notes](#).

Domain specific

History

search datasets

Unnamed history

0 B 0 This history is empty. You can load your own data or get data from an external source.

Hands On

Exercise 1

History

Galaxy France

Workflow Visualize Shared Data Help User

Tools

search tools

Upload Data

Get Data

Send Data

Collection Operations

GENERAL TEXT TOOLS

Text Manipulation

Filter and Sort

Join, Subtract and Group

GENOMIC FILE MANIPULATION

Convert Formats

FASTA/FASTQ

FASTQ Quality Control

SAM/BAM

BED

VCF/BCF

Nanopore

COMMON GENOMICS TOOLS

Operate on Genomic Intervals

Fetch Alignments/Sequences

Workflow4MetabolomicsCovid19

4 Wm

Variant

Welcome to usegalaxy.fr

By using this Galaxy instance, we assume that you have read and accepted the [Terms of Use](#). For any questions or support: [community.cluster.france-bioinformatique.fr/c/galaxy](#)

- 28/11/2022: usegalaxy.fr is now running the **22.05 release** of Galaxy. Please check the [22.05 user release notes](#).

Domain specific subdomains:

The same instance, usegalaxy.fr, but with only the dedicated tools in order to focus on the domain

Using 79%

History

search datasets

Unnamed history

0 B 0 0

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

History panel
Keep track of each job run

History

History options

Switch history

Create new history

Search datasets

History name

Unnamed history

0 B

Refresh History

The screenshot shows the 'History' interface. At the top is a search bar labeled 'search datasets'. Below it is a section labeled 'Unnamed history' with a size indicator of '0 B'. At the bottom are several icons: a checkmark, a download arrow, a location pin, a refresh symbol, and a gear. Orange arrows point from the text labels to their corresponding elements in the interface.

You have 45 histories.

- Show Histories Side-by-Side
- Resume Paused Jobs
- Copy this History
- Delete this History
- Export Tool Citations
- Export History to File
- Extract Workflow
- Share or Publish
- Set Permissions
- Make Private

View all histories

Galaxy France Using 50%

Tools

VCFtoTab-delimited

Upload Data

Show Sections

VCFtoTab-delimited: Convert VCF data into TAB-delimited format

WORKFLOWS

All workflows

search datasets in selected histories

RNA-seq data analysis

- 10 : Filter on data 9
- 9 : siMitfVssiLuc.up.annot.txt
- 8 : siMitfVssiLuc.up.txt
- 7 : mart_export.txt.gz
- 6 : FastQC on data 1: Raw Data
- 5 : FastQC on data 1: Web page
- 4 : FastQC on data 1: Raw Data
- 3 : FastQC on data 1: Web page
- 1 : siLuc3_S12040.fastq.gz

NGS data analysis training Strasbourg

- 27 : Ctrl_chr10.fastq
- 26 : Mitf_chr10.fastq
- 23 : ctrl.bam
- 22 : mitf.bam
- 21 : siMitfVssiLuc.up.annot.txt
- 20 : htseq-count on siMitf 4
- 19 : htseq-count on siMitf 3
- 18 : htseq-count on siLuc 3
- 17 : htseq-count on siLuc 2
- 16 : RNA STAR on siLuc2_ other_protocol: mapped.b

Unnamed history

- 16 : RNA STAR on collection 4: mapped.bam
- 15 : RNA STAR on collection 4: splice_junctions.bed
- 14 : RNA STAR on collection 4: log
- 7 : Homo_sapiens.GRCh3 8.105.chr.gtf.gz
- 4 : Trimmomatic across collection 3
- 3 : test collection

Copy of 'NGS data analysis training Strasbourg - RNASeq'

- 65 : UCSC Main on Mous e: knownGene (chr4:17,57 7,198-17,607,970)
- 64 : Cut on data 63
- 63 : Join two Datasets on data 62 and data 55
- 62 : pi3k_akt_signalling_genes.txt
- 61 : Cut on data 60
- 58 : Cut on data 57
- 57 : Filter on data 55
- 56 : Filter on data 55
- 55 : siMitfVssiLuc.up.annot.txt
- 54 : mart_export.txt.gz
- 53 : siMitfVssiLuc.up.txt

History

DNA-seq data analysis

- 12 : FreeBayes on data 1 and d ata 10 (variants)
- 11 : CaptureDesign_chr4.bed
- 10 : MarkDuplicates on data 8: BA M
- 9 : MarkDuplicates on data 8: tabular
- 8 : Map with BWA-MEM on data 3 and data 2 (mapped reads in BA M format)
- 7 : FastQC on data 3: RawData
- 6 : FastQC on data 3: Webpage
- 5 : FastQC on data 2: RawData
- 4 : FastQC on data 2: Webpage
- 3 : CRN-107_11-R2.fastq.gz
- 2 : CRN-107_11-R1.fastq.gz
- 1 : sample.bed.gz

<https://usegalaxy.fr/datasets/964b95495a4315be/edit>

Hands On

Exercise 2

Import data into Galaxy

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

Import your own data to Galaxy

The screenshot shows the Galaxy France web interface. On the left, a sidebar lists "Tools", "Get Data", "Send Data", "Collection Operations", and "GENERAL TEXT TOOLS". A red arrow points from the "Upload Data" button in the main toolbar to the "Upload Data" button in the sidebar. Another red arrow points from the "Upload Data" button in the sidebar to the "Upload Data" button in the main toolbar. The main area has a dark header with the text "Display the drag and drop utility used to upload local files". Below the header, there's a "Download from web or upload from disk" section with tabs for "Regular", "Composite", "Collection", and "Rule-based". A message says "You added 1 file(s) to the queue. Add more files or click 'Start' to proceed.". A dashed box highlights a dataset in the queue table:

Name	Size	Type	Genome	Settings	Status
CRN-107_11-R1.fastq	18.5 MB	Auto-de...	----- Additional ...	⚙️	0% (progress bar)

Annotations point to specific columns in the table:

- "Name of the dataset" points to the "Name" column.
- "Size of the dataset" points to the "Size" column.
- "File format" points to the "Type" column.
- "Genome" points to the "Genome" column.

At the bottom, there are buttons for "Choose local files", "Choose remote files", "Paste/Fetch data", "Start", "Pause", "Reset", and "Close".

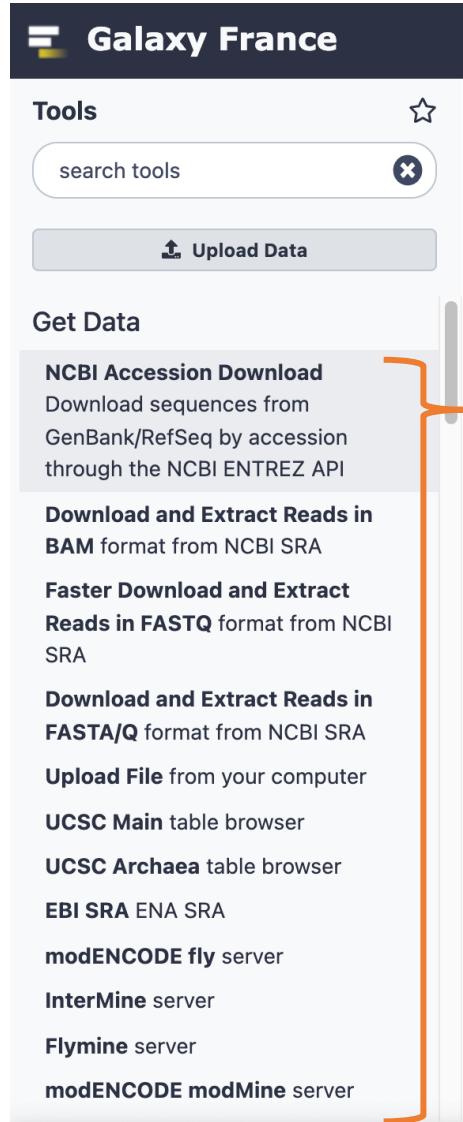
Import shared data (data libraries)

The screenshot shows the Galaxy France web interface. At the top, there is a navigation bar with links for Workflow, Visualize, Shared Data, Help, User, and a grid icon. A dropdown menu for 'Shared Data' is open, showing options: Histories, Access published resources (which is highlighted with a red arrow labeled '1.'), Workflows, Visualizations, and Pages. Below the navigation bar is a search bar with a 'Search' button and a checkbox for 'exclude restricted'. The main content area displays a table titled 'List of data libraries' with columns for 'Name' and 'Description'. The table lists seven entries:

Name	Description
ProteoRE	ProteoRE datasets
covid-19	
GTN - Material	Galaxy Training Network Material Galaxy Training Network Material. See ht ... (more)
workflow4metabolomics	Workflow4Metabolomics referenced histori ... (more) https://workflow4metabolomics.org/refere ... (more)
Roscoff 2021	Data for Assembly and Annotation trainin ... (more)
EBAII A&A 2022	Ecole EBAII Assemblage & Annotation sept ... (more)
Formation sRNA 2022	Data for Formation sRNA 2022

At the bottom, there is a pagination control with buttons for '«', '<', '1', '>', and '»', followed by '10' and 'per page, 7 total'.

Import public data



The screenshot shows the Galaxy France interface. At the top, there's a search bar labeled "search tools" and a button "Upload Data". Below that, under "Get Data", there are several options:

- NCBI Accession Download**: Download sequences from GenBank/RefSeq by accession through the NCBI ENTREZ API.
- Download and Extract Reads in BAM format from NCBI SRA**
- Faster Download and Extract Reads in FASTQ format from NCBI SRA**
- Download and Extract Reads in FASTA/Q format from NCBI SRA**
- Upload File** from your computer
- UCSC Main table browser**
- UCSC Archaea table browser**
- EBI SRA ENA SRA**
- modENCODE fly server**
- InterMine server**
- Flymine server**
- modENCODE modMine server**

Browse and import external data from public databases



The screenshot shows the Table Browser interface. At the top, there's a navigation bar with links for Genomes, Genome Browser, Tools, Mirrors, Downloads, My Data, Help, and About Us. The main section is titled "Table Browser" and contains the following text:

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.

Below this, there are several input fields and buttons:

- clade: Mammal
- genome: Mouse
- assembly: Dec. 2011 (GRCm38/mm10)
- group: Genes and Gene Predictions
- track: UCSC Genes
- table: KnownGene
- region: genome position chr1:27557-121432936
- identifiers (names/acccessions):
- 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

At the bottom, there are buttons for "get output" and "summary/statistics", and a link to reset user cart settings.

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.

Hands On

Exercise 3

Datasets/Jobs in the History

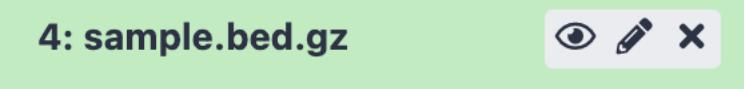
Grey: the job is waiting to run



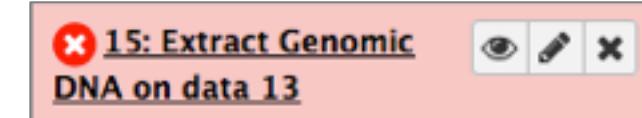
Orange: the job is running



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

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

The screenshot shows a Nextflow history entry for a dataset named "sample.bed.gz". The entry includes the following details:

- Name: 1 : sample.bed.gz
- Description: 32,561 regions, format bed, database mm9
- Type: uploaded bed file
- Genome: mm9

The table displays the first few lines of the BED file:

1.Chrom	2.Start	3.End	4.Name
chr1	193580486	193580686	chr1-1934573
chr1	64972363	64972563	chr1-6486016
chr1	134238383	134238583	chr1-1341694
chr1	51991430	51991630	chr1-5187923
chr1	53880739	53880939	chr1-53768546

Datasets/Jobs in the History

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

Download dataset

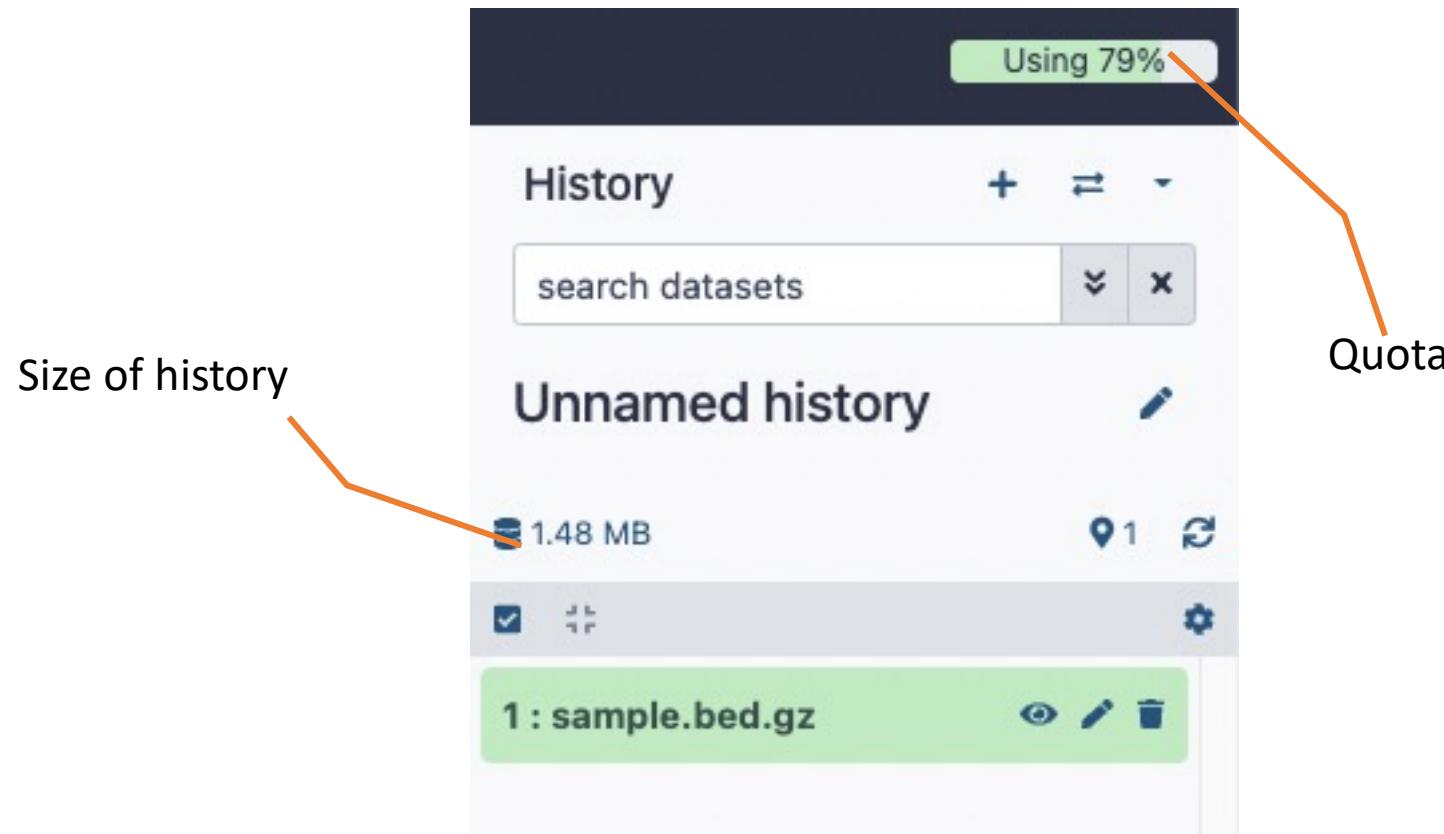
The screenshot shows a Galaxy dataset card for a file named 'sample.bed.gz'. The card has a light green header with the file name '1 : sample.bed.gz' and three icons: a magnifying glass, a pencil, and a trash can. Below the header, there's a small icon of a key and the text '32,561 regions format bed, database mm9'. A large button labeled 'uploaded bed file' is centered below this information. At the bottom of the card, there's a table with four columns: '1.Chrom', '2.Start', '3.End', and '4.Name'. The table contains five rows of genomic data for chromosome 1.

1.Chrom	2.Start	3.End	4.Name
chr1	193580486	193580686	chr1-19345731
chr1	64972363	64972563	chr1-64860161
chr1	134238383	134238583	chr1-13416941
chr1	51991430	51991630	chr1-51879231
chr1	53880739	53880939	chr1-53768541

Delete dataset

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

Size of histories and quota



Hands On

Exercise 4

Tool Panel / Run analyses

The screenshot shows the Galaxy France web interface. On the left, a vertical orange-highlighted sidebar titled "Galaxy France" contains a "Tools" section with a search bar and an "Upload Data" button, followed by a list of tool categories: Get Data, Send Data, Collection Operations, GENERAL TEXT TOOLS, GENOMIC FILE MANIPULATION, Convert Formats, FASTA/FASTQ, FASTQ Quality Control, SAM/BAM, BED, VCF, NGS, and CO. A large orange callout box labeled "Tool panel" points to this sidebar. At the bottom of the sidebar, there are buttons for Fetch Aliases, Sequences, and Other. The main content area features a colorful bar chart graphic and the text "Welcome to usegalaxy.fr". Below this, a message states: "By using this Galaxy instance, we assume that you have read and accepted the [Terms of Use](#). For any questions or support: community.cluster.france-bioinformatique.fr/c/galaxy". A blue callout box highlights a recent release note: "• 28/11/2022: usegalaxy.fr is now running the **22.05 release** of Galaxy. Please check the [22.05 user release notes](#)." To the right, a "History" panel shows an "Unnamed history" entry with 0 B of data. A message in this panel says: "This history is empty. You can load your own data or get data from an external source." At the bottom, there are links for "Workflow4MetabolomicsCovid19", "Wm", and "Variant".

Using 79%

Galaxy France

Tools

search tools

Upload Data

Get Data

Send Data

Collection Operations

GENERAL TEXT TOOLS

Text Manipulation

Filter and Sort

Join, Subtract and Group

GENOMIC FILE MANIPULATION

Convert Formats

FASTA/FASTQ

FASTQ Quality Control

SAM/BAM

BED

VCF

NGS

CO

Fetch Aliases

Sequences

Other

Tool panel

Welcome to usegalaxy.fr

By using this Galaxy instance, we assume that you have read and accepted the [Terms of Use](#). For any questions or support: community.cluster.france-bioinformatique.fr/c/galaxy

- 28/11/2022: usegalaxy.fr is now running the **22.05 release** of Galaxy. Please check the [22.05 user release notes](#).

Domain specific subdomains:

The same instance, usegalaxy.fr, but with only the dedicated tools in order to focus on the domain

Workflow4MetabolomicsCovid19

4 Wm

Variant

History

search datasets

Unnamed history

0 B

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

Using 79%

42

Tool Panel / Run analyses

The screenshot shows the Tool Panel interface. On the left, there's a sidebar with a search bar labeled "Search a tool" and a dropdown menu labeled "Tools". Below these are several categories: "Get Data", "Send Data", "Collection Operations", "GENERAL TEXT TOOLS" (which is highlighted), "Text Manipulation", "Filter and Sort", "Join, Subtract and Group", "GENOMIC FILE MANIPULATION" (highlighted), "Convert Formats", "FASTA/FASTQ", "FASTQ Quality Control", and "SAM/BAM". On the right, a large orange arrow points from the "Text Manipulation" section of the sidebar to a detailed view of the same section. This detailed view includes a title "Text Manipulation" and a list of tools: "Add Header", "csvtk-cut and keep selected columns", "annotateMyIDs", "Regex Find And Replace", "Column Regex Find And Replace", "Histogram of a numeric column", "Compute on rows", and "Split by group".

Upload data

Search a tool

Tool category

Tools

search tools

Upload Data

Get Data

Send Data

Collection Operations

GENERAL TEXT TOOLS

Text Manipulation

Filter and Sort

Join, Subtract and Group

GENOMIC FILE MANIPULATION

Convert Formats

FASTA/FASTQ

FASTQ Quality Control

SAM/BAM

Tool

Text Manipulation

Add Header Add header row with column names to tabular data

csvtk-cut and keep selected columns

annotateMyIDs annotate a generic set of identifiers

Regex Find And Replace

Column Regex Find And Replace

Histogram of a numeric column

Compute on rows

Split by group

Tools dialog window

The screenshot shows the Galaxy France web interface. The top navigation bar includes links for Workflow, Visualize, Shared Data, Help, User, and a History section indicating "Using 79%". The left sidebar contains a "Tools" section with a search bar and an "Upload Data" button, followed by categories: Get Data, Send Data, Collection Operations, GENERAL TEXT TOOLS (Text Manipulation, Filter and Sort, Join, Subtract and Group), GENOMIC FILE MANIPULATION (Convert Formats, FASTA/FASTQ, FASTQ Quality Control, SAM/BAM, BED, VCF/BCF, Nanopore), and COMMON GENOMICS TOOLS (Operate on Genomic Intervals, Fetch Alignments/Sequences). The main content area features a colorful graphic and the heading "Welcome to usegalaxy.fr". It includes a message about accepting terms of use and links to user release notes. A large orange box at the bottom lists four key functions: Set parameters, Run tools, Get help on tools, and Display content of dataset.

Galaxy France

Workflow Visualize Shared Data Help User History Using 79%

Tools

search tools

Upload Data

Get Data

Send Data

Collection Operations

GENERAL TEXT TOOLS

- Text Manipulation
- Filter and Sort
- Join, Subtract and Group

GENOMIC FILE MANIPULATION

- Convert Formats
- FASTA/FASTQ
- FASTQ Quality Control
- SAM/BAM
- BED
- VCF/BCF
- Nanopore

COMMON GENOMICS TOOLS

- Operate on Genomic Intervals
- Fetch Alignments/Sequences

Welcome to usegalaxy.fr

By using this Galaxy instance, we assume that you have read and accepted the [Terms of Use](#). For any questions or support: community.cluster.france-bioinformatique.fr/c/galaxy

- 28/11/2022: usegalaxy.fr is now running the **22.05 release** of Galaxy. Please check the [22.05 user release notes](#).

History

search datasets

Unnamed history

0 B 0

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

Hands On

Exercise 5

Hands On

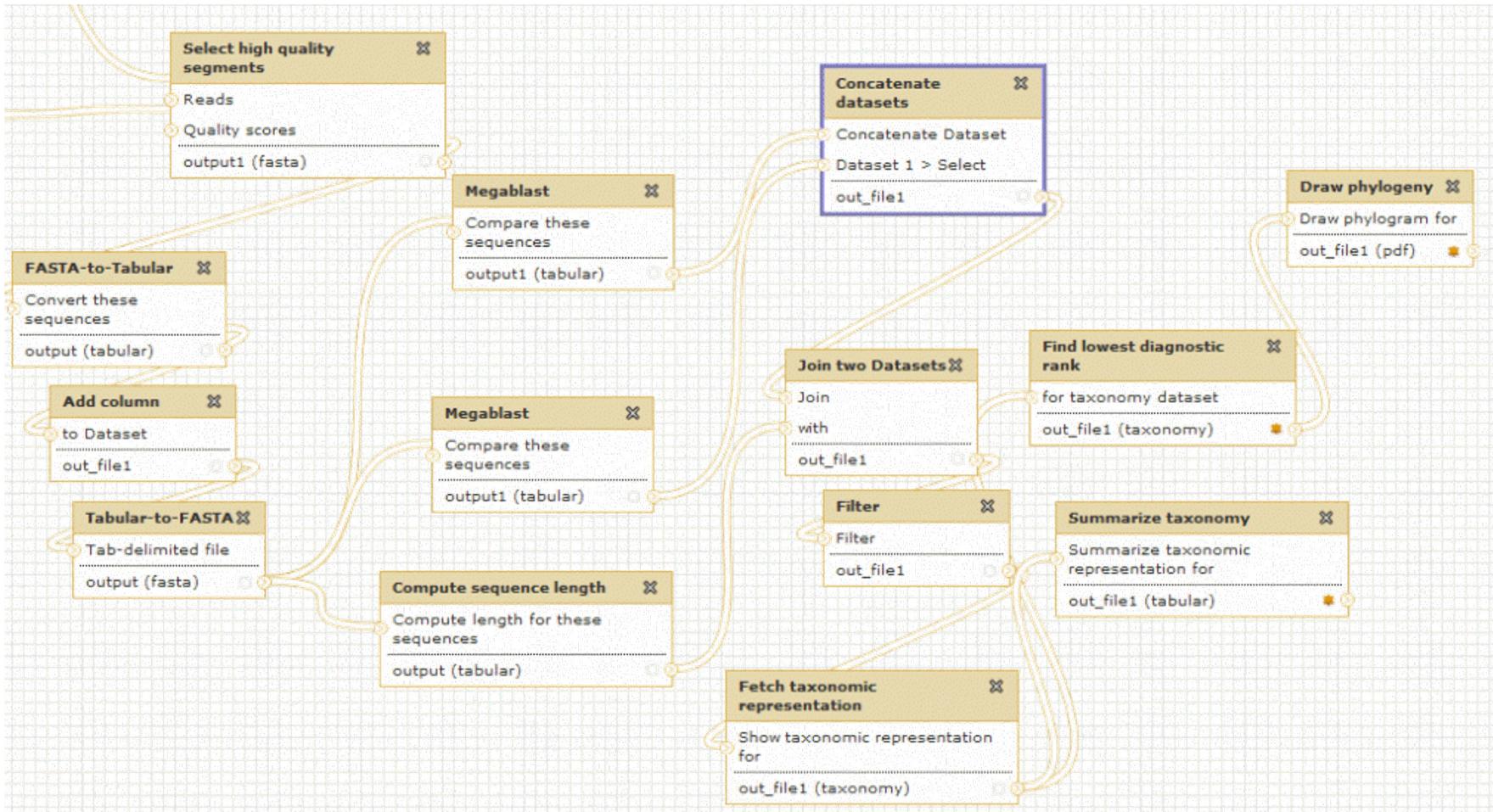
Exercise 6

Workflow

What if we'd mix all together



Galaxy workflow



Galaxy workflows

- Workflow:
 - Analysis protocol with several steps (tools)
 - The output of a step is used as the input of the next so file formats between two steps should be compatible!
- Workflows are often made general so that they can be run on various datasets
- Some of the parameters are pre-defined while others are set at runtime

Workflows

The screenshot shows the Galaxy France web interface. The top navigation bar includes links for Home, Workflow, Visualize, Shared Data, Help, User, and a search bar indicating "Using 79%". The left sidebar lists various tool categories: Tools, Get Data, Send Data, Collection Operations, GENERAL TEXT TOOLS (Text Manipulation, Filter and Sort, Join, Subtract and Group), GENOMIC FILE MANIPULATION (Convert Formats, FASTA/FASTQ, FASTQ Quality Control, SAM/BAM, BED, VCF/BCF, Nanopore), COMMON GENOMICS TOOLS (Operate on Genomic Intervals, Fetch Alignments/Sequences), and GENOMICS ANALYSIS (Annotation, Assembly). A central panel displays a welcome message: "Welcome to usegalaxy.fr" with a note about using the instance and a link to the community. Below this is a "Domain specific subdomains:" section with a subdomain for "Workflow4MetabolomicsCovid19" featuring a logo with "4 Wm" and a COVID-19 variant icon. To the right, a "History" panel shows a dataset named "1 : sample.bed.gz". An orange callout box highlights the text "Create, run, edit (...) workflows". The bottom right corner contains the number "51".

Galaxy France

Workflow Visualize Shared Data Help User

Using 79%

Tools

search tools

Upload Data

Get Data

Send Data

Collection Operations

GENERAL TEXT TOOLS

Text Manipulation

Filter and Sort

Join, Subtract and Group

GENOMIC FILE MANIPULATION

Convert Formats

FASTA/FASTQ

FASTQ Quality Control

SAM/BAM

BED

VCF/BCF

Nanopore

COMMON GENOMICS TOOLS

Operate on Genomic Intervals

Fetch Alignments/Sequences

GENOMICS ANALYSIS

Annotation

Assembly

Welcome to usegalaxy.fr

By using this Galaxy instance, we assume that you have read and acc
For any questions or support: [community.cluster.france-bioinformatics](#)

- 28/11/2022: usegalaxy.fr is now running the **22.05 release** of Galaxy. Please check the [22.05 user release notes](#).

Create, run, edit (...) workflows

Domain specific subdomains:

The same instance, usegalaxy.fr, but with only the dedicated tools in order to focus on the domain

Workflow4MetabolomicsCovid19

4 Wm

Variant analysis, consensus using community

Metabolomics data processing, analysis and annotation for Metabolomics

1 : sample.bed.gz

History

search datasets

Unnamed history

1.48 MB

51

Workflows

Create workflows

The screenshot shows the Galaxy France web interface. On the left, a sidebar lists various genomic tools and analysis categories. A red box highlights the "List of Workflows" section under the "SHIVI/DAWI" heading. An orange arrow points from this box to the "Create workflows" button at the top right of the main content area. The main content area displays a list of available workflows, each with a name, updated date, sharing status, and a play button for execution. The history panel on the right shows an unnamed history containing a single dataset named "sample.bed.gz".

Galaxy France

Workflow Visualize Shared Data Help User

Using 79%

Tools

search tools

+ Create Import

Search Workflows

Name Tags Updated Sharing Bookmarked

- ChIP-seq data analysis workflow
- BED Ensembl to BED UCSC
- Unnamed workflow
- ChIP-seq data analysis workflow
- Run "Prepare RNA-seq data for seqMINER"
- Run ChIP-seq hands-on
- ChIP-seq data analysis workflow
- BED Ensembl to BED UCSC
- DNA-seq data analysis (DU Dijon)

History

search datasets

Unnamed history

1.48 MB 1

1 : sample.bed.gz

52

Workflow creation

The screenshot shows the Galaxy France interface for workflow creation. On the left, a sidebar lists various tool categories: Tools, Inputs, Get Data, Send Data, Collection Operations, Expression Tools, GENERAL TEXT TOOLS, GENOMIC FILE MANIPULATION, COMMON GENOMICS TOOLS, and others like FASTA/FASTQ, BED, VCF/BCF, Nanopore, and Fetch Alignments/Sequences. A red arrow points from the text "Add tools or input datasets to the workflow" at the bottom left towards the "COMMON GENOMICS TOOLS" section. The main workspace is titled "My workflow" and contains a grid for tool placement. To the right, configuration options include:

- Name:** My workflow
- Version:** 1: Dec 13th 2022, 0 steps
- Annotation:** (Empty box with note: "These notes will be visible when this workflow is viewed.")
- License:** Specify a license for this workflow.
- Creator:** Add a new creator - either a person or an organization.
- Tags:** (Icon of a tag) Apply tags to make it easy to search for and find items with the same tag.

At the bottom right, there is a navigation bar with icons for back, forward, and search.

Add tools or input datasets to the workflow

Workflow creation

The screenshot shows the Galaxy France workflow creation interface. On the left, a sidebar lists various tools under categories like 'Inputs', 'Collection Operations', and 'FASTA/FASTQ'. The main workspace displays two workflow steps: '1: Input Dataset' and '2: Filter'. Step 1 has an output named 'output (input)'. Step 2 has a filter tool with an input named 'out_file1 (input)'. To the right of the steps is a panel for 'Input Dataset' configuration, which includes fields for 'Label', 'Step Annotation', 'Optional' status (set to 'No'), 'Format(s)', and 'Tag filter'. A large orange box highlights the 'Input dataset' step, and another orange box highlights the 'Tool to be run' section of the 'Filter' step.

Input dataset.
Most of the time, a workflow starts with an input dataset to which analyses are applied. In Galaxy, the file format of the input dataset will be limited to the input file format of the subsequent step

Tool to be run

Workflow creation

The screenshot shows the Galaxy France workflow creation interface. On the left, there's a sidebar with a search bar for 'Tools' containing 'filter'. Below it, under 'Inputs', is a section titled 'Filter and Sort' with several options like 'Filter data on any column using simple expressions' and 'Filter sequences by ID from a tabular file'. There are also sections for 'Collection Operations', 'FASTA/FASTQ', and 'SAM/BAM'. A central workspace shows two workflow steps: '1: Input Dataset' and '2: Filter'. Step 1 has an output labeled 'output (input)' and 'output is dataset'. Step 2 has an input labeled 'out_file1 (input)'. A green arrow connects the output of step 1 to the input of step 2. To the right of the workspace is a panel for 'Input Dataset' with fields for 'Label', 'Step Annotation', 'Optional' (set to 'No'), 'Format(s) - optional', and 'Tag filter - optional'. A large orange callout box points to the green arrow with the text: 'If two steps can be chained, the output arrow of one box and the inputs of the other boxes are green'.

If two steps can be chained, the output arrow of one box and the inputs of the other boxes are green

Workflow creation

Galaxy France Workflow Visualize Shared Data Help User Using 50%

Tools Unnamed workflow

Inputs

Filter and Sort

- Filter data on any column using simple expressions
- Filter sequences by ID from a tabular file
- Filter GFF data by attribute using simple expressions
- Filter GFF data by feature count using simple expressions
- Filter GTF data by attribute values_list

Collection Operations

- Filter failed datasets
- Filter empty datasets
- Filter collection

FASTA/FASTQ

- Filter FASTA on the headers and/or the sequences
- Filter by quality
- Filter FASTQ reads by quality score and length
- Filter sequences by length

BAM filter

- Removes reads from a BAM file based on criteria

SAM/BAM

- Filter BAM datasets on a variety of attributes

filter

The screenshot shows the Galaxy web interface for workflow creation. On the left, a sidebar lists various tools under categories like 'Inputs', 'Collection Operations', and 'FASTA/FASTQ'. A search bar at the top of the sidebar contains the word 'filter'. In the center, a workflow graph is displayed with two nodes: '1: Input Dataset' and '2: Filter'. The 'Input Dataset' node has an output port labeled 'output (input)'. The 'Filter' node has an input port labeled 'Filter' and an output port labeled 'out_file1 (input)'. An orange arrow points from a callout box containing text about pre-configuration to the 'With following condition' section of the 'Filter' tool's configuration panel. The configuration panel includes fields for 'Tool Parameters' (e.g., 'Filter *', 'Data input 'input' (tabular)', 'c1==chr22'), 'With following condition *' (containing 'c1==chr22'), and 'Number of header lines to skip *'. The top right of the interface shows a status bar with 'Using 50%'.

Pre-configure tool parameters and configure parameters to be set at run time

Filter
data on any column using simple expressions
(Galaxy Version 1.1.1)

Label

Add a step label.

Step Annotation

Add an annotation or notes to this step.
Annotations are available when a workflow is viewed.

Conditionally skip step?

No

Set to true and connect a boolean parameter that determines whether step will be skipped

Tool Parameters

Filter *

Data input 'input' (tabular)
Dataset missing? See TIP below.

With following condition *

c1==chr22

Double equal signs, ==, must be used as shown above. To filter for an arbitrary string, use the Select tool.

Number of header lines to skip *

Workflow creation

Galaxy France Workflow Visualize Shared Data Help User Tools Unnamed workflow Using 50%

Inputs Filter and Sort

- Filter data on any column using simple expressions
- Filter sequences by ID from a tabular file
- Filter GFF data by attribute using simple expressions
- Filter GFF data by feature count using simple expressions
- Filter GTF data by attribute values_list

Collection Operations

- Filter failed datasets
- Filter empty datasets
- Filter collection

FASTA/FASTQ

- Filter FASTA on the headers and/or the sequences
- Filter by quality
- Filter FASTQ reads by quality score and length
- Filter sequences by length

SAM/BAM

- Filter BAM datasets on a variety of attributes

Save workflow Run

Filter data on any column using simple expressions (Galaxy Version 1.1.1)

Label

Add a step label.

Step Annotation

Add an annotation or notes to this step. Annotations are available when a workflow is viewed.

Conditionally skip step?

No

Set to true and connect a boolean parameter that determines whether step will be skipped

Tool Parameters

Filter *

Data input 'input' (tabular)
Dataset missing? See TIP below.

With following condition *

c1=='chr22'

Double equal signs, ==, must be used as shown above. To filter for an arbitrary string, use the Select tool.

Number of header lines to skip *

```
graph LR; A[1: Input Dataset] --> B[2: Filter]; B -- "c1=='chr22'" --> C[Number of header lines to skip *]
```

Run workflows

The screenshot shows the Galaxy France web interface. On the left, a sidebar lists various tool categories. In the center, a workflow titled "Workflow: DNA-seq data analysis" is being configured. The workflow consists of four steps:

- Step 1: Read 2 (fastq)**
Input: CRN-107_11-R1.fastq.gz
- Step 2: Read 1 (fastq)**
Input: CRN-107_11-R1.fastq.gz
- Step 3: Capture design (bed)**
Input: CaptureDesign_chr4.bed
- Step 4: FastQC (Galaxy Version 0.73+galaxy0)**
Raw read data from your current history * required
Connected to 'output' from Step 1
Contaminant list - optional
Nothing selected

On the right, the "History" panel shows an unnamed history containing three datasets: CaptureDesign_chr4.bed, CRN-107_11-R1.fastq.gz, and CRN-107_11-R2.fastq.gz. A "Run Workflow" button is located at the top right of the workflow area.

Annotations with orange arrows and boxes:

- An arrow points from the text "Set input file(s). It has to be a dataset from your current history" to the input field for Step 1.
- An arrow points from the text "Run workflow" to the "Run Workflow" button.
- An arrow points from the text "Set parameters" to the "FastQC" step parameters.

Hands On

Exercise 7

Hands
On

Exercise 8

Hands
On

Exercise 9

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