

Introduction to Galaxy

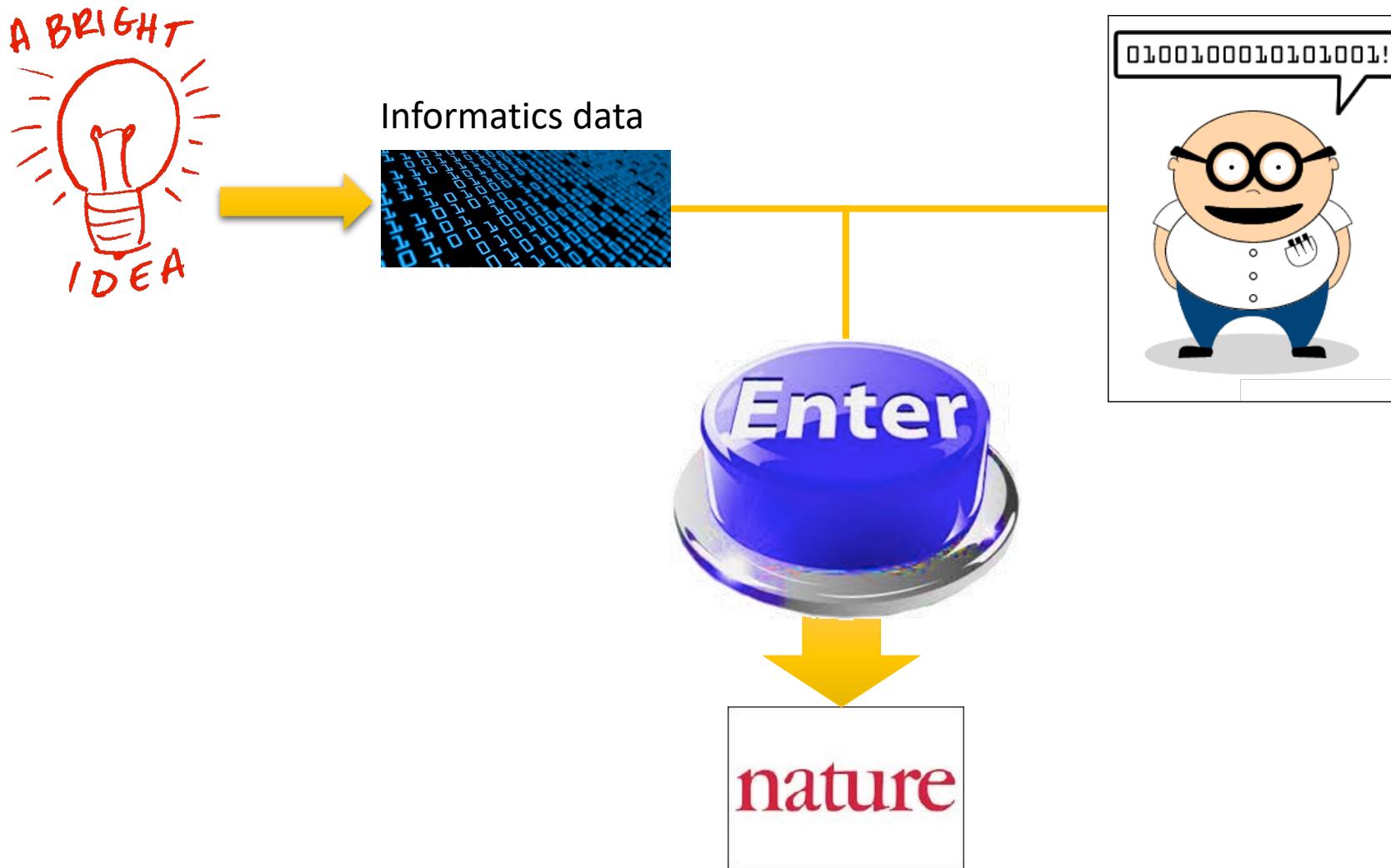
Stéphanie Le Gras
slegras@igbmc.fr

Guidelines

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

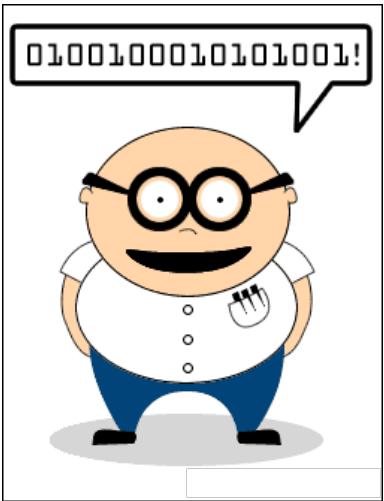
Analyzing biological data with informatics tools

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 );

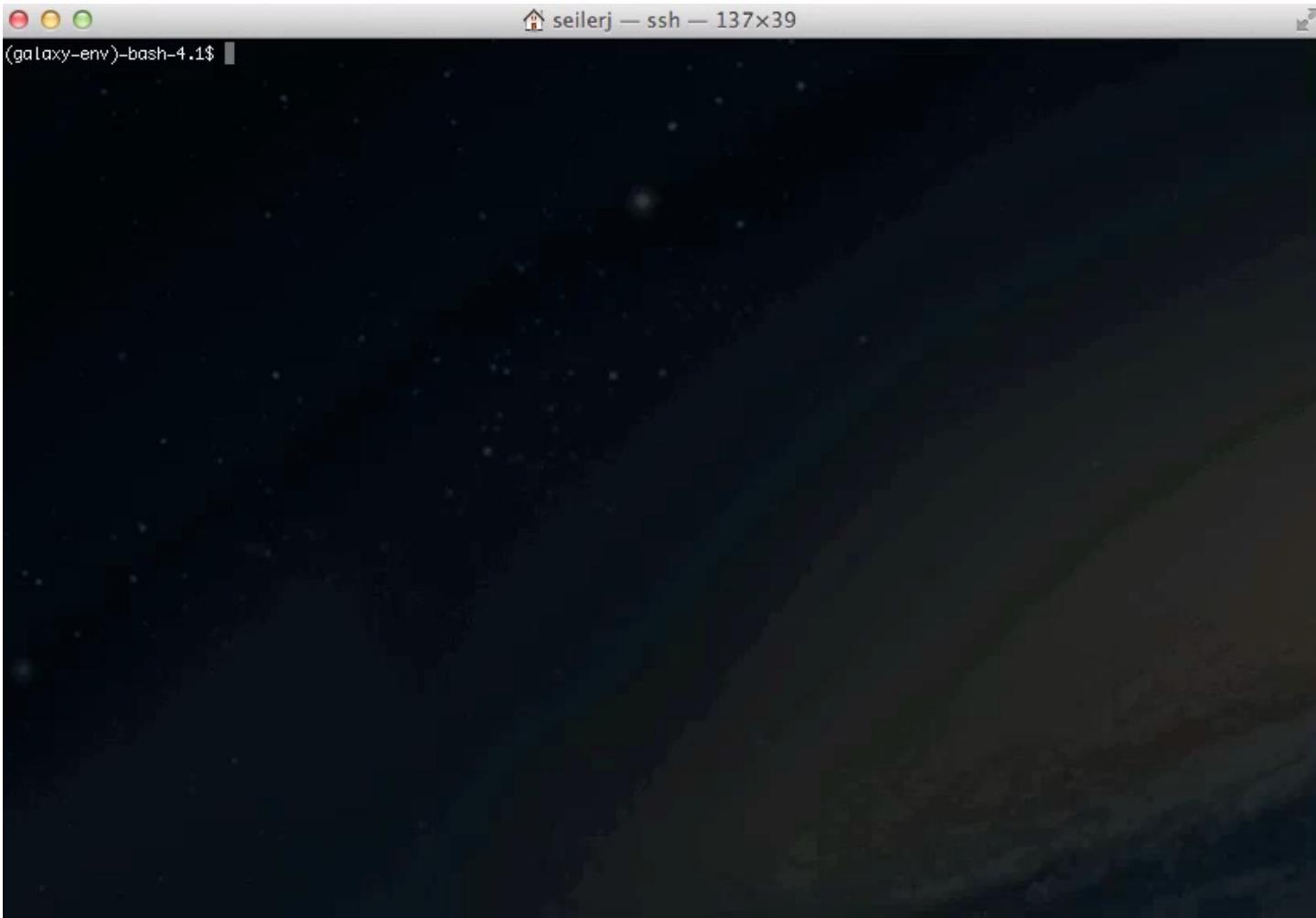
my %ids;
$out = ( defined $out ) ? $out : "results.txt";

## 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.
```

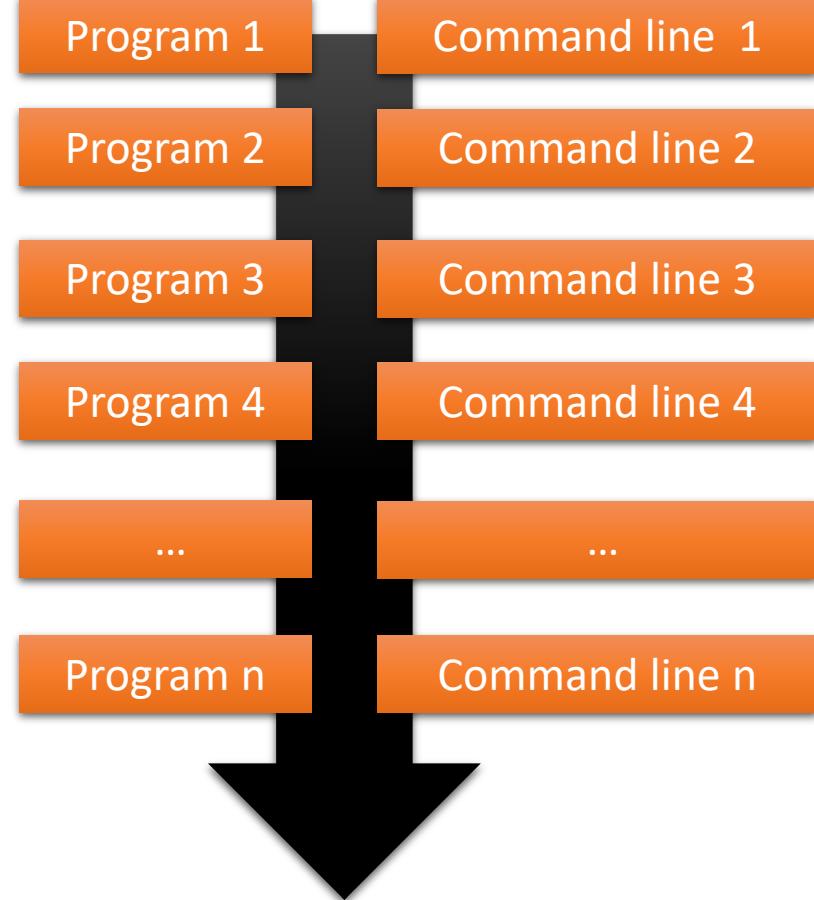
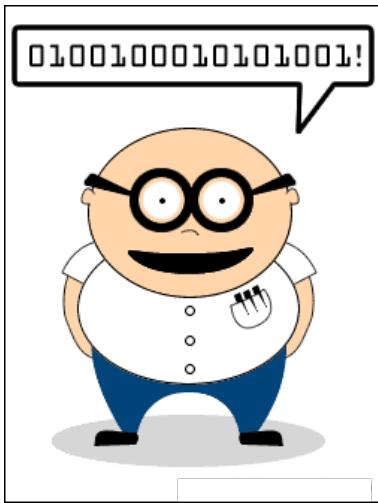
Command line

```
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 ?

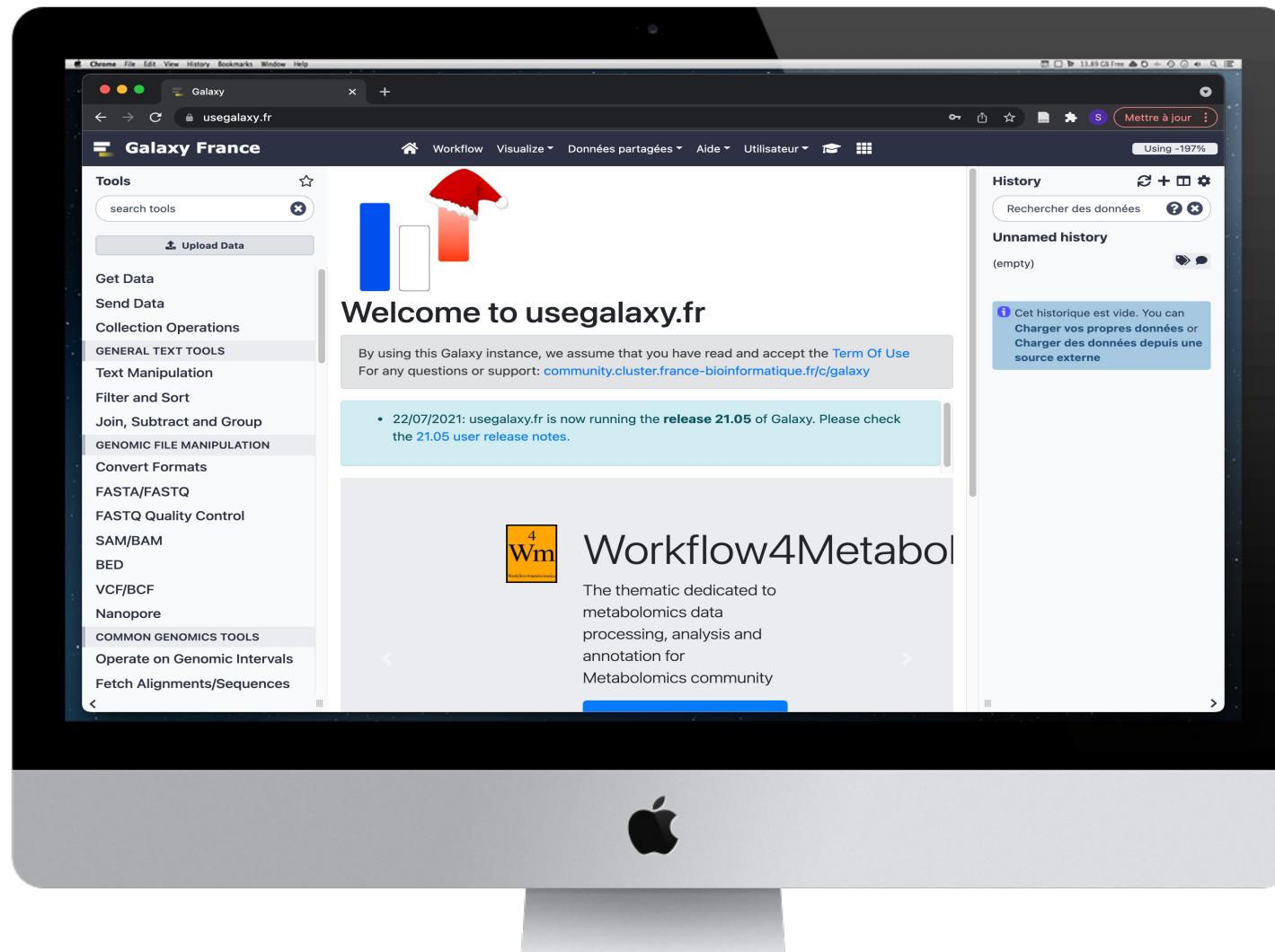
Galaxy is a **computing platform** that enables people to run **complex bioinformatics tools** on a **compute cluster** through a **simple web interface**.



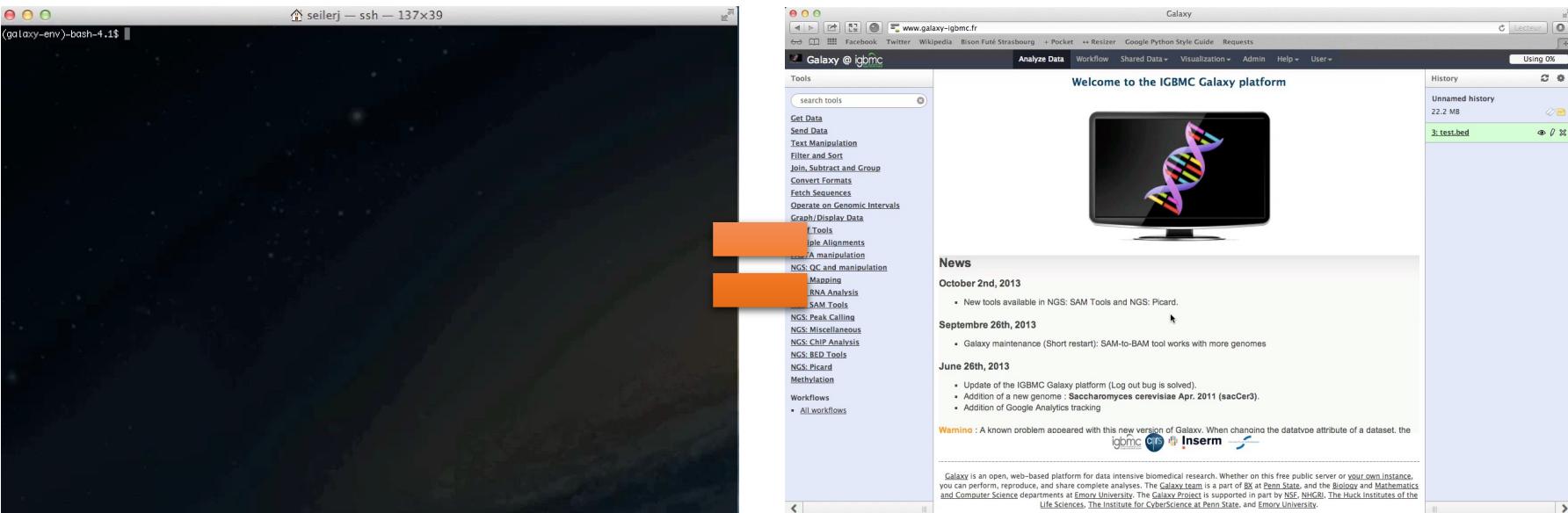
EMORY
UNIVERSITY



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 (160)
 - Genomics Servers
 - Domain Servers
 - Tool Publishing Servers

Public Galaxy Servers list :
<https://galaxyproject.org/use/>
Last Update on: 2020, December 16th

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 (including intermediate) total > 250GB	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

* TlaaS: Training Infrastructure as a Service

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

Galaxy public servers

Your research institute



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

Download and
upload of the data

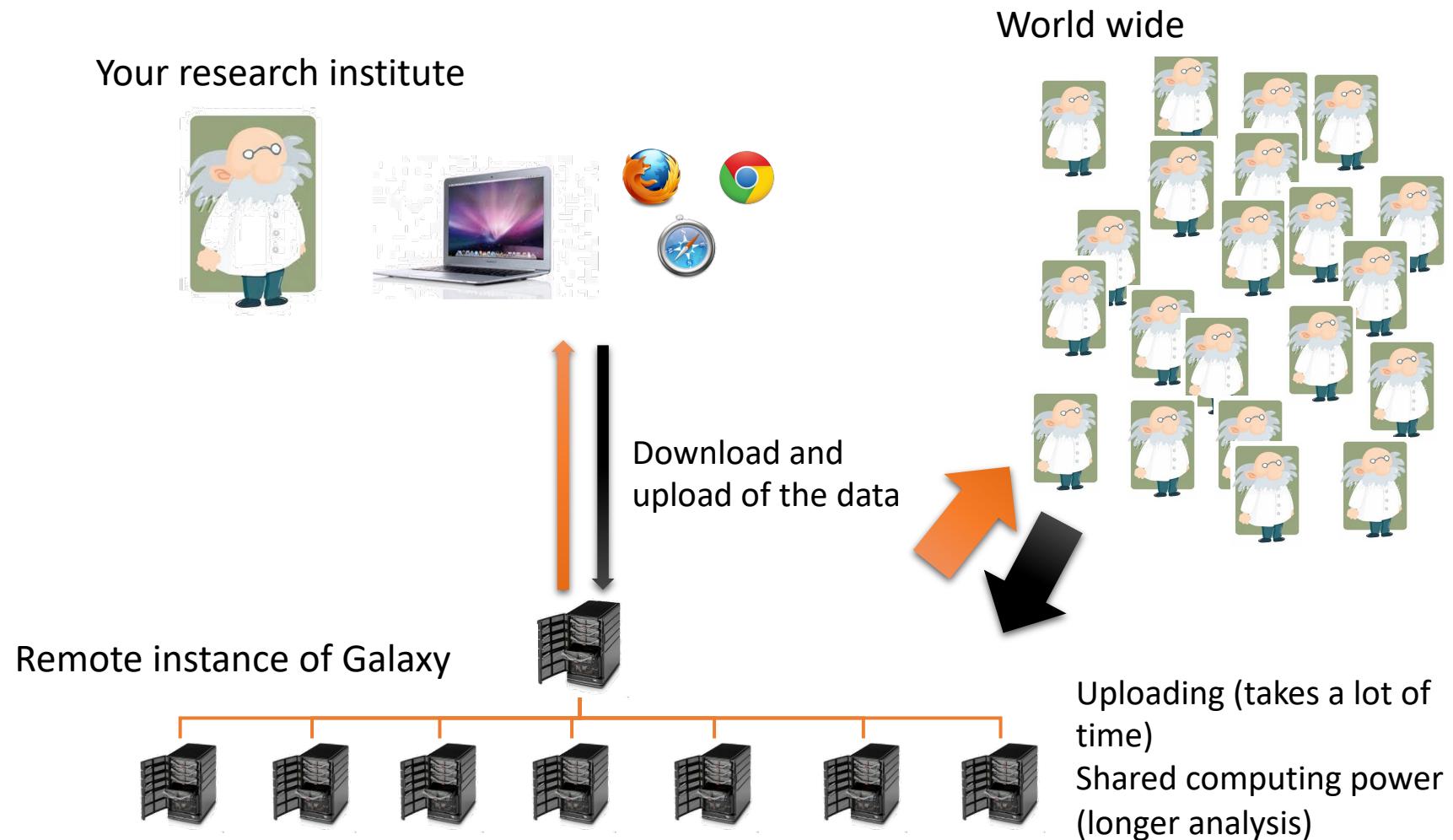
Remote instance of Galaxy



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
 - 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:** On the left, under the "Galaxy France" header, it 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), and Nucleic acids (Fetch Alignments, Sequences).
- 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 and a "Workflow4Metabolomics" tool.
- History panel:** On the right, it shows an empty history titled "Unnamed history" with 0 B of data and 0 items.

Annotations with orange arrows point to each of these sections: "Top menu" points to the top navigation bar; "Tool panel" points to the left sidebar; "Data display and tools dialog window" points to the central content area; and "History panel" points to the right sidebar.

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 a list of operations: 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). The "GENERAL TEXT TOOLS" section is currently selected. The main content area also includes a "History" panel on the right showing an empty history named "Unnamed history". The top navigation bar contains several items with orange callout boxes: "Workflow" (labeled "Run workflows"), "Shared Data" (labeled "Access public data"), "Help" (labeled "Get Help"), "User" (labeled "Get access to training materials"), and "Log in/out, manage your account" (labeled "Log in/out, manage your account").

Run workflows

Access public data

Get Help

Get access to training materials

Run analyses

Log in/out, manage your account

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

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

Refresh History

You have 29 histories.

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

Return to legacy panel

The screenshot shows the 'History' panel with a list of histories. At the top right is a context menu titled 'History options'. The menu includes options like 'Show Histories Side-by-Side', 'Copy this History', and 'Delete this History'. Below the menu is a list of 29 histories, with the first one, 'Unnamed history', selected. This history card displays its name, size (0 B), and a refresh icon. Labels with orange arrows point to various UI elements: 'History options' points to the menu; 'Switch history' and 'Create new history' are menu items; 'Search datasets' points to the search bar; 'History name' points to the history card; and 'Refresh History' points to the refresh icon on the history card.

View all histories

The screenshot displays the Galaxy France interface with several history panels:

- Current History**: An unnamed history panel labeled "(empty)". It includes a "search datasets" button and a message: "Drag datasets here to copy them to the current history". Below this, it says "This history is empty".
- DNA-seq data analysis**: A history panel containing 14 datasets. Recent items include:
 - 15: VCFtoTab-delimited: on data 13
 - 14: SnpEff eff: on data 12 - H TML stats
 - 13: SnpEff eff: on data 12
 - 12: FreeBayes on data 11 and data 10 (variants)
 - 11: CaptureDesign_chr4.bed
 - 10: MarkDuplicates on data 8: MarkDuplicates BAM output
 - 9: MarkDuplicates on data 8: MarkDuplicate metrics
 - 8: Map with BWA-MEM on data 3 and data 2 (mapped reads in BAM 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
- Neuro-epigenetics training**: A history panel containing 17 datasets. Recent items include:
 - 17: bamCoverage on data 15
 - 16: bamCoverage on data 14
 - 15: WT_320_St.chr19.bam
 - 14: R6_1_387_St.chr19.bam
 - 13: RNA STAR on data 7 and data 2: mapped.bam
 - 12: RNA STAR on data 7 and data 2: splice junctions.bed
 - 11: RNA STAR on data 7 and data 2: log
 - 10: RNA STAR on data 7 and data 1: mapped.bam
 - 9: RNA STAR on data 7 and data 1: splice junctions.bed
 - 8: RNA STAR on data 7 and data 1: log
 - 7: Mus_musculus.NCBIM37.67_UCSConlychr.gtf
 - 6: FastQC on data 2: RawData
 - 5: FastQC on data 2: Webpage
 - 4: FastQC on data 1: RawData
- Neuro-epigenetics training (data)**: A history panel containing 5 datasets. Recent items include:
 - 5: WT_320_St.chr19.bam
 - 4: R6_1_387_St.chr19.bam
 - 3: R6_1_387_St.chr19.fastq.gz
 - 2: WT_320_St.chr19.fastq.gz
 - 1: Mus_musculus.NCBIM37.67_UCSConlychr.gtf
- imported: Neuro-epigenetics training (data)**: A history panel containing 5 datasets. Recent items include:
 - 5: WT_320_St.chr19.bam
 - 4: R6_1_387_St.chr19.bam
 - 3: R6_1_387_St.chr19.fastq.gz
 - 2: WT_320_St.chr19.fastq.gz
 - 1: Mus_musculus.NCBIM37.67_UCSConlychr.gtf
- Neuro-epigenetics training**: A history panel containing 15 datasets. Recent items include:
 - 15: bamCoverage on data 15
 - 14: bamCoverage on data 14
 - 13: WT_320_St.chr19.bam
 - 12: R6_1_387_St.chr19.bam
 - 11: RNA STAR on data 5 and data 2: mapped.bam
 - 10: RNA STAR on data 5 and data 2: splice junctions.bed
 - 9: RNA STAR on data 5 and data 2: log
 - 8: RNA STAR on data 5 and data 1: mapped.bam
 - 7: RNA STAR on data 5 and data 1: splice junctions.bed
 - 6: RNA STAR on data 5 and data 1: log
 - 5: Mus_musculus.NCBIM37.67_UCSConlychr.gtf
 - 4: FastQC on data 2: RawData
 - 3: FastQC on data 2: Webpage
 - 2: R6_1_387_St.chr19.fastq.gz
 - 1: WT_320_St.chr19.fasta

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 categories: Tools, Get Data, Send Data, Collection Operations, and GENERAL TEXT TOOLS. The GENERAL TEXT TOOLS category is currently selected. In the main area, there's a search bar labeled "search tools" and a button labeled "Upload Data". A callout points to the "Upload Data" button with the text "Display the drag and drop utility used to upload local files". Below the search bar, there's a section titled "Download from web or upload from disk" 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 table displays the dataset in the queue:

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". There are also dropdown menus for "Type (set all): Auto-detect" and "Genome (set all): ----- Additional ...".

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), 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 of data libraries:

Name	Description	Actions
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'.

Annotations in the image:

- A red arrow points from the text "List of data libraries" to the "covid-19" library entry in the table.
- An orange arrow labeled '1.' points to the 'Access published resources' option in the Shared Data dropdown menu.
- An orange arrow labeled '2.' points to the 'Histories' option in the same dropdown menu.

Import public data

The screenshot shows the Galaxy France interface. In the top left corner, there is a logo and the text "Galaxy France". Below it, a "Tools" section has a search bar labeled "search tools" and a button labeled "Upload Data". Under the "Get Data" heading, there is a list of 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

A vertical orange bracket on the right side of the page groups the "Get Data" section and the "Table Browser" section.

Browse and import external data from public databases

The screenshot shows the "Table Browser" interface. At the top, there is a navigation bar with links for Home, Genomes, Genome Browser, Tools, Mirrors, Downloads, My Data, Help, and About Us. The main content area 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 text 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 chr1:21427557-121432936
- 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

At the bottom of the main content area, there is a link: "To reset all user cart settings (including custom tracks), [click here](#)".

Below the main content area, there is a section titled "Using the Table Browser" with the following text:

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

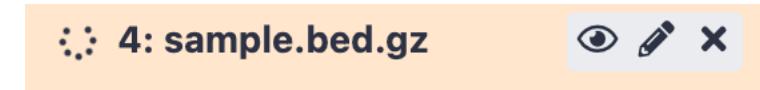
Exercise 3

Datasets/Jobs in the History

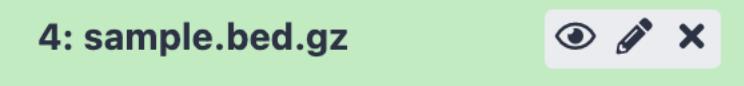
Grey: the job is waiting to run



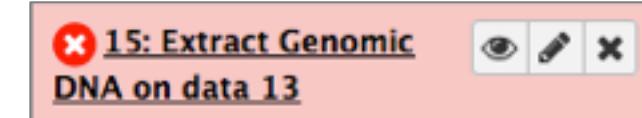
Orange: the job is running



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

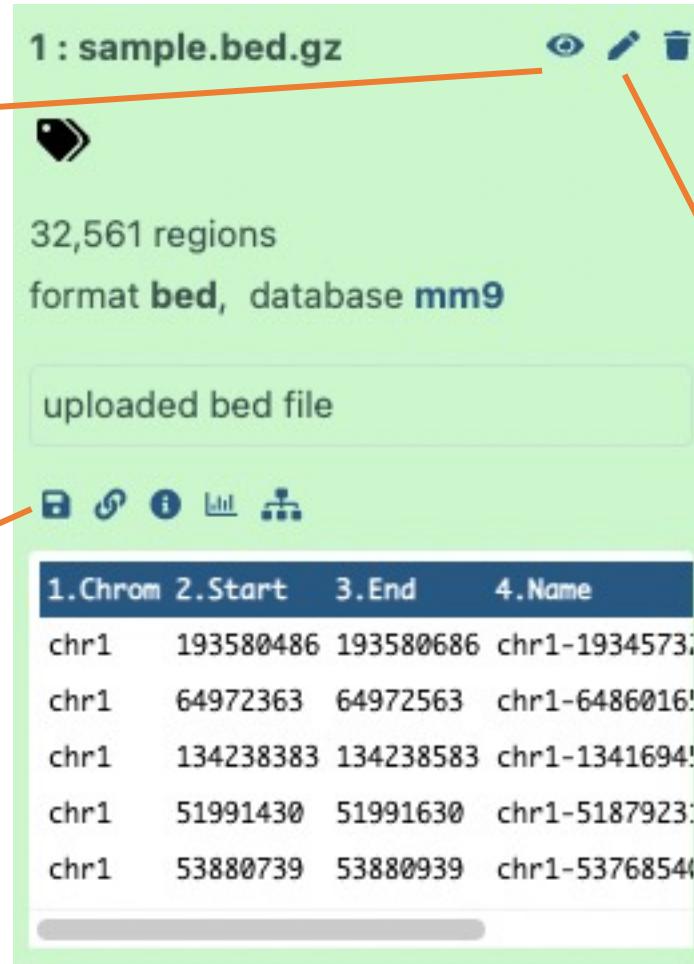
Format

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

The screenshot shows a dataset entry in a Nextflow history. The title is "1 : sample.bed.gz". Below it, there's a black download icon, the text "32,561 regions", and "format bed, database mm9". A green box contains the text "uploaded bed file". At the bottom, there's a table with four columns: "1.Chrom", "2.Start", "3.End", and "4.Name". The table displays five rows of genomic data from chromosome 1.

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



The screenshot shows a Galaxy dataset history panel for a file named "1 : sample.bed.gz". The panel includes the following information:

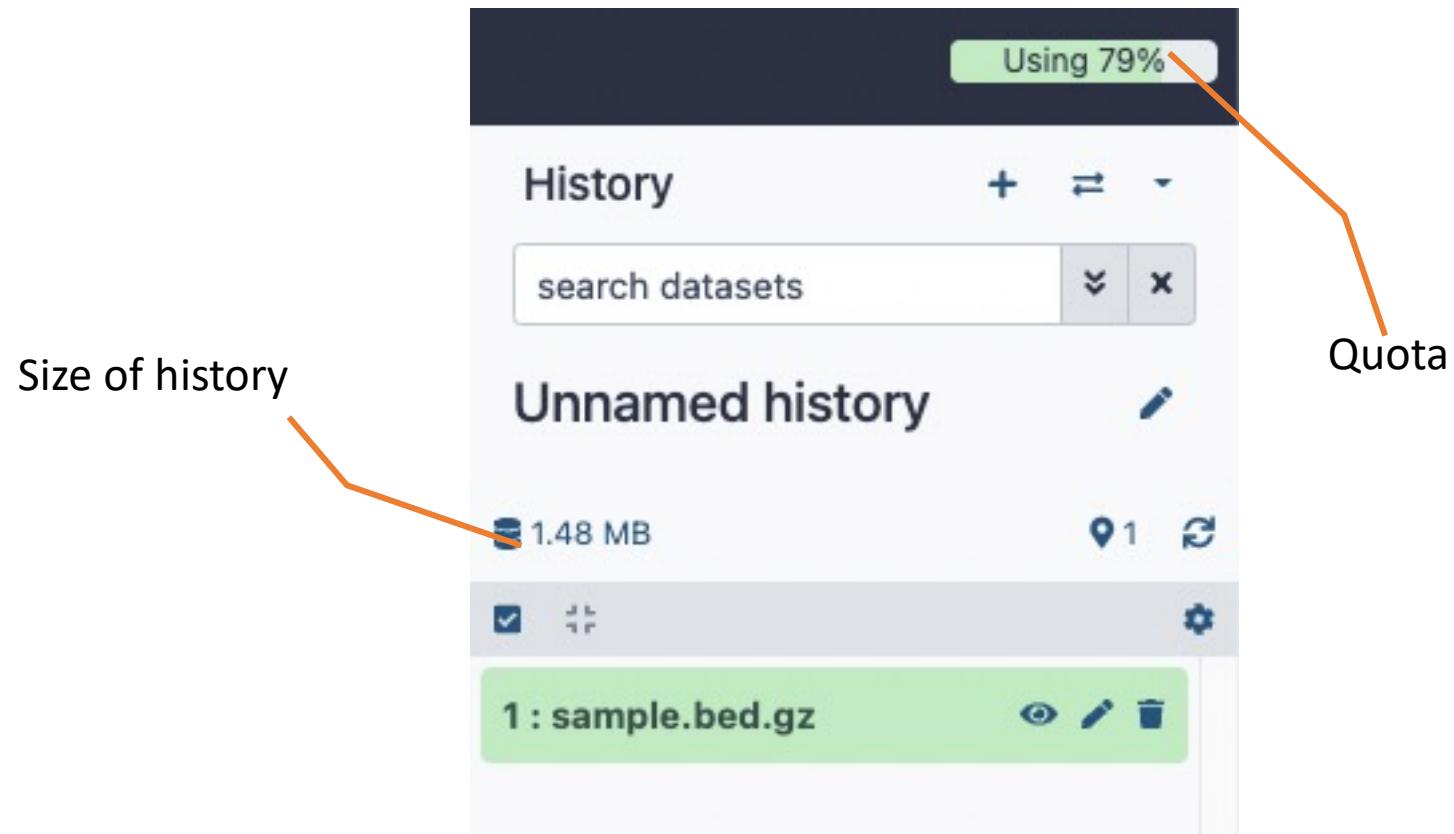
- File name: 1 : sample.bed.gz
- Icon: A small icon representing the dataset.
- Summary: 32,561 regions, format bed, database mm9
- Type: uploaded bed file
- Actions: A row of icons for managing the dataset.
- Data table: A table showing genomic coordinates for chromosome 1.

Annotations with orange arrows point to specific features:

- Above the file name: "View dataset (if possible) in the middle panel of Galaxy"
- To the right of the file name: "Delete dataset"
- Below the "uploaded bed file" type: "Edit attributes of the dataset (change name, format, genome, permission)"
- Below the "uploaded bed file" type: "Download dataset"

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

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 Fetch Aliases. A large orange callout box labeled "Tool panel" points to this sidebar. At the bottom of the sidebar, there are "Workflow", "Visualize", "Shared Data", "Help", "User", and a gear icon. The main content area features a colorful bar chart graphic and a "Welcome to usegalaxy.fr" message. Below the welcome message is a terms of use notice and a list of recent releases. To the right, a vertical sidebar titled "History" shows an "Unnamed history" section with a message stating "This history is empty. You can load your own data or get data from an external source." The top navigation bar includes icons for home, workflow, visualize, shared data, help, user, and a gear.

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

Fetch Aliases

Tool panel

Workflow Visualize Shared Data Help User

Using 79%

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%

Tool Panel / Run analyses

The screenshot shows the Tool Panel interface. At the top left is a search bar labeled "Search a tool". To its right is a "Tools" section with a "search tools" input field and a star icon. Below these are two buttons: "Upload Data" with an upward arrow icon and a large "x" button. A vertical scroll bar is on the right side of the panel.

Below the search bar is a list of categories:

- Upload data
- 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

An orange arrow points from the "Text Manipulation" category in the main list to a detailed view of the same category on the right. Another orange arrow points from the "Tool" heading in that detailed view back to the "Text Manipulation" category in the main list.

Text Manipulation

Query Tabular using sqlite sql

annotateMyIDs annotate a generic set of identifiers

Compute an expression on every row

Replace column by values which are defined in a convert file

Sort a row according to their columns

Column Regex Find And Replace

Regex Find And Replace

Tools dialog window

The screenshot shows the Galaxy France web interface. The top navigation bar includes links for Home, 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 sidebar on the right shows an empty history titled "Unnamed history" with a note: "This history is empty. You can load your own data or get data from an external source." The central content area has a large orange background with the text "Data display and tools dialog window" and a bulleted list: "– 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

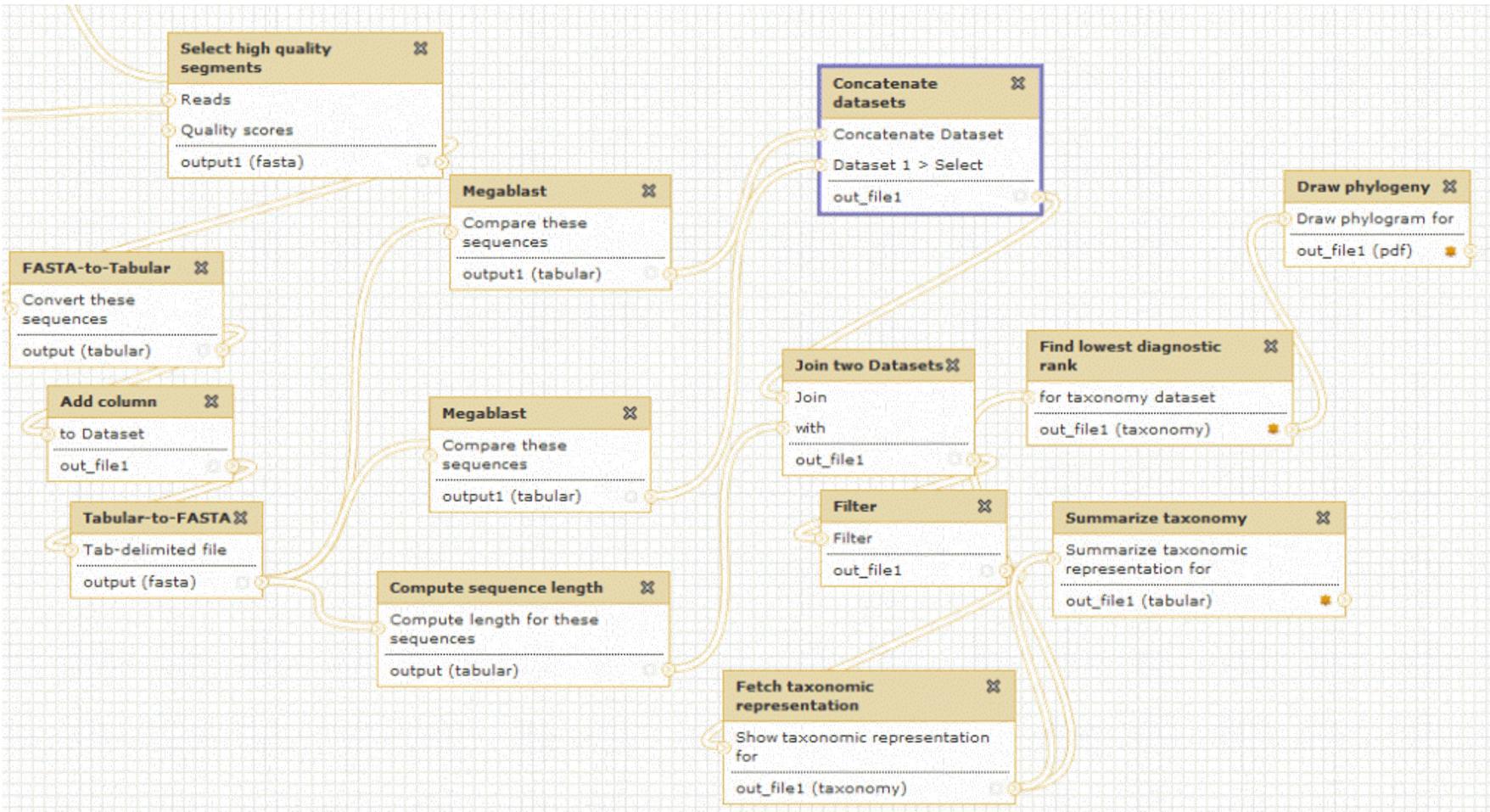
Exercise 6

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. A progress bar at the top right indicates "Using 79%". The left sidebar contains a "Tools" section with a search bar and "Upload Data" button, followed by sections for "Get Data", "Send Data", "Collection Operations", "GENERAL TEXT TOOLS" (selected), "Text Manipulation", "Filter and Sort", "Join, Subtract and Group", "GENOMIC FILE MANIPULATION" (selected), "Convert Formats", "FASTA/FASTQ", "FASTQ Quality Control", "SAM/BAM", "BED", "VCF/BCF", "Nanopore", "COMMON GENOMICS TOOLS" (selected), "Operate on Genomic Intervals", "Fetch Alignments/Sequences", "GENOMICS ANALYSIS" (selected), "Annotation", and "Assembly". The main content area features a "Welcome to usegalaxy.fr" message with a "Create, run, edit (...) workflows" callout. It also displays a "Domain specific subdomains" section for "Workflow4MetabolomicsCovid19" with a logo and a "Variant analysis, consensus using community" section. The right sidebar shows a "History" panel with an "Unnamed history" entry containing a dataset named "1 : sample.bed.gz".

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 accepted our Terms of Service. For any questions or support: [community.cluster.france-bioinformatics.fr](#)

• 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

Metabolomics data processing, analysis and annotation for Metabolomics

Variant analysis, consensus using community

History

search datasets

Unnamed history

1.48 MB

1 : sample.bed.gz

Using 79%

Workflows

Create workflows

The screenshot shows the Galaxy France web interface. On the left, a sidebar lists various tool categories under 'Tools'. A red box highlights the 'List of Workflows' section. At the top center, there's a search bar for 'Search Workflows' and a prominent '+ Create' button. To the right, a 'History' panel shows an 'Unnamed history' containing a single dataset named 'sample.bed.gz'. The main area displays a list of workflows, each with a name, updated date, sharing status, and a play button for execution.

Name	Updated	Sharing	Bookmarked
CHIP-seq data analysis workflow	9 months ago	☆	▶
BED Ensembl to BED UCSC	9 months ago	☆	▶
Unnamed workflow	9 months ago	☆	▶
CHIP-seq data analysis workflow	9 months ago	☆	▶
Run "Prepare RNA-seq data for seqMINER"	10 months ago	☆	▶
Run ChIP-seq hands-on	10 months ago	☆	▶
ChIP-seq data analysis workflow	10 months ago	☆	▶
BED Ensembl to BED UCSC	10 months ago	☆	▶
DNA-seq data analysis (DU Dijon)	12 months ago	☆	▶

Workflow creation

The screenshot shows the Galaxy France workflow creation interface. On the left, a sidebar titled "Tools" lists various categories: 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, etc. 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 panels are visible for "Name" (set to "My workflow"), "Version" (set to "1: Dec 13th 2022, 0 steps"), "Annotation" (a large empty text area), "License" (a placeholder "Specify a license for this workflow."), "Creator" (a placeholder "Add a new creator - either a person or an organization."), and "Tags" (a placeholder "Apply tags to make it easy to search for and find items with the same tag."). The top navigation bar includes links for Home, Workflow, Visualize, Shared Data, Help, User, and a search bar indicating "Using 79%".

Add tools or input datasets to the workflow

Galaxy France

Workflow Visualize Shared Data Help User

Using 79%

My workflow

Tools

search tools

Inputs

Get Data

Send Data

Collection Operations

Expression Tools

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

Name

My workflow

Version

1: Dec 13th 2022, 0 steps

Annotation

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

Apply tags to make it easy to search for and find items with the same tag.

100% +

52

Workflow creation

The screenshot shows the Galaxy France workflow creation interface. On the left, a sidebar lists various tools under the heading "Tools". A search bar at the top of the sidebar contains the text "filter". The main workspace is titled "My workflow" and contains two workflow steps:

- Step 1: Input Dataset** (highlighted with a blue border):
 - Icon: A blue square with a white arrow pointing right.
 - Label: "1: Input Dataset".
 - Input: "output (input)"
- Step 2: Filter**:
 - Icon: A blue square with a white wrench icon.
 - Label: "2: Filter".
 - Input: "out_file1 (input)"

A blue line connects the output of Step 1 to the input of Step 2. To the right of the workspace, there is a panel for "Input Dataset" configuration:

- Label:** A text input field with placeholder "Add a step label."
- Step Annotation:** A text area with placeholder "Add an annotation or notes to this step. Annotations are available when a workflow is viewed."
- Optional:** A toggle switch set to "No".
- Format(s):** A text input field with placeholder "Leave empty to auto-generate filtered list at runtime based on connections."
- Tag filter:** A text input field with placeholder "Tags to automatically filter inputs"

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

Using 79%

Workflow Visualize Shared Data Help User

Workflow creation

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

The screenshot shows the Galaxy France workflow creation interface. On the left, a sidebar lists various tools under categories like 'Inputs', 'Filter and Sort', 'SAM/BAM', 'Metagenomic Analysis', 'VCF/BCF', and 'Qiime2'. In the center, a workflow canvas displays two steps: '1: Input Dataset' and '2: Filter'. A green line links the 'output (input)' of step 1 to the 'out_file1 (input)' of step 2. To the right of the canvas is a panel for step 2, which includes sections for 'Input Dataset' (with 'Label' and 'Step Annotation' fields), 'Optional' (with a 'No' radio button), 'Format(s)' (with a note about auto-generation), and 'Tag filter'.

Workflow creation

Galaxy France Workflow Visualize Shared Data Help User Using 79%

Tools My workflow

Inputs

Filter and Sort
SAM/BAM
Metagenomic Analysis
VCF/BCF
Qiime2

1: Input 2: Filter

Dataset Filter

output (input) out_file1 (input)

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

Add a step label.

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

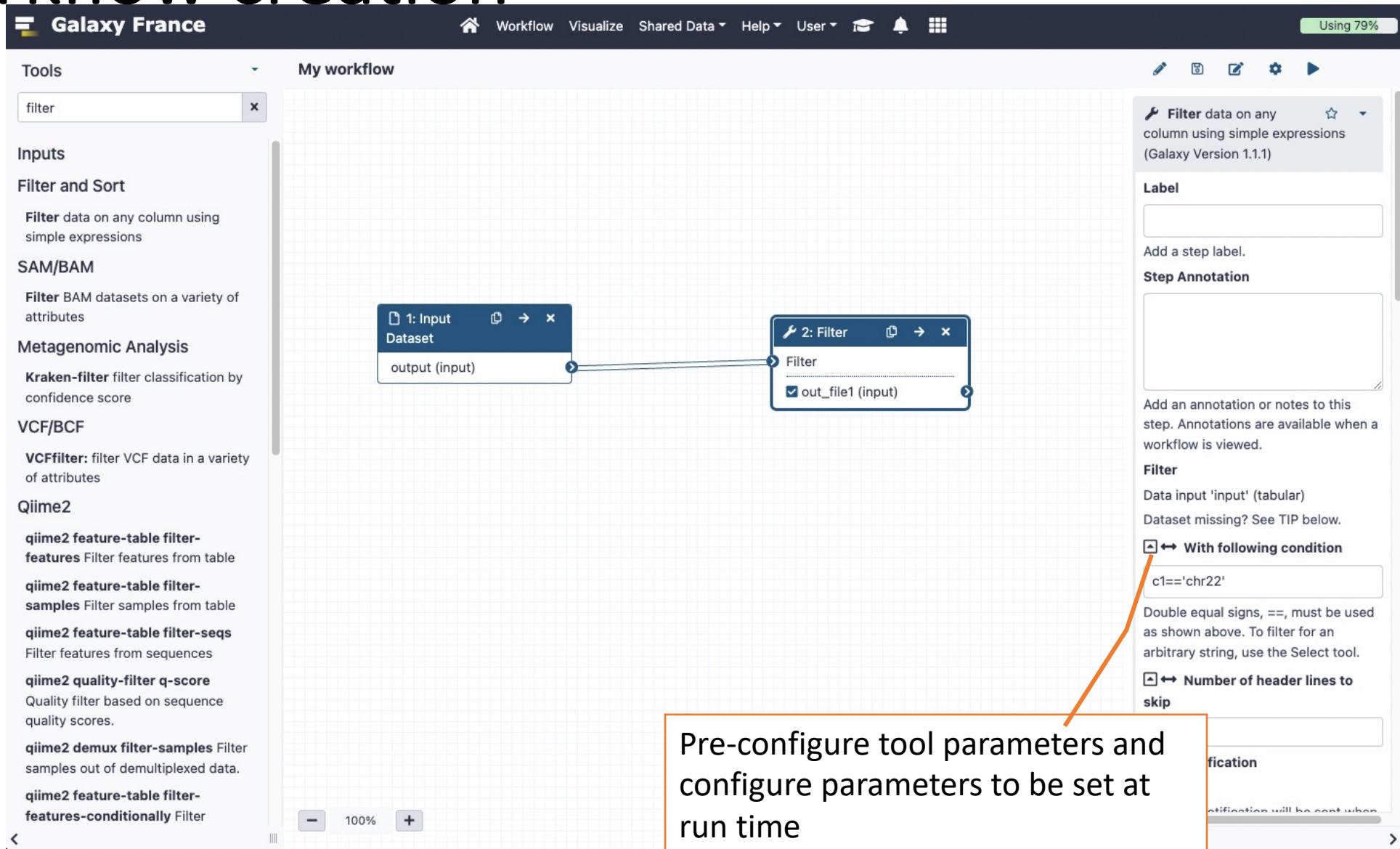
Data input 'input' (tabular)
Dataset missing? See TIP below.

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

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



Workflow creation

Save workflow

Run

The screenshot shows the Galaxy France workflow creation interface. On the left, a sidebar lists various tools under categories like Tools, Inputs, Filter and Sort, SAM/BAM, Metagenomic Analysis, VCF/BCF, and Qiime2. The main workspace displays a workflow consisting of two steps: '1: Input' and '2: Filter'. Step 1 is a 'Dataset' tool with an output named 'output (input)'. Step 2 is a 'Filter' tool with an output named 'out_file1 (input)'. An arrow connects the output of step 1 to the input of step 2. To the right of the workflow, a detailed configuration panel for the '2: Filter' step is open. It includes sections for 'Label' (empty), 'Step Annotation' (empty), 'Filter' (described as 'Data input 'input' (tabular) Dataset missing? See TIP below.'), 'With following condition' (containing the expression 'c1==\'chr22\''), 'Number of header lines to skip' (set to 0), and 'Email notification' (disabled). The top right of the interface shows a progress bar 'Using 79%' and several action icons.

Galaxy France

Workflow Visualize Shared Data Help User

My workflow

Tools

filter

Inputs

Filter and Sort

Filter data on any column using simple expressions

SAM/BAM

Filter BAM datasets on a variety of attributes

Metagenomic Analysis

Kraken-filter filter classification by confidence score

VCF/BCF

VCFfilter: filter VCF data in a variety of attributes

Qiime2

qiime2 feature-table filter-features Filter features from table

qiime2 feature-table filter-samples Filter samples from table

qiime2 feature-table filter-sequences Filter features from sequences

qiime2 quality-filter q-score Quality filter based on sequence quality scores.

qiime2 demux filter-samples Filter samples out of demultiplexed data.

qiime2 feature-table filter-features-conditionally Filter

1: Input

Dataset

output (input)

2: Filter

Filter

out_file1 (input)

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.

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

0

Email notification

No

An email notification will be sent when

Using 79%

Save workflow

Run

100%

56

Run workflows

The screenshot shows the Galaxy France web interface for running workflows. On the left, a sidebar lists various tool categories. The main area displays a workflow titled "Workflow: My workflow". The workflow consists of two steps: Step 1, "Input dataset", which has an input field containing "1: sample.bed.gz"; Step 2, "Filter (Galaxy Version 1.1.1)", which includes a "With following condition" section with the value "c1=='chr22'" and a "Number of header lines to skip" section set to "0". To the right, a history panel shows an "Unnamed history" containing a dataset named "1 : sample.bed.gz". A red arrow points from the "sample.bed.gz" entry in the history to the "Run Workflow" button at the top right of the workflow editor. Another red arrow points from the "Set parameters" text to the "Number of header lines to skip" input field. A callout box with the text "Set input file(s). It has to be a dataset from your current history" is positioned above the history panel.

Set input file(s). It has to be a dataset from your current history

Run workflow

Set parameters

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

GENOMICS ANALYSIS

Annotation

Assembly

Workflow: My workflow

History Options

Send results to a new history

No

1: Input dataset

input

1: sample.bed.gz

2: Filter (Galaxy Version 1.1.1)

Filter

Connected to 'output' from Step 1

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

0

Using 79%

history

search datasets

Unnamed history

1 : sample.bed.gz

57

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