

Introduction to Galaxy

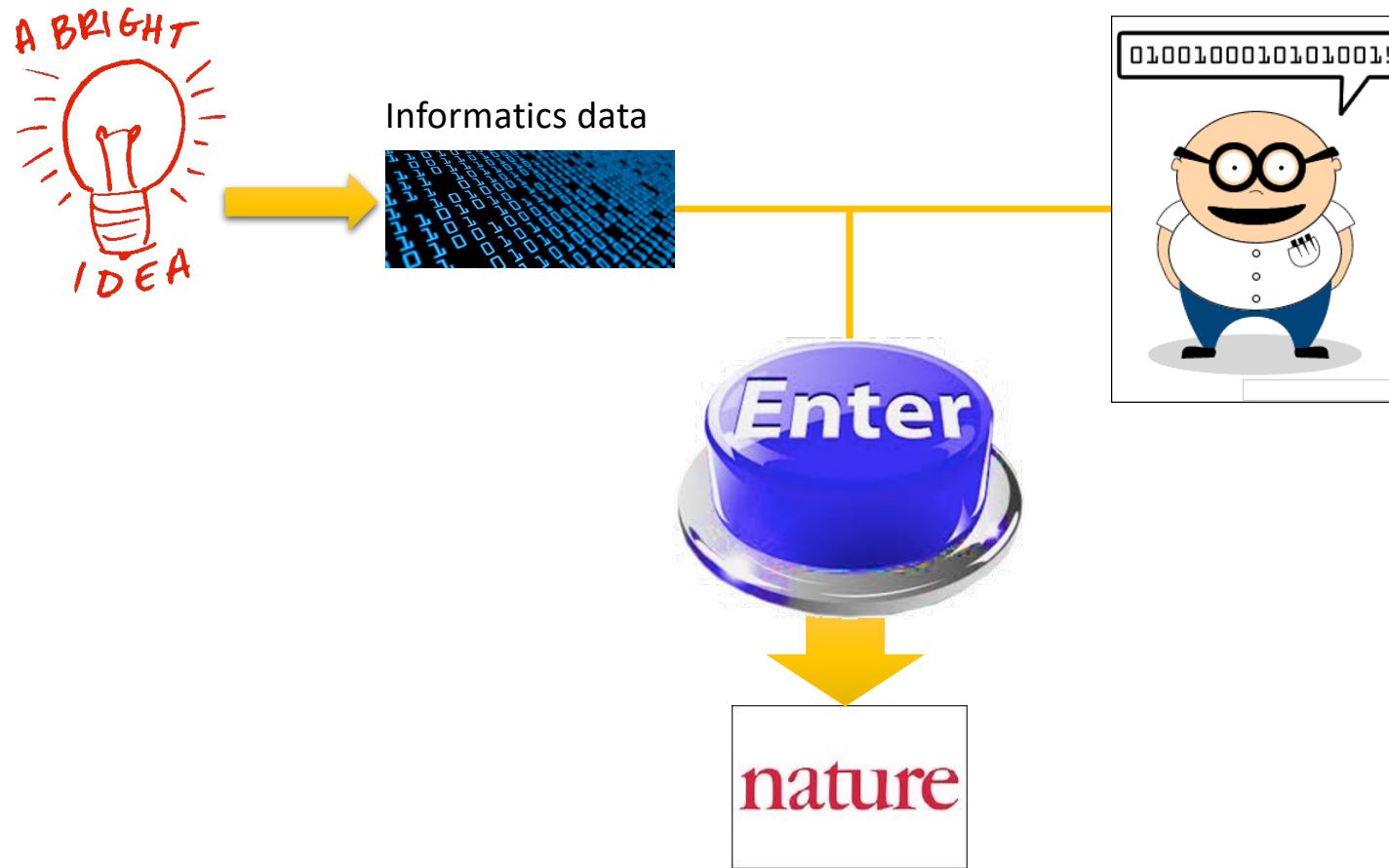
Stéphanie Le Gras
[\(slegras@unistra.fr\)](mailto:slegras@unistra.fr)

Guidelines

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

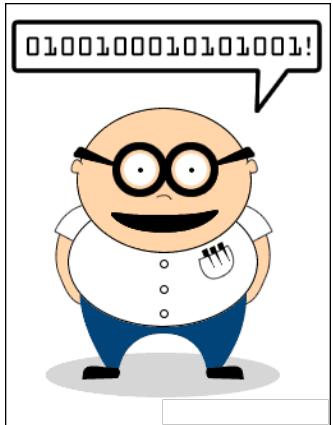
Analyzing biological data with informatics tools

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 $progrname = "ExtractID.pl";
my $input;
my $out;
my $id;

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

my $usage = <<END;
Usage: $progrname --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 );

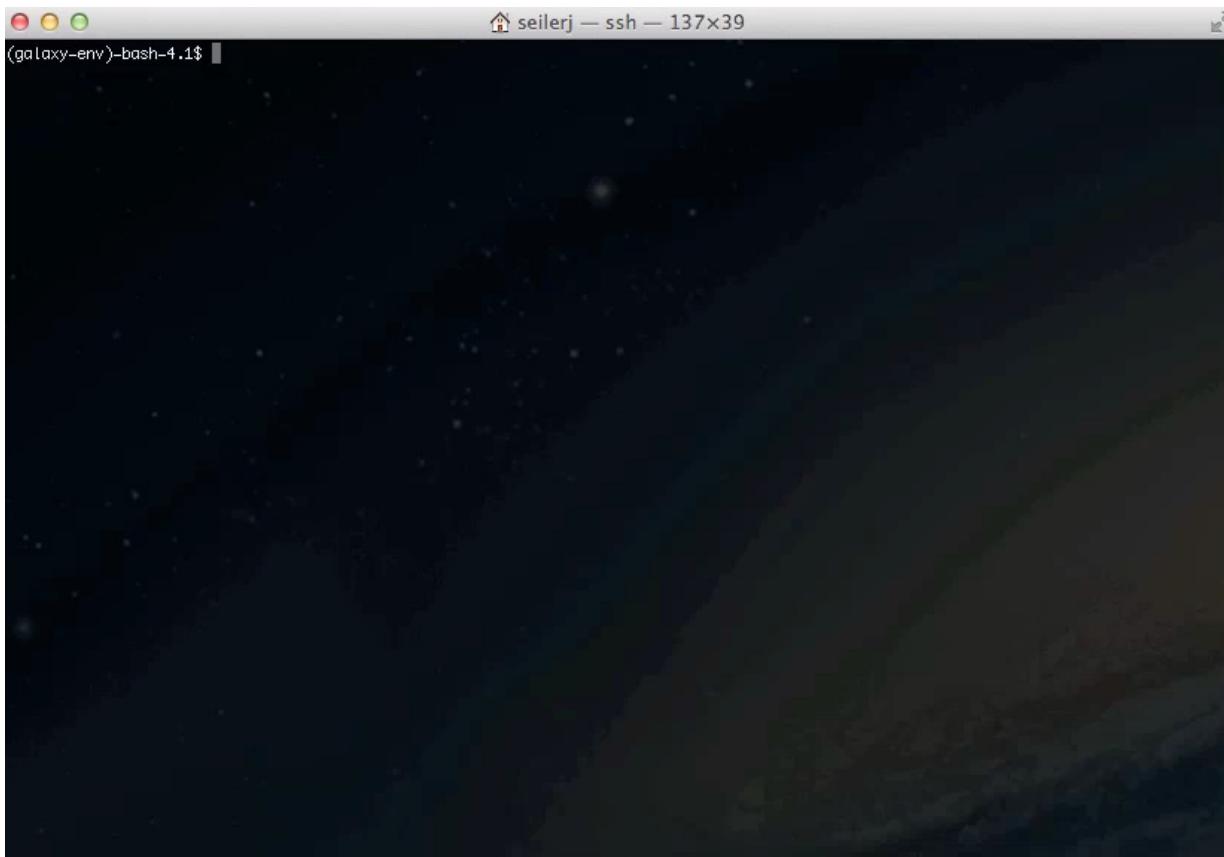
my %ids;
$out = ( defined $out ) ? $out : "results.txt";

## 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.
```

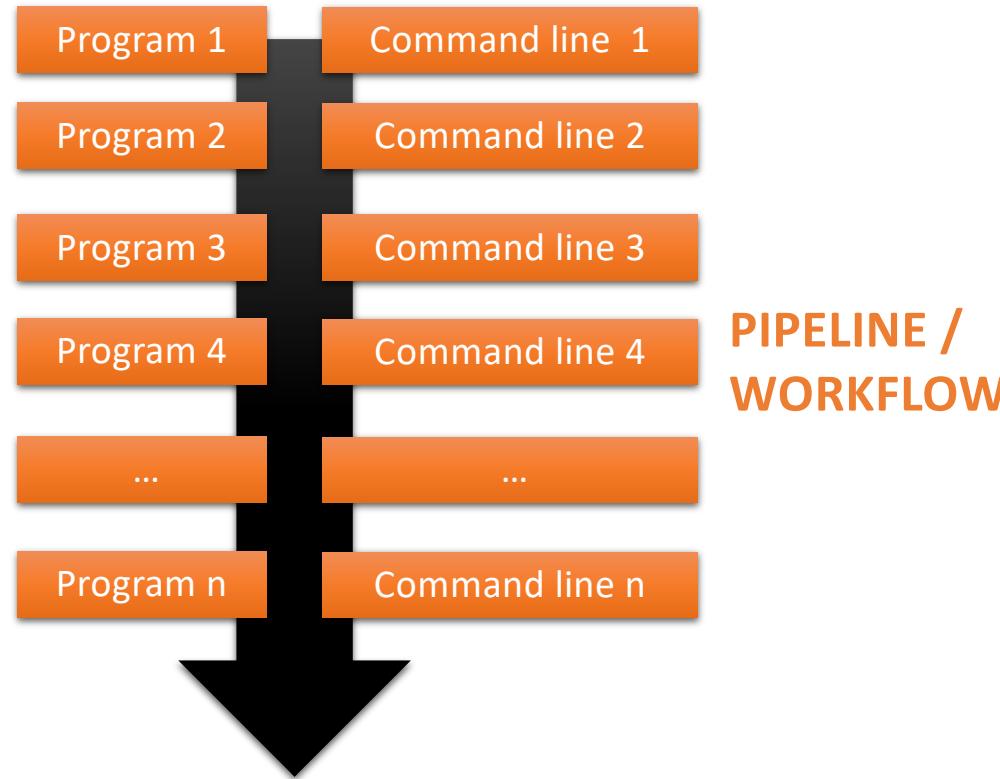
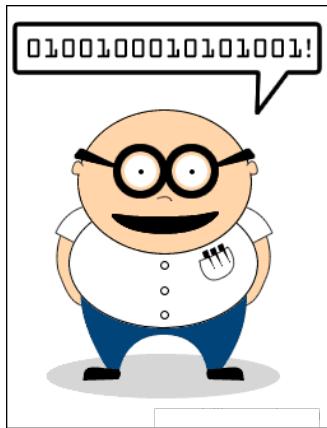
Command line

```
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



Galaxy ?





Galaxy PROJECT

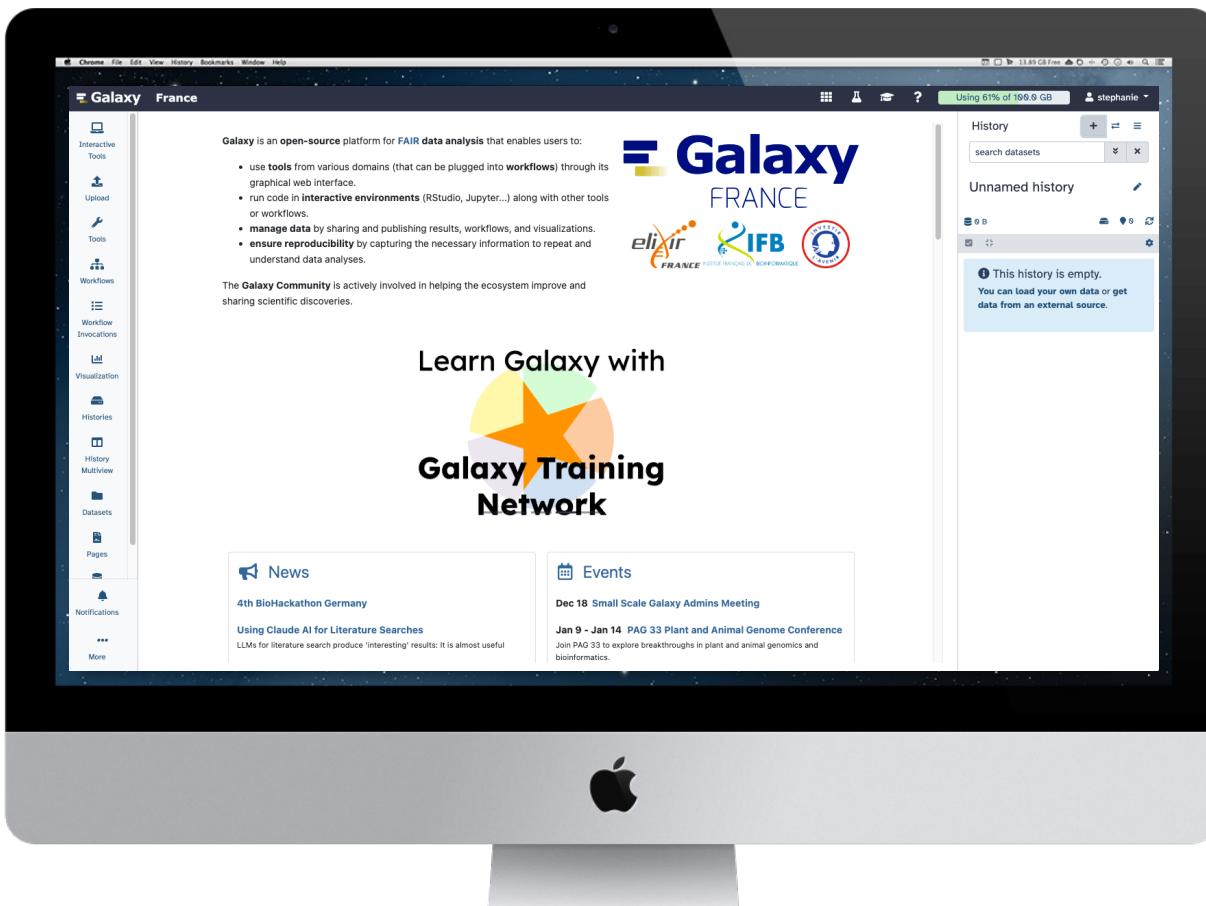
The Galaxy project

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

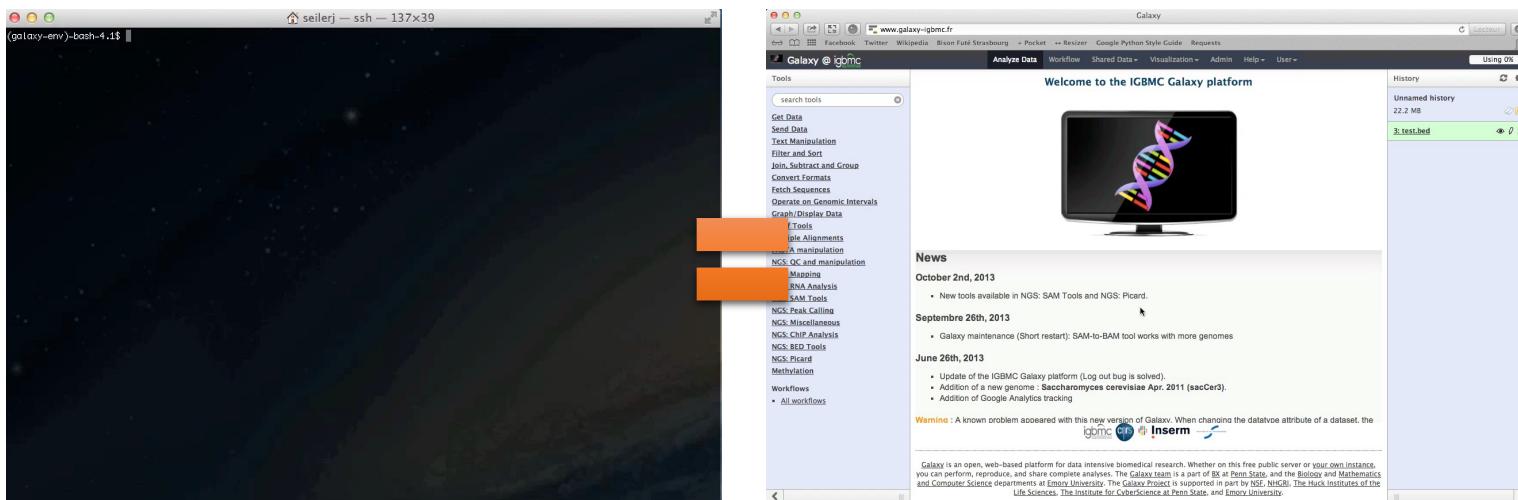


National
Human Genome
Research Institute

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

Galaxy public servers

- Galaxy Project's public server (<https://usegalaxy.org/>) (4)
- There are several public remote Galaxy instances worldwide (137)
 - Genomics Servers
 - Domain Servers
 - Tool Publishing Servers

Public Galaxy Servers list :
<https://galaxyproject.org/use/>
Last Update on: 2025, December 17th

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	Tlaas	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.

* Tlaas: Training Infrastructure as a Service

- <https://galaxyproject.org/use/>
 - 167 resources for using Galaxy (Last Update on: 2025, December 17th)

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 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 UI elements highlighted by orange arrows and labels:

- Top menu:** Located at the top right of the header.
- Tool panel:** A sidebar on the left containing links for Interactive Tools, Upload, Tools, Workflows, Workflow Invocations, Visualization, Histories, History Multiview, Datasets, Notifications, and More. The "Tools" link is highlighted.
- Shortcuts:** A label pointing to the top of the Tool panel sidebar.
- Data display and tools dialog window:** A central area featuring the Galaxy logo, a brief introduction, logos for elixir FRANCE, IFB, and INVESTISSEMENT D'AVENIR, and a "Collections!" section with a cartoon character.
- History panel:** A sidebar on the right listing a history of datasets, such as "bamCoverage on data 26", "WT_320_St.chr19.bam", and "R6_1_387_St.chr19.bam".

Left menu



Hands
On

Exercise 1

History

Screenshot of the Galaxy France interface showing the History panel.

The interface includes a sidebar with navigation links: Interactive Tools, Upload, Tools, Workflows, Workflow Invocations, Visualization, Histories, History Multiview, Datasets, Notifications, and More.

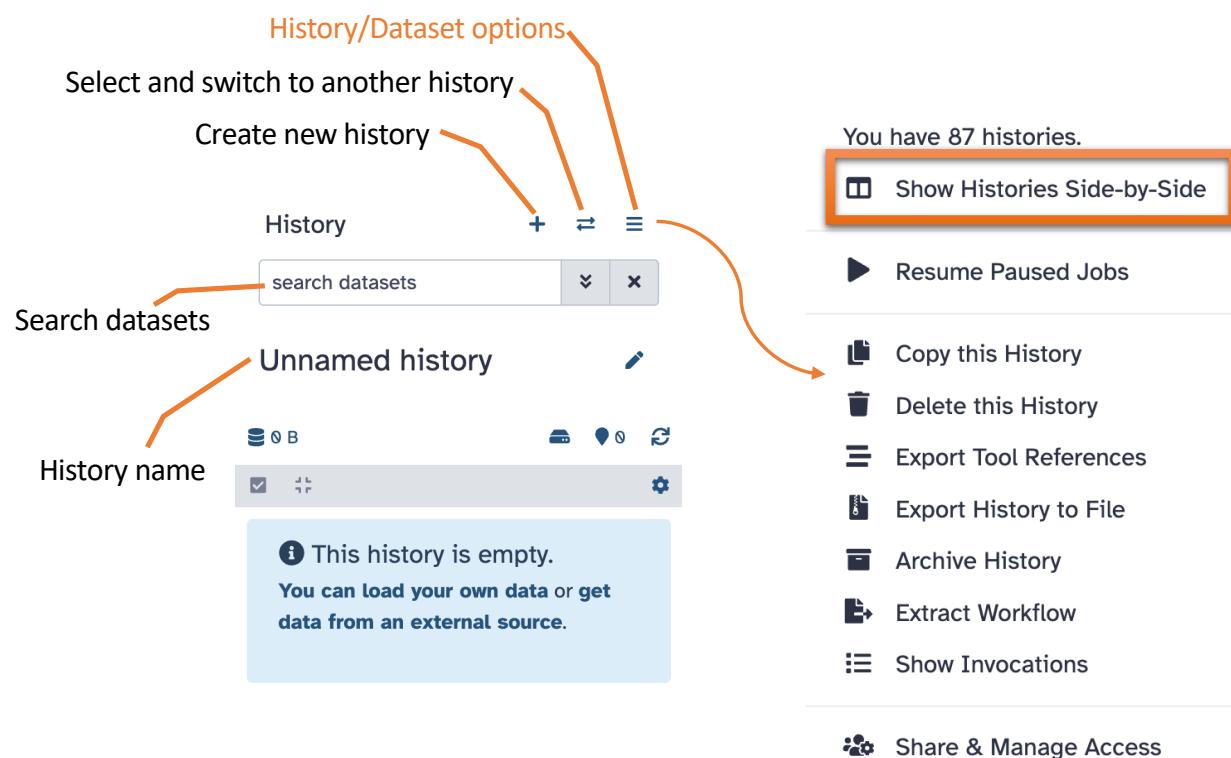
The main content area displays information about Galaxy, its open-source nature, and its role in FAIR data analysis. It also highlights the Galaxy Community's involvement in improving the ecosystem.

A central illustration features two cartoon characters: one holding a small dataset and another holding a large collection of datasets, with the text "Don't be like this guy. Use collections!" above them.

The History panel on the right shows an empty history named "Unnamed history". It includes a search bar, a summary of 0 B, and a message encouraging users to load their own data or get data from an external source.

An orange sidebar on the right provides a summary of the History panel: "History panel" and "Keep track of each job run".

History



View all histories

The screenshot shows the Galaxy web interface with the title "Galaxy France". The left sidebar contains links for Interactive Tools, Upload, Tools, Workflows, Workflow Invocations, Visualization, Histories (which is selected), Datasets, Notifications, and More. The main area is titled "History Multiview" and displays two pinned histories: "DNA-seq data analysis" and "Unnamed history".

DNA-seq data analysis (Last edited 11 months ago): 147 MB, 14 items, 15 revisions. Contains datasets: 29: VCFtoTab-delimited: on data 27, 28: SnpEff eff: on data 16 - HTML stats, 27: SnpEff eff: on data 16, 16: FreeBayes on data 15 and data 14 (variants), 15: CaptureDesign_chr4.bed, 14: MarkDuplicates on data 12: BAM, 13: MarkDuplicates on data 12: tabular.

Unnamed history (Last edited about 1 year ago): 39.5 MB, 3 items, 3 revisions. Contains datasets: 3: CaptureDesign_chr4.bed, 2: CRN-107_11-R2.fastq.gz, 1: CRN-107_11-R1.fastq.gz.

On the right, there is a "History" panel showing a list of 31 pinned histories:

- 29: VCFtoTab-delimited: on data 27
- 28: SnpEff eff: on data 16 - HTML stats
- 27: SnpEff eff: on data 16
- 16: FreeBayes on data 15 and data 14 (variants)
- 15: CaptureDesign_chr4.bed
- 14: MarkDuplicates on data 12: BAM
- 13: MarkDuplicates on data 12: tabular
- 12: Map with BWA-MEM on data 3 and data 2 (mapped reads in BAM format)
- 11: FastQC on data 3: RawData

At the bottom left of the main area, it says "Loaded 31 out of 87 histories".

**Hands
On**

Exercise 2

Import data into Galaxy

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

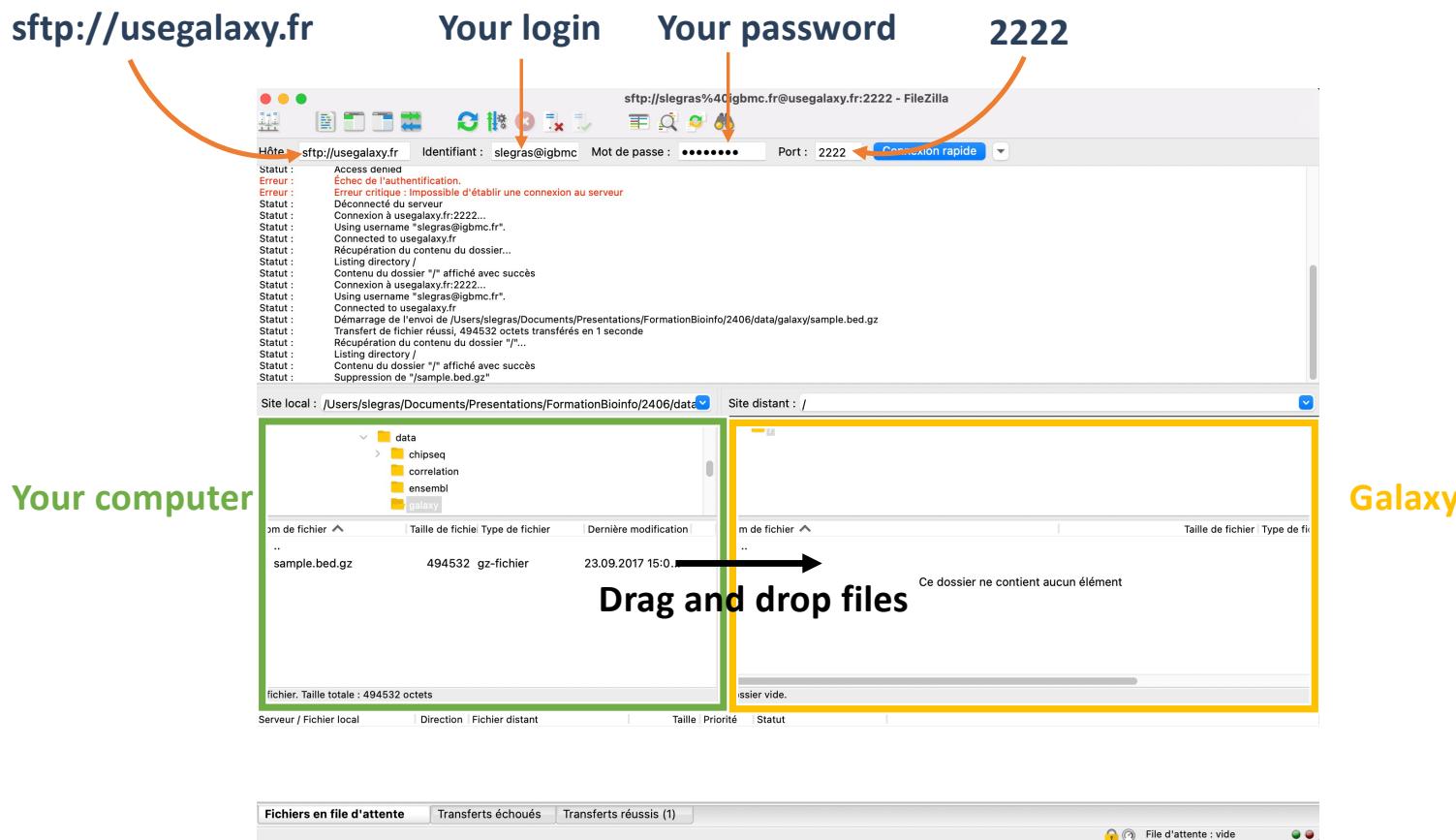
Import your own data to Galaxy

The screenshot shows the Galaxy web interface with the following elements:

- Left Sidebar:** A vertical sidebar with icons and labels for: Interactive Tools, Upload (highlighted with an orange arrow), Tools, Workflows, Workflow Invocations, Visualization, Histories, History Multiview, Datasets, Notifications, and More.
- Header:** "Galaxy" logo.
- Main Content Area:**
 - Section Title:** "Upload from Disk or Web to **DNA-seq data analysis**".
 - File Queue:** Shows one file: "CRN-107_11-R1.fastq" (18.5 MB). It includes fields for "Type (set all): Auto-detect", "Reference (set all): unspecified (?)", and controls for "Choose local file", "Choose remote files", "Paste/Fetch data", "Start", "Pause", "Reset", and "Close".
 - Annotations:** Labels with orange arrows pointing to specific fields:
 - "Name of the dataset" points to the file name "CRN-107_11-R1.fastq".
 - "Size of the dataset" points to the file size "18.5 MB".
 - "File format" points to the "Type (set all)" dropdown set to "Auto-detect".
 - "Genome" points to the "Reference (set all)" dropdown set to "unspecified (?)".

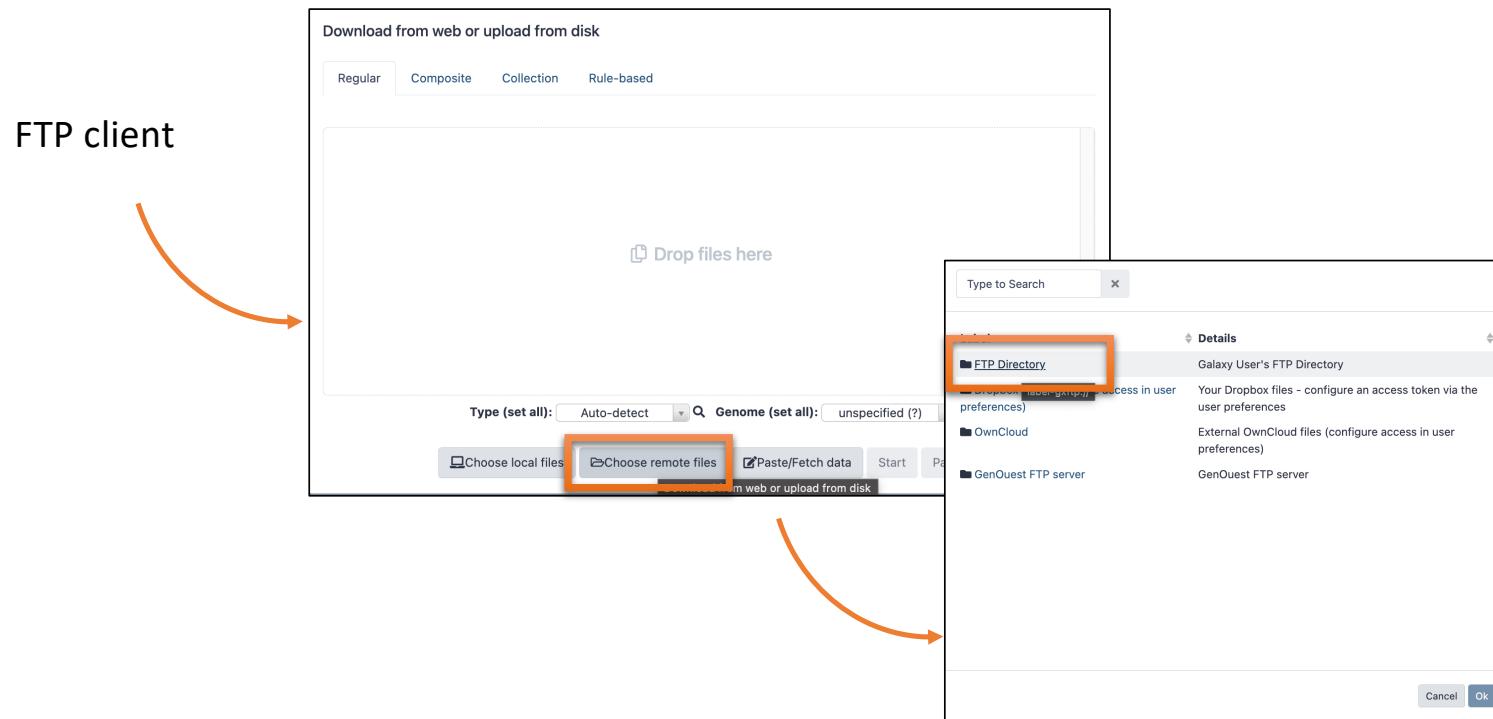
Import your own data to Galaxy (ftp - 1)

- For big files (>1GB) use FTP



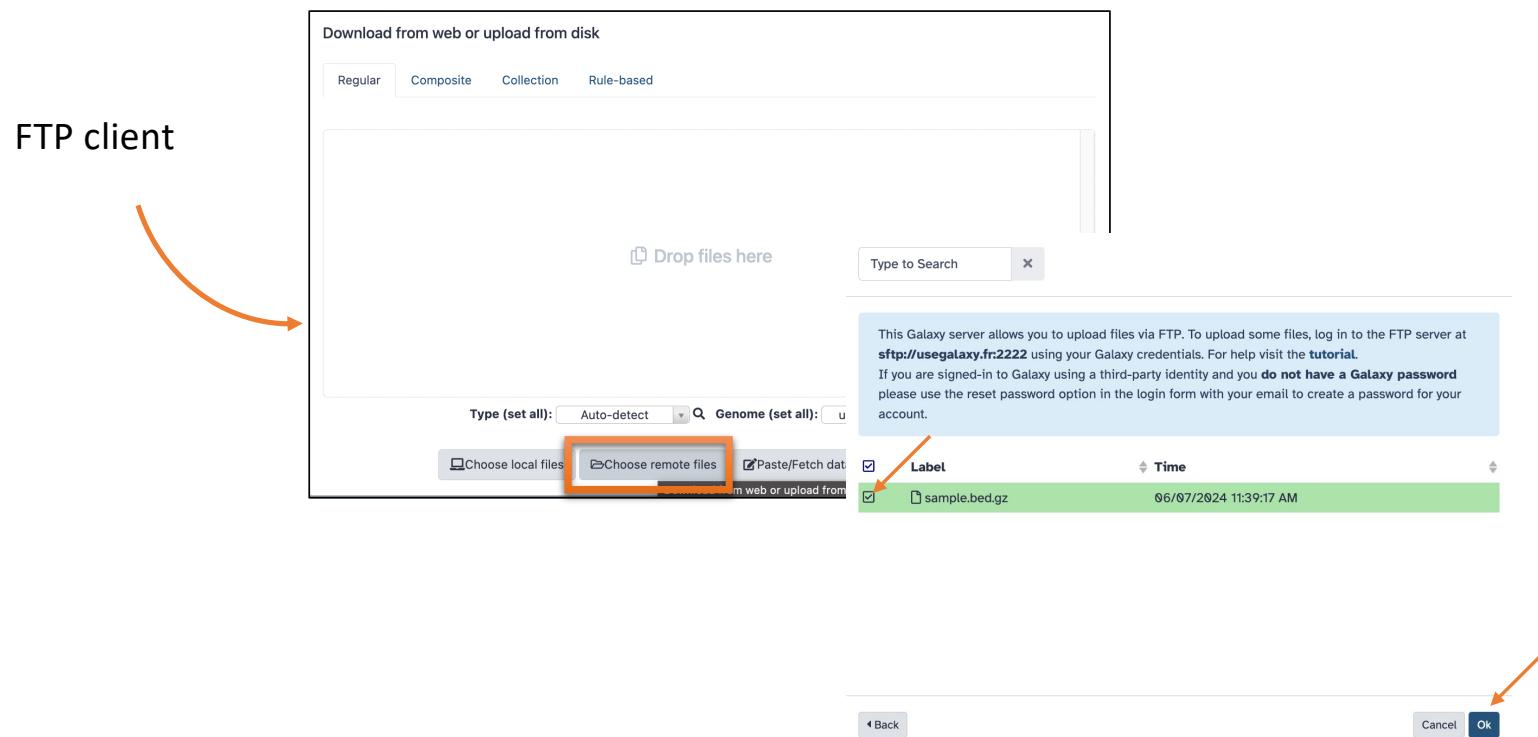
Import your own data to Galaxy (ftp - 2)

- For big files (>1GB) use FTP

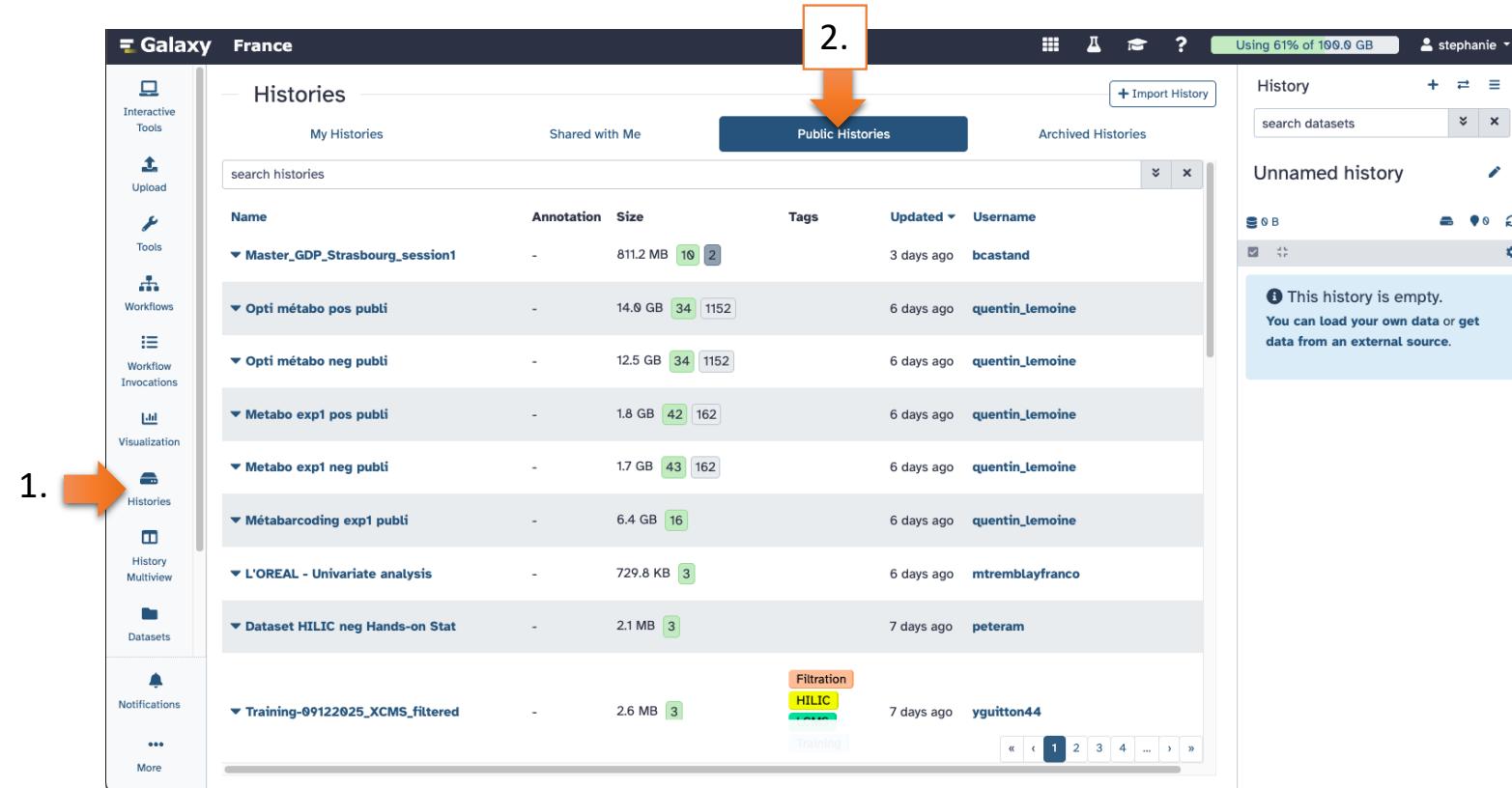


Import your own data to Galaxy (ftp - 3)

- For big files (>1GB) use FTP



Import shared data (Histories)



The screenshot shows the Galaxy web interface with the following details:

- Left Sidebar:** Includes icons for Interactive Tools, Upload, Tools, Workflows, Workflow Invocations, Visualization, Histories (highlighted with an orange arrow), History Multiview, Datasets, Notifications, and More.
- Top Bar:** Displays "Galaxy France", the user name "stephanie", and "Using 61% of 100.0 GB".
- Navigation:** Tabs include "My Histories", "Shared with Me", "Public Histories" (highlighted with an orange arrow), and "Archived Histories".
- Search:** A search bar for histories.
- Table:** A list of histories with columns for Name, Annotation, Size, Tags, Updated, and Username. Examples include "Master_GDP_Strasbourg_session1", "Opti métabo pos publi", and "Metabolomic exp1 neg publi".
- Right Panel:** Shows an "Unnamed history" section which is currently empty, with a message: "This history is empty. You can load your own data or get data from an external source."

Import public data

The screenshot shows the Galaxy web interface with the title "Galaxy France". The left sidebar contains links for Interactive Tools, Upload, Tools (which is selected), Workflows, Workflow Invocations, Visualization, Histories, and History Multiview. The main content area is titled "Tools" and includes a search bar. Below it is a section titled "Get Data" containing several options:

- NCBI Datasets Gene** download gene sequences and metadata
- NCBI Datasets Genomes** download genome sequence, annotation and metadata
- NCBI Accession Download** Download sequences from GenBank/RefSeq by accession through the NCBI ENTREZ API
- Download and Generate Pileup** Format from NCBI SRA
- Faster Download and Extract Reads in FASTQ format** from NCBI SRA
- Download and Extract Reads in FASTQ format** from NCBI SRA
- Download and Extract Reads in BAM format** from NCBI SRA

Browse and import external data from public databases

The screenshot shows the UCSC Main Table Browser interface. At the top, there are tabs for Genomes, Genome Browser, Tools, Mirrors, Downloads, My Data, Projects, Help, and About Us. The main content area is titled "Table Browser" and includes the following controls:

- Select dataset**: clade: Mammal, genome: Mouse, assembly: Dec. 2011 (GRCm38/mm10)
- Define region of interest**: region: genome (radio button selected), position: chr12:56,694,976-56,714,605, lookup, define regions
- Optional: Subset, combine, compare with another track**: filter: create, intersection: create
- Retrieve and display data**: output format: BED - browser extensible data, Send output to: Galaxy, GREAT, output filename: mm39_rmsk.fasta.gz, file type returned: plain text, gzip compressed
- 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](#).

UCSC Main Table Browser

**Hands
On**

Exercise 3

Datasets/Jobs in the History

Grey: the job is waiting to run

1: sample.bed.gz 

Orange: the job is running

1: sample.bed.gz 

Green: the job is successfully done

1: sample.bed.gz 

Red: the job encountered a problem

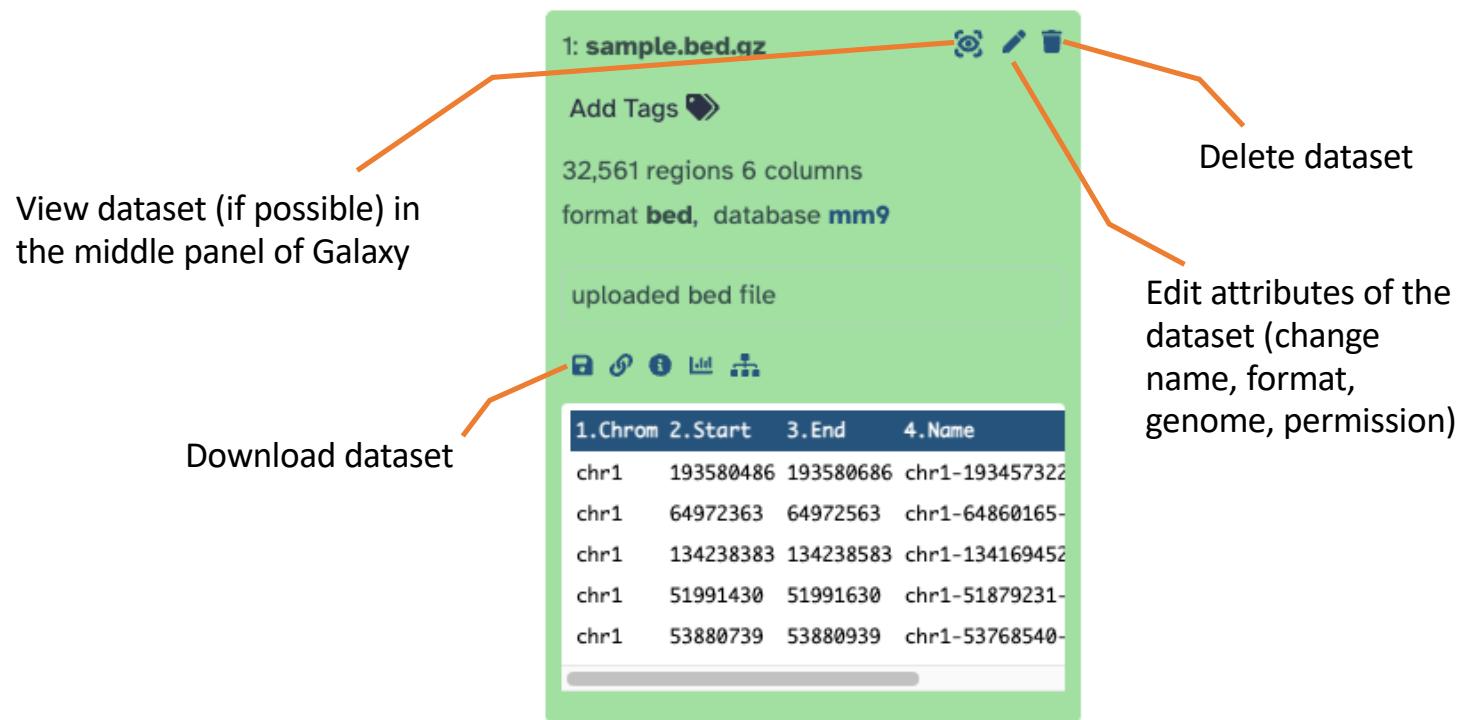
98: MultiQC on data 96, d 
ata 95, and others: Stats

Datasets/Jobs in the History

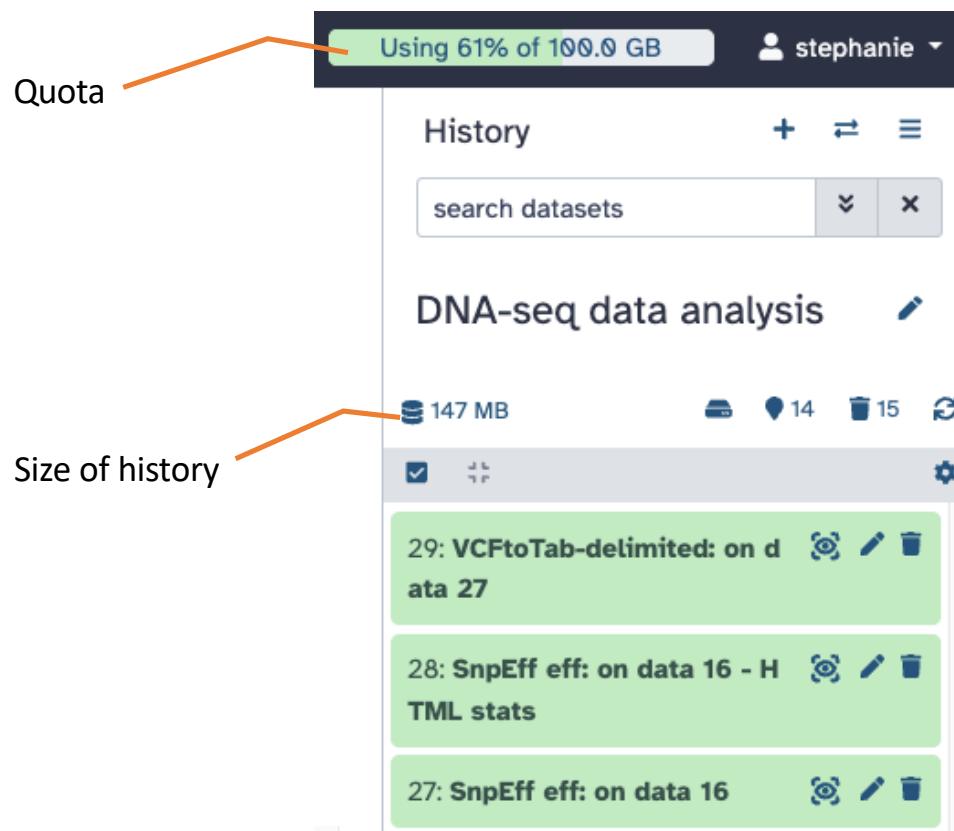
The screenshot shows a dataset card for a file named "sample.bed.gz". The card has a green header and footer and a white central area. In the top right corner of the header are three icons: a magnifying glass, a pencil, and a trash can. Below these are buttons for "Add Tags" and a small icon. The main text area displays the file's characteristics: "32,561 regions 6 columns", "format bed, database mm9", and "uploaded bed file". At the bottom, there is a table preview with columns labeled "1.Chrom", "2.Start", "3.End", and "4.Name". The first five rows of the table are shown, all corresponding to "chr1". Orange arrows point from the text "Number of lines in the file or size of the file" to the line count in the header, from "Format" to the "format bed" text, and from "Genome" to the "mm9" database name. Another orange arrow points from the text "If the dataset is a text file, the first lines of the file are displayed" to the table preview.

1.Chrom	2.Start	3.End	4.Name
chr1	193580486	193580686	chr1-193457322
chr1	64972363	64972563	chr1-64860165-
chr1	134238383	134238583	chr1-134169452
chr1	51991430	51991630	chr1-51879231-
chr1	53880739	53880939	chr1-53768540-

Datasets/Jobs in the History



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 sidebar menu is displayed with various icons and labels. An orange arrow points from the text "Show/hide Tool panel" to the "Tools" icon in the sidebar. Another orange arrow points from the text "Tool panel" to the expanded "Tools" panel on the right, which is highlighted with a red border. The "Tools" panel contains a search bar and a list of tool categories: General Text Tools, Genomic File Manipulation, and Community. The "General Text Tools" category is currently selected. The main content area features the Galaxy France logo and a list of features. To the right of the main content is a "History" panel, which is currently empty.

Show/hide
Tool panel

Tool panel

Galaxy France

Using 61% of 100.0 GB stephanie

History + ⌂ ⌂

search datasets

Unnamed history

0 B

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

Galaxy FRANCE

elixir FRANCE IFB INSTITUT FRANÇAIS DE BIOMÉTIQUE INVESTISSEMENT D'ETAT

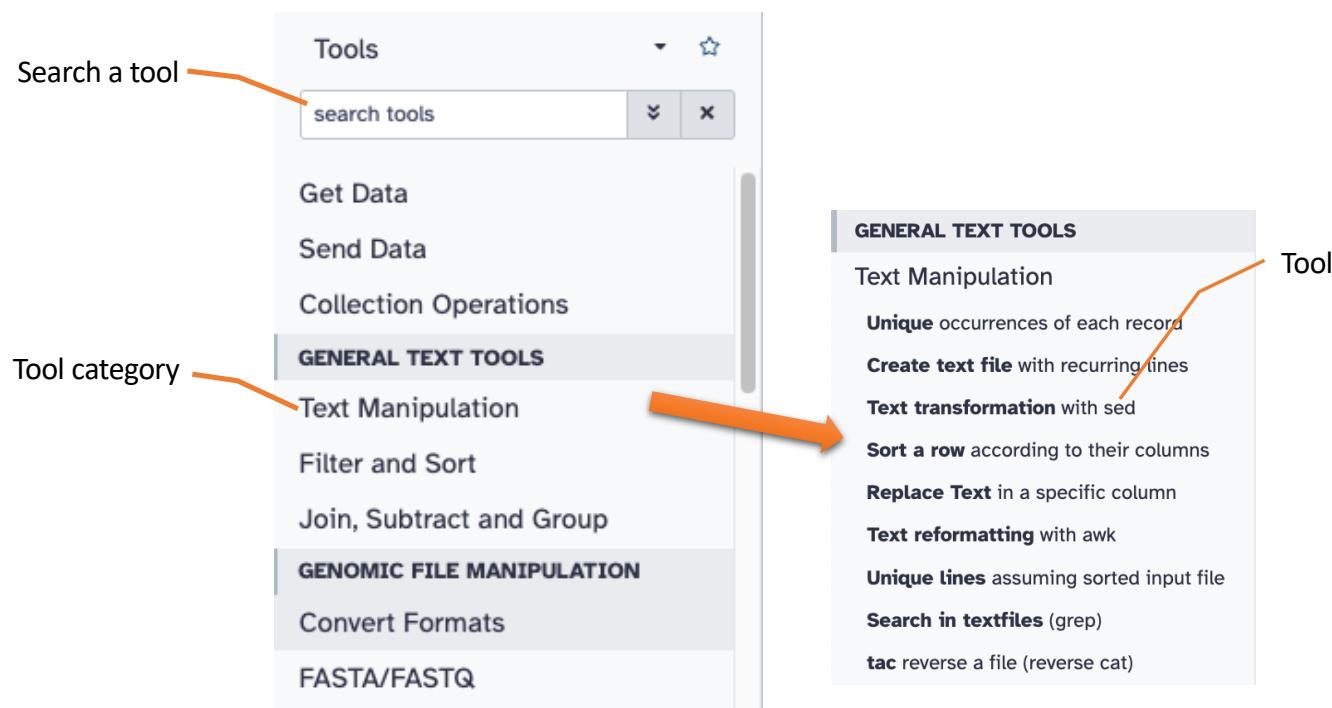
Galaxy is an open-source platform for FAIR data analysis that enables users to:

- use tools from various domains (that can be plugged into workflows) through its graphical web interface.
- run code in interactive environments (RStudio, Jupyter...) along with other tools or workflows.
- manage data by sharing and publishing results, workflows, and visualizations.
- ensure reproducibility by capturing the necessary information to repeat and understand data analyses.

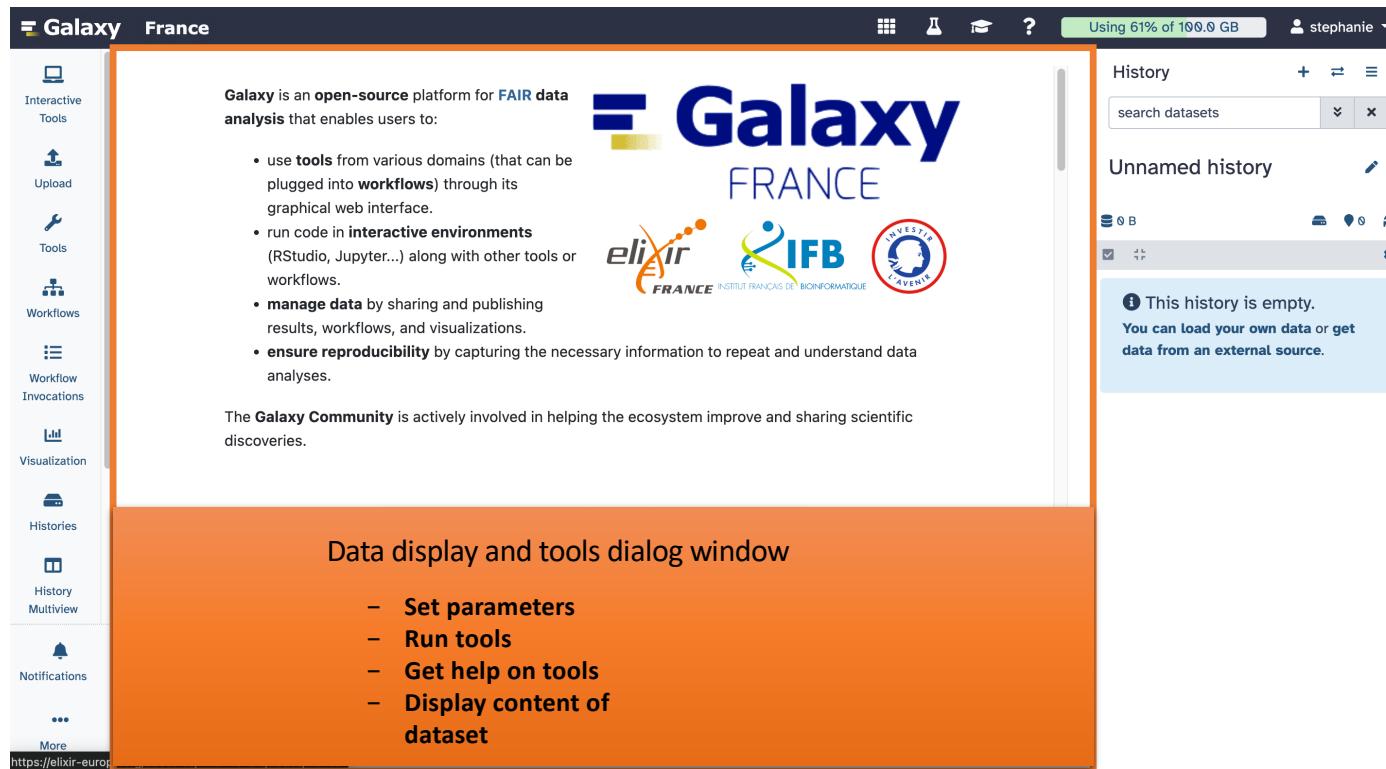
The Galaxy Community is actively involved in helping the ecosystem improve and sharing scientific discoveries.

JXTX James

Tool Panel / Run analyses



Tools dialog window



**Hands
On**

Exercise 5

**Hands
On**

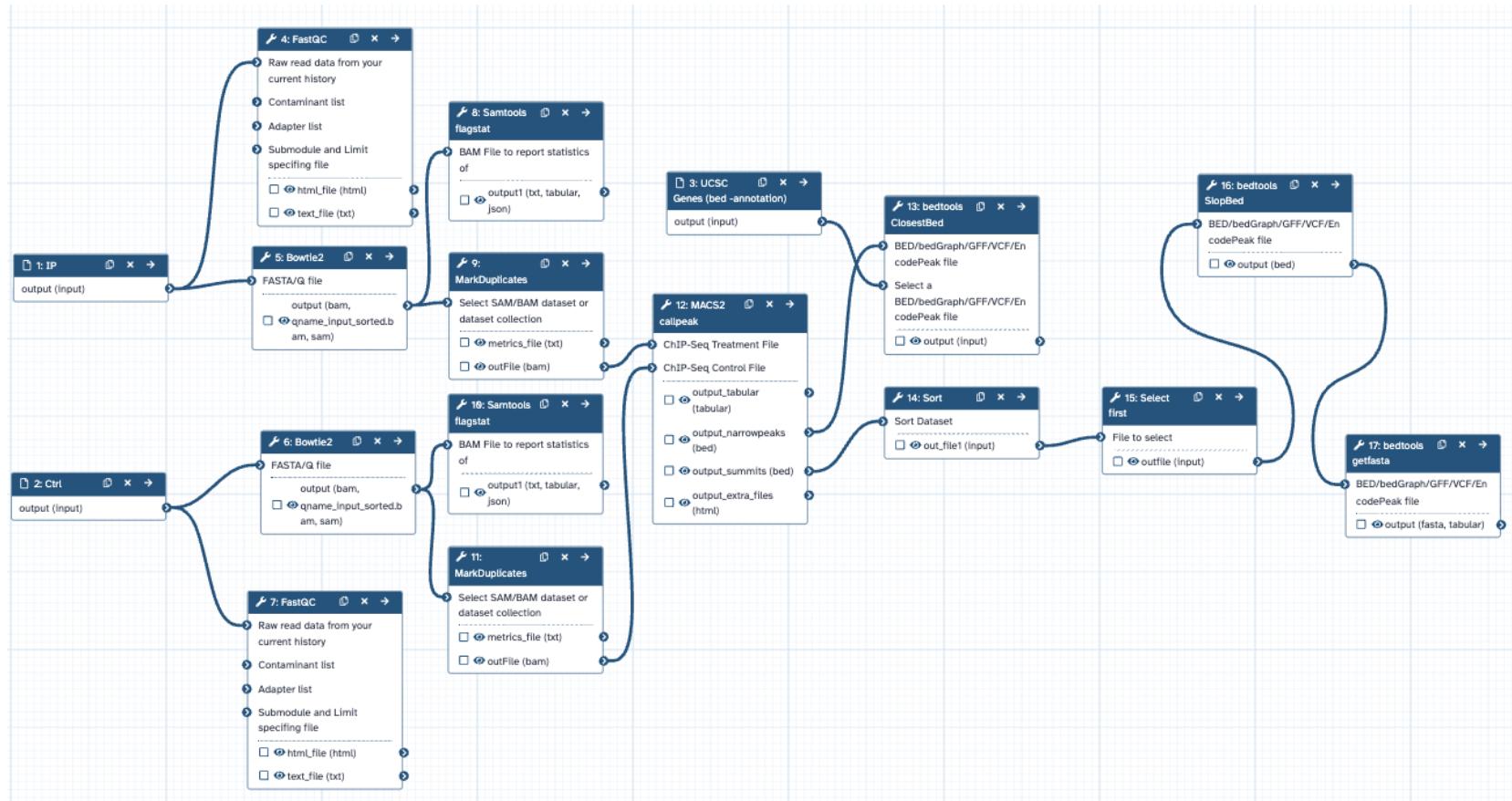
Exercise 6

Workflow

**Hands
On**

Exercise 7

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 interface for a user named stephanie. The left sidebar has a 'Workflows' icon highlighted with an orange box. The main area is titled 'Workflows' and contains a list of workflow entries. Each entry includes a checkbox, the workflow name, a timestamp, and a set of actions (Edit, Run, Download, etc.). A red arrow points from the 'Create workflows' button in the center-right area to the 'Create' button in the top right corner of the header bar.

Galaxy France

Workflows

My workflows Workflows shared with me Public workflow

+ Create Import

History search datasets

Unnamed history

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

Workflows

Interactive Tools Upload Tools Workflow Invocations Visualization Histories Multiview Datasets Pages Notifications More

Workflows

My workflows

Search my workflows by query or use the advanced filtering options

Select all Sort by: Name Update time Filter: Show deleted Show bookmarked Display:

BED Ensembl to BED UCSC (workflow runs: 6) RNAseq data analysis (Workflow runs: 1)

DNA-seq data analysis (11 Décembre 2024) (Workflow runs: 1) DNA-seq data analysis (Workflow runs: 1)

DNA-seq data analysis (Workflow runs: 1) ChIP-seq data analysis workflow (11 juin 2024) (Workflow runs: 1)

DNA-seq workflow (Workflow runs: 1) ChIP-seq data analysis workflow (2 juin 2023) (Workflow runs: 1)

BED Ensembl to BED UCSC (Workflow runs: 2) Run "Prepare RNA-seq data for seqMINER" (Workflow runs: 1)

Add Tags 6 months ago Add Tags 6 months ago

Add Tags about 1 year ago Add Tags about 1 year ago

Add Tags about 1 year ago Add Tags about 1 year ago

Add Tags about 2 years ago Add Tags over 2 years ago

Add Tags

Create workflows

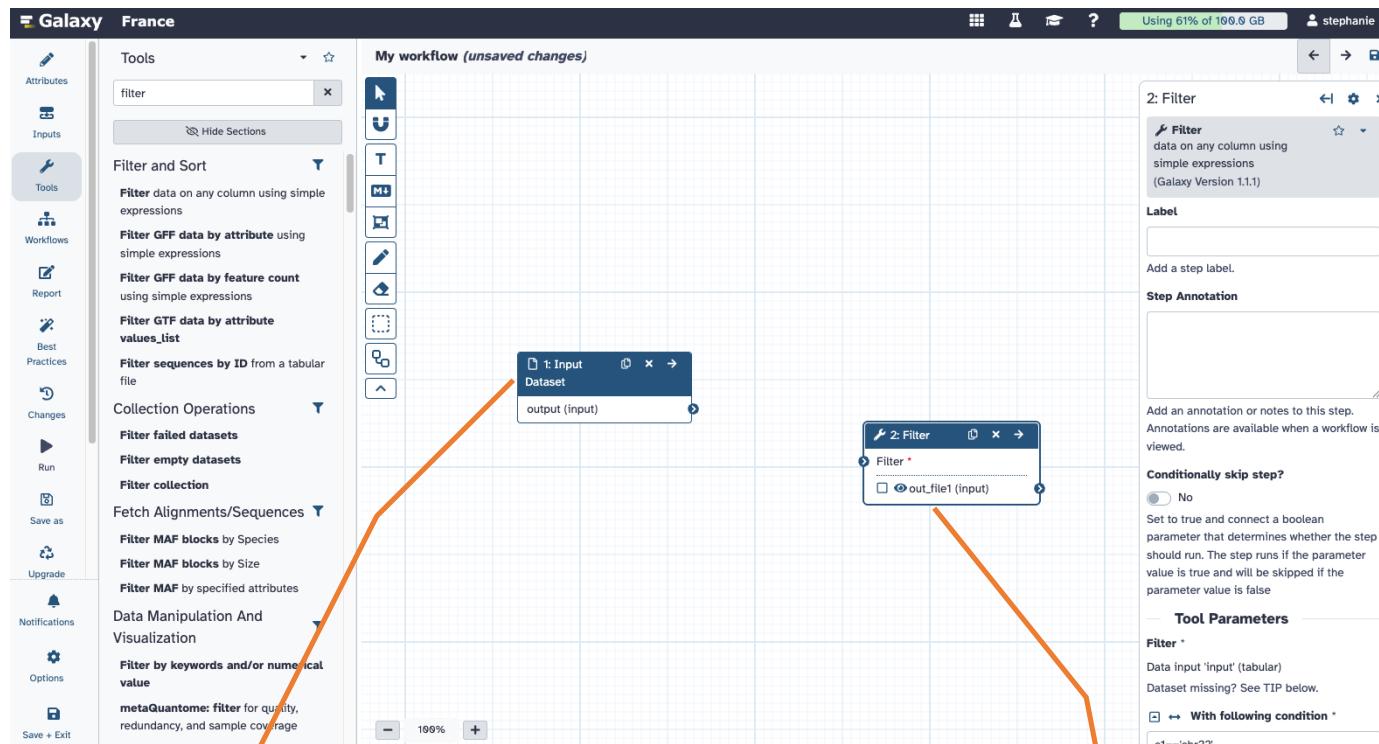
Workflow creation

Add tools or
input datasets
to the workflow

The screenshot shows the Galaxy software interface for creating a workflow. On the left, there's a sidebar with various options: Attributes (selected), Inputs, Tools, Workflows, Report, Best Practices, Changes, Run, Save as, Upgrade, Notifications, Options, and Save + Exit. The main area is titled "Unnamed Workflow". It has a "Name" field containing "Unnamed Workflow" and a "Short Description" field with placeholder text about visibility and length. Below these are sections for "License" (set to "Do not specify a license"), "Creator" (with a link to add a new creator), "Digital Object Identifier (DOI)" (with a link to add a new DOI), "Tags" (with a "Add Tags" button), and "Readme" (with a "Show Readme" button and a detailed description of its purpose). A large, light-blue grid workspace occupies the right side of the screen. An orange arrow points from the text "Add tools or input datasets to the workflow" towards the central workspace area.

Give a name to the workflow

Workflow creation

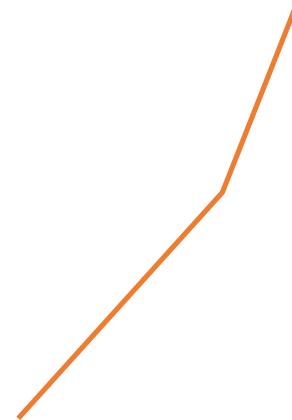


Input dataset.

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

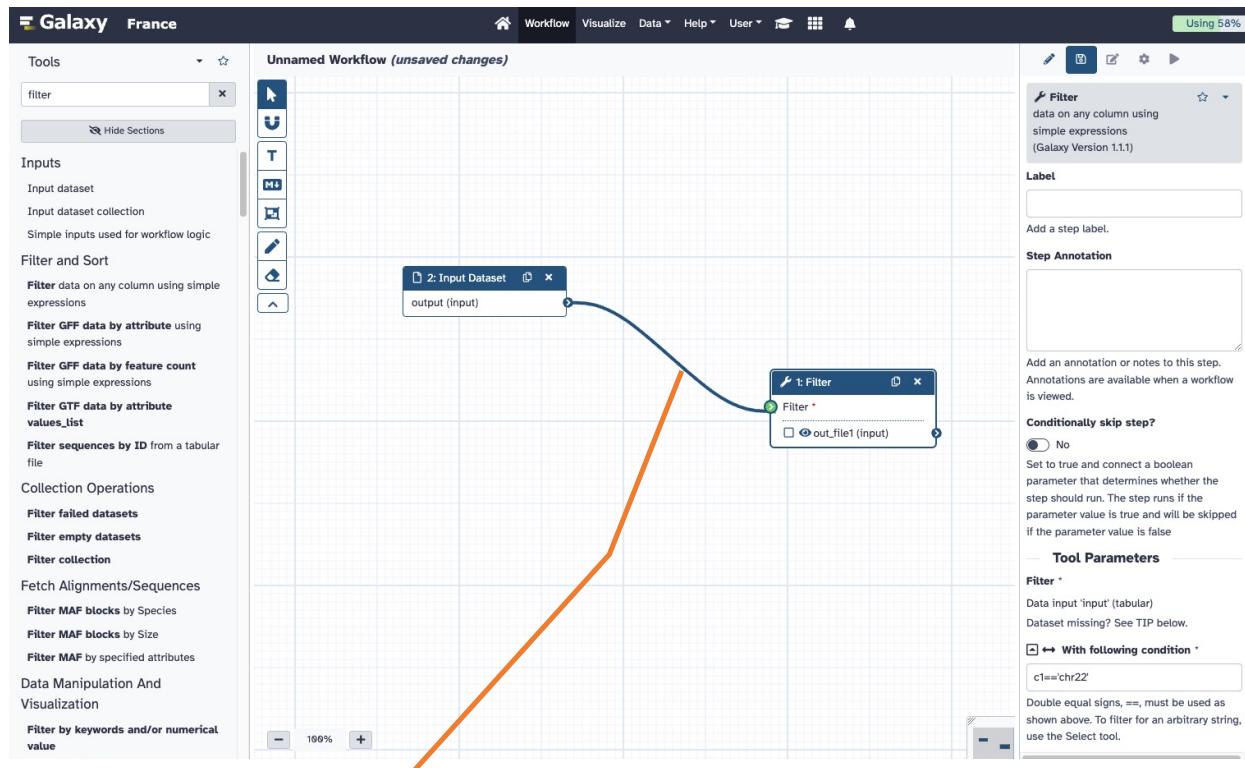
Tool to be run

Workflow creation



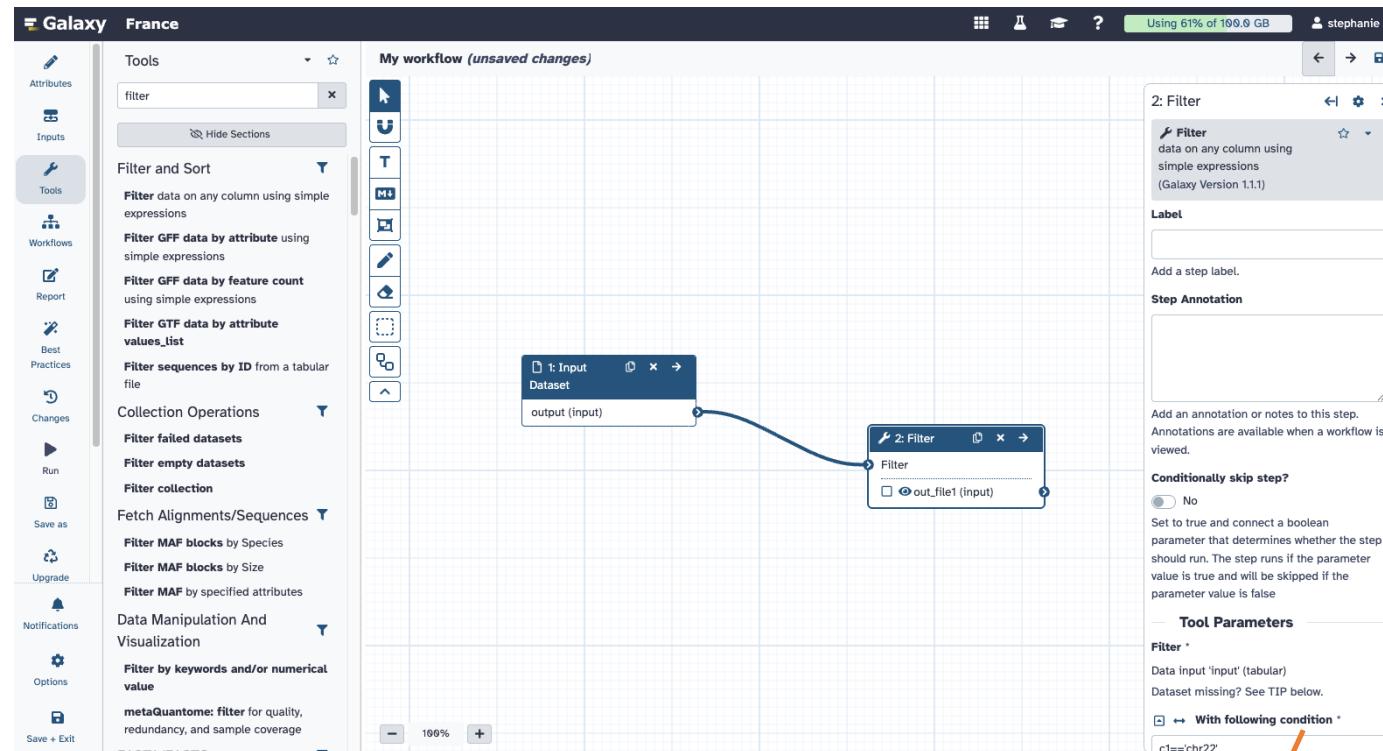
If two steps can be linked together,
the link between the two boxes is
green

Workflow creation



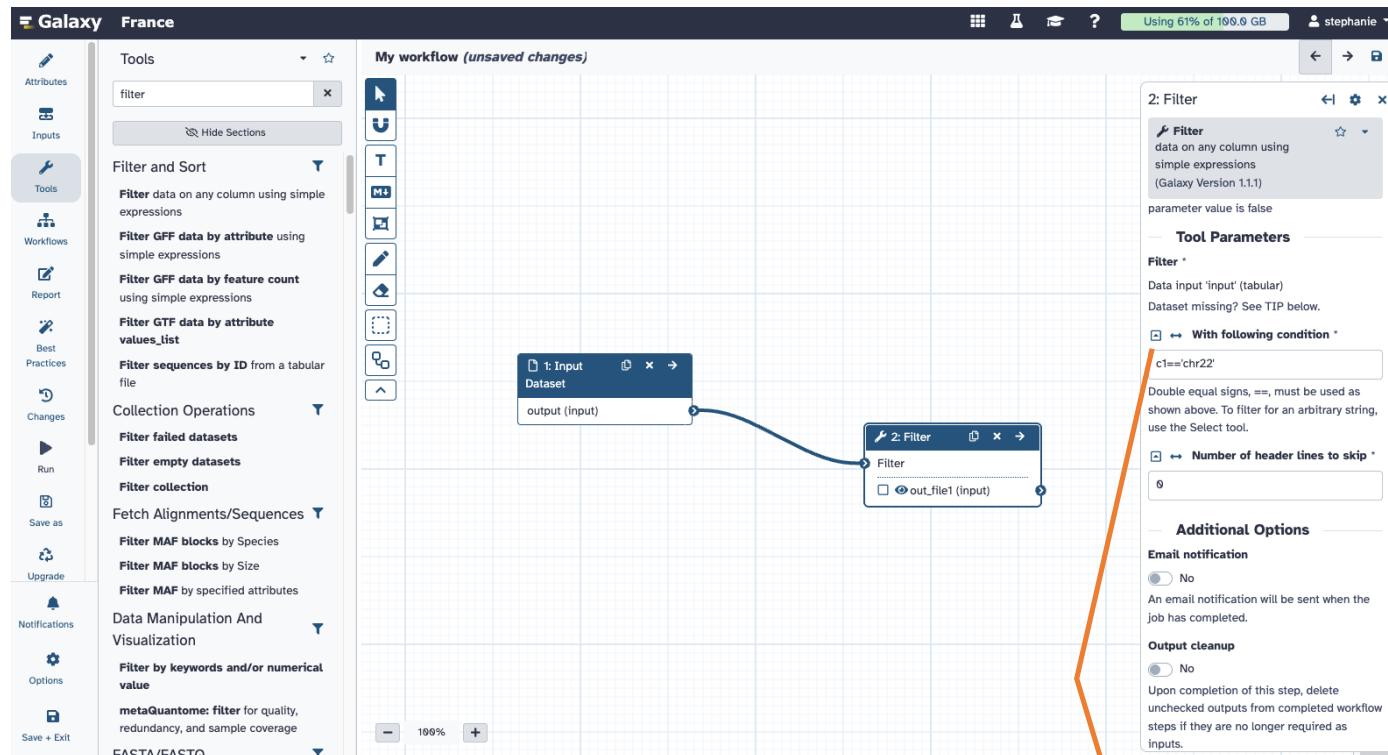
If two steps can be linked together,
the link between the two boxes is
green

Workflow creation



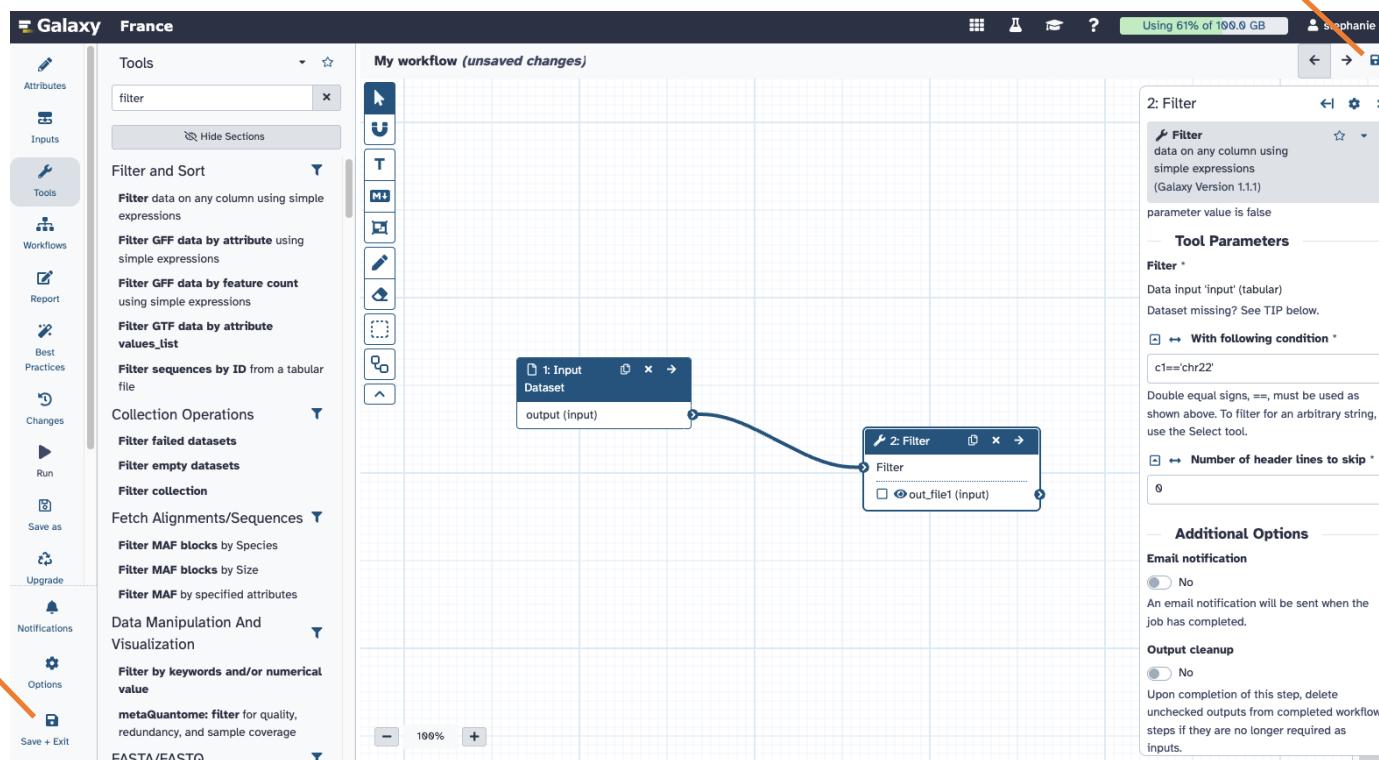
Pre-configure tool parameters and
configure parameters to be set at
run time

Workflow creation



Click to get the parameter to be set at runtime

Workflow creation



Save and exit

Save

Run workflow

The screenshot shows the Galaxy web interface with the following steps highlighted:

- Step 1:** A large orange arrow points to the "Workflows" icon in the left sidebar.
- Step 2:** An orange box highlights the play button icon (play icon with a playhead) in the workflow list, indicating the action to run the workflow.

Workflow Details:

- Workflow Name:** BED Ensembl to BED UCSC (Version: 6)
- Last Edit:** edited 6 months ago
- Inputs:** 1: hg38_ens105.bed
- Outputs:** (not visible)
- Run Workflow:** A blue button labeled "Run Workflow" is visible at the top right of the workflow card.

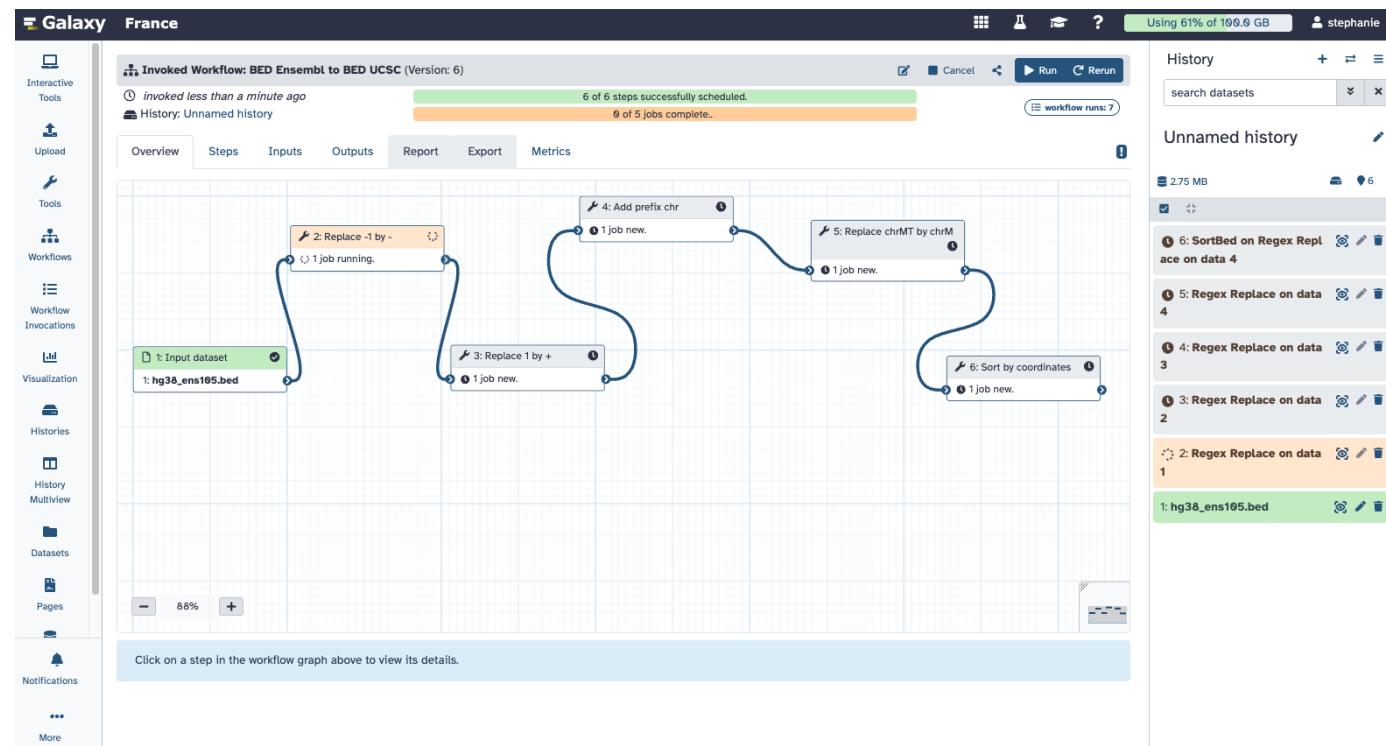
History Panel:

- History:** search datasets
- Unnamed history:** This history is empty. You can load your own data or get data from an external source.
- Dataset:** 1: hg38_ens105.bed (green status)

Annotations:

- Step 2. Run workflow:** An orange box with an arrow pointing to the "Run Workflow" button.
- Step 1. Set input file(s). Found in current history!** An orange box with an arrow pointing to the input field containing "1: hg38_ens105.bed".

Workflow processing



**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