

Introduction to Galaxy

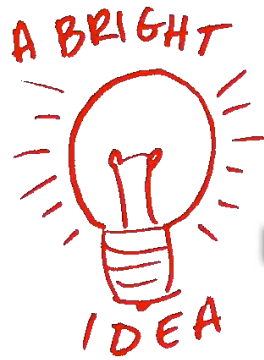
Stéphanie Le Gras
(slegras@igbmc.fr)

Guidelines

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

Analyzing biological data with informatics tools

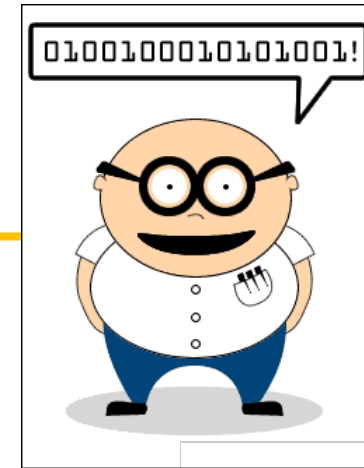
Bioinformatics analyses



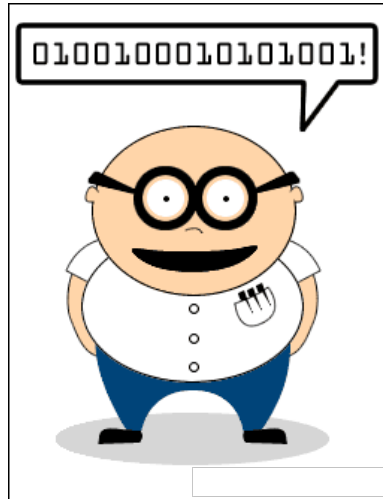
Informatics data



nature



Bioinformatics analyses



Scripts, softwares

```
#!/usr/bin/perl

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

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

## Objectives :

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

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

my $usage = <<END;

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

END

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

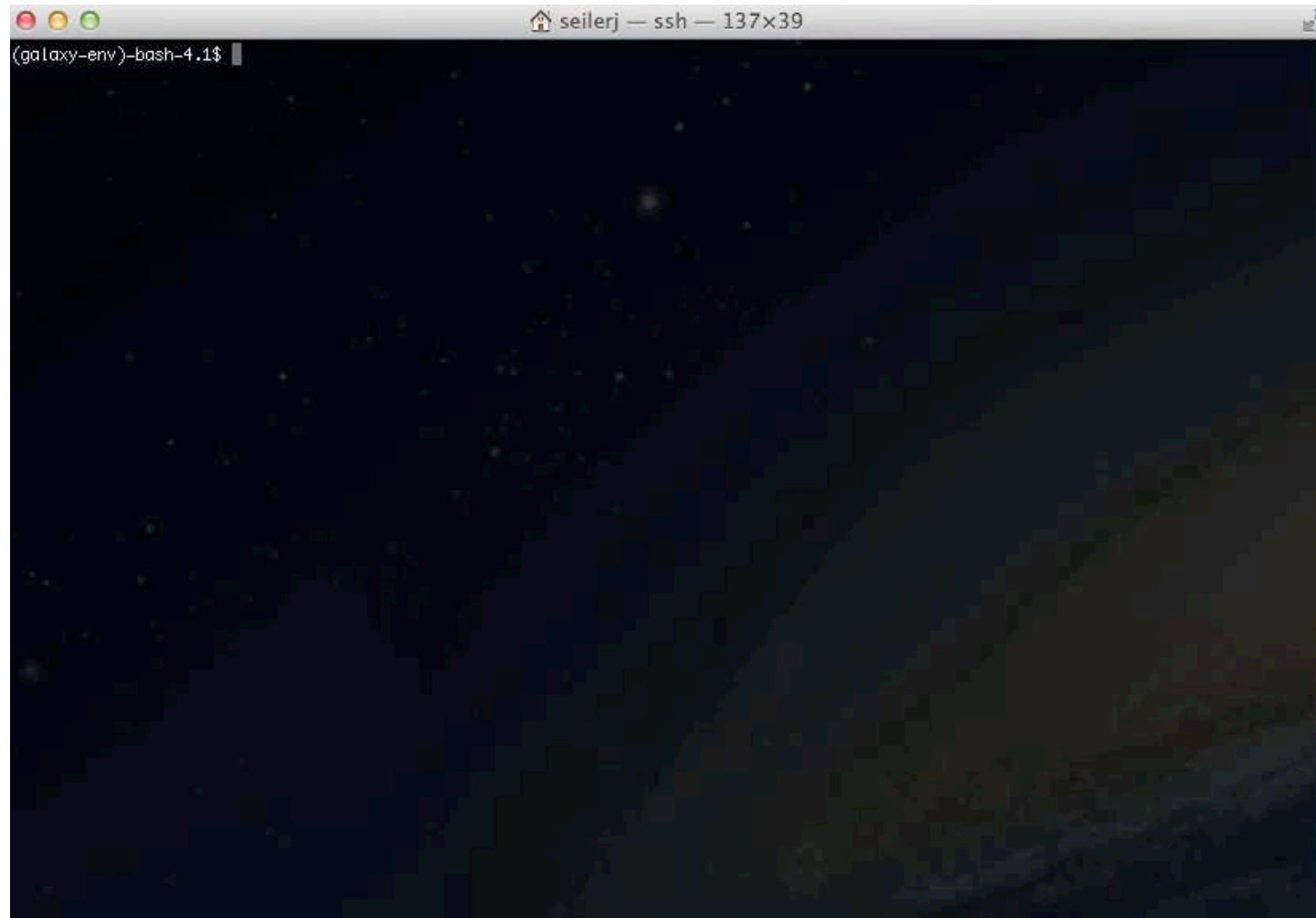
my %ids;
$out = ( defined $out ) ? $out : "results.txt";

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

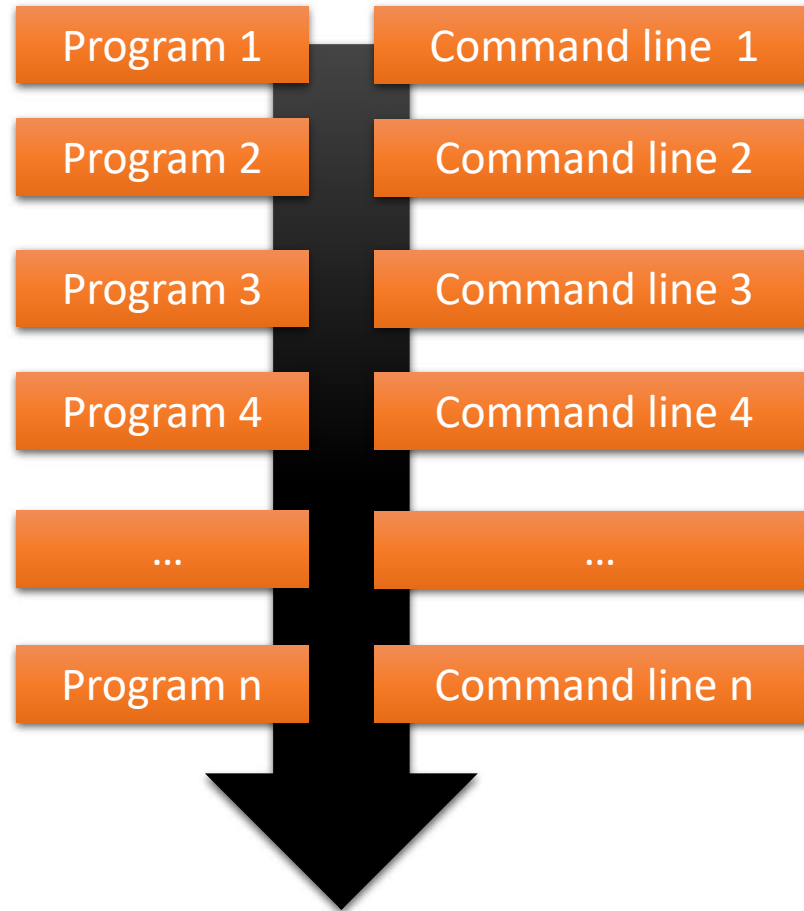
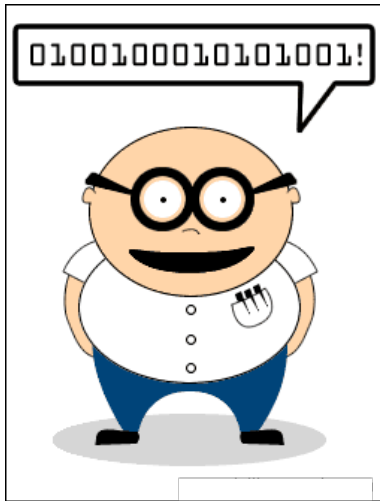
Command line

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

Bioinformatics analyses



Bioinformatics analyses



**PIPELINE /
WORKFLOW**

Galaxy ?





Galaxy

PROJECT

Galaxy project

What is Galaxy ?

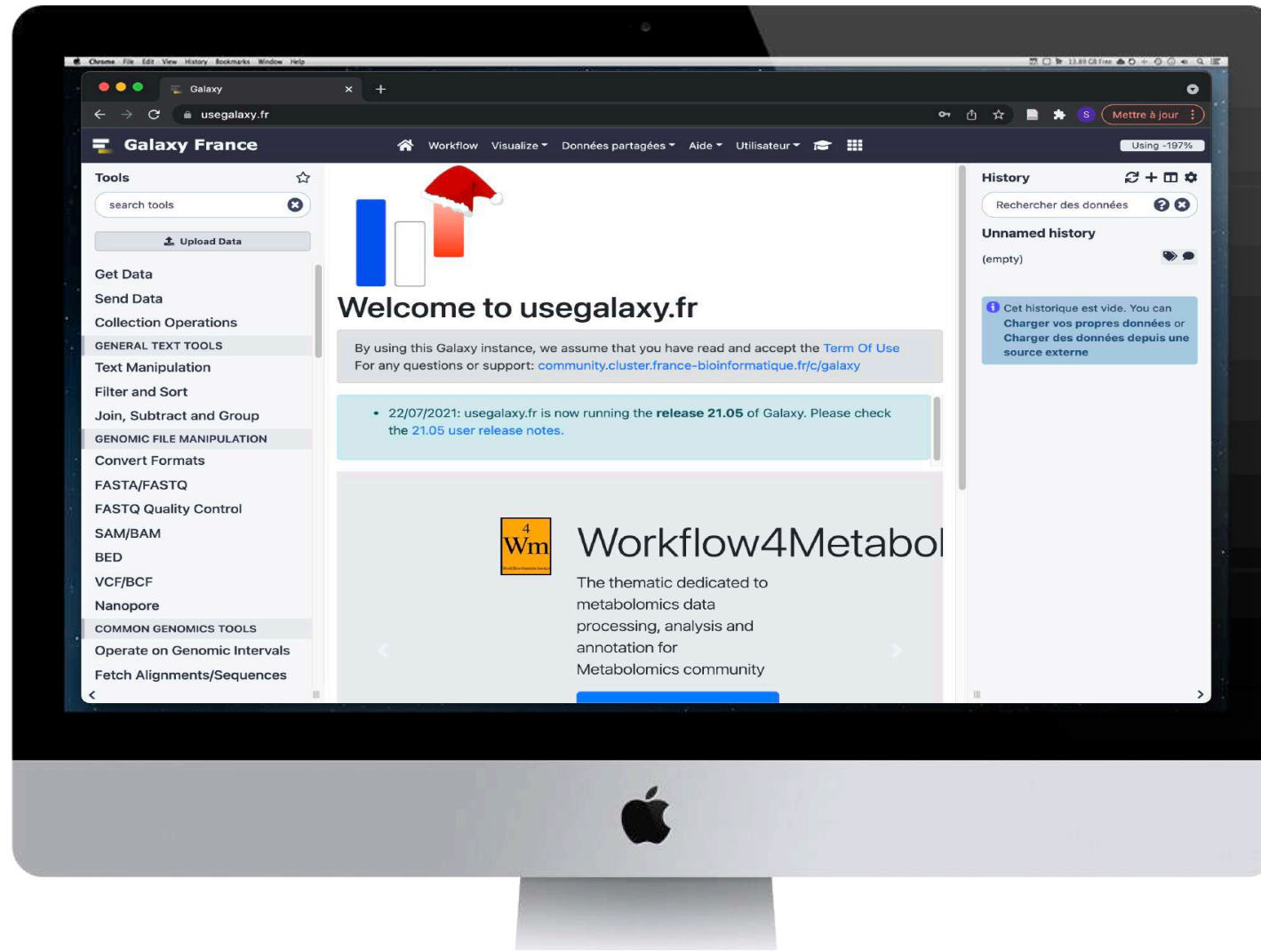
Galaxy is a **computing platform** that enables people to **run complex bioinformatics tools** on a **compute cluster** through a **simple web interface**.



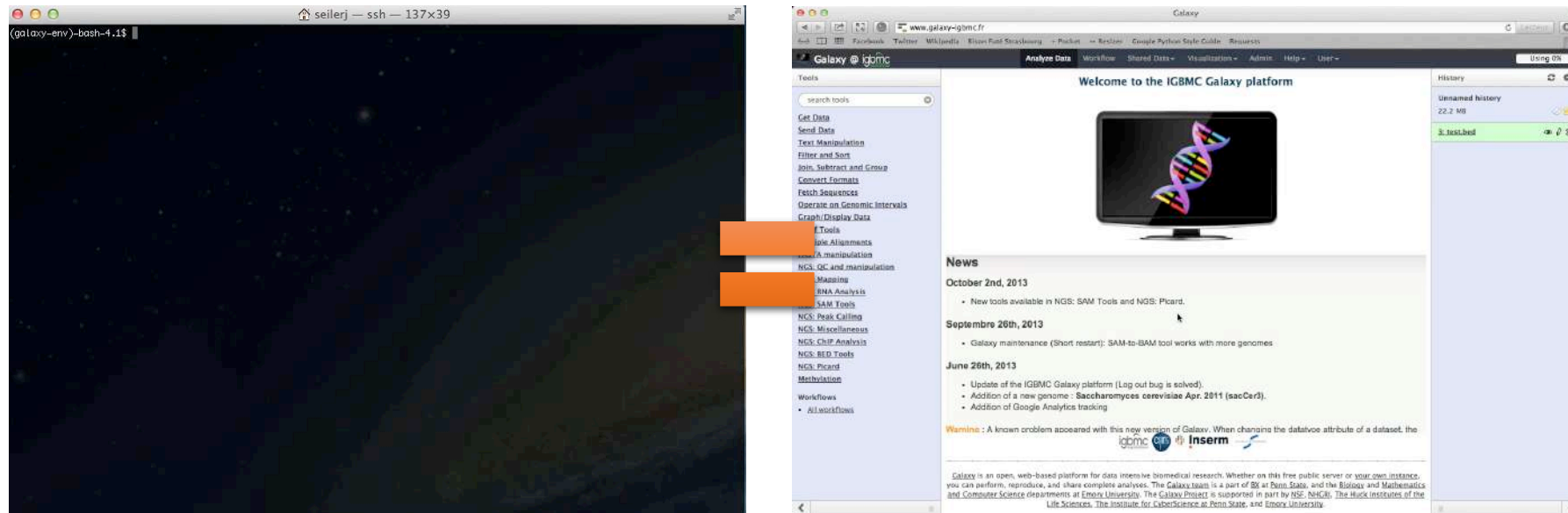
EMORY
UNIVERSITY



This is Galaxy



Running analyses with tools



The image shows two side-by-side screenshots. The left screenshot is a terminal window titled 'seilerj - ssh - 137x39' with the prompt '(galaxy-env)-bash-4.1\$'. The right screenshot is a web browser window titled 'Galaxy' showing the 'IGBMC Galaxy platform' homepage. The browser address bar is 'www.galaxy-igbmc.fr'. The page features a 'Welcome to the IGBMC Galaxy platform' message with a DNA double helix icon on a monitor. Below this is a 'News' section with dates and updates: 'October 2nd, 2013' (New tools available in NGS: SAM Tools and NGS: Picard), 'September 26th, 2013' (Galaxy maintenance (Short restart): SAM-to-BAM tool works with more genomes), and 'June 26th, 2013' (Update of the IGBMC Galaxy platform (Log out bug is solved), Addition of a new genome: Saccharomyces cerevisiae Apr. 2011 (sacCer3), Addition of Google Analytics tracking). A 'Warnings' section at the bottom states: 'A known problem appeared with this new version of Galaxy. When changing the datavoe attribute of a dataset, the'. Logos for 'igbmc', 'EBS', and 'Inserm' are visible. A 'History' sidebar on the right shows 'Unnamed history' with '22.2 MB' and a file named '3_test.bed'. The terminal window on the left has two orange arrows pointing to the 'Tools' sidebar of the Galaxy interface, which lists categories like 'Get Data', 'Text Manipulation', 'Filter and Sort', 'Join, Subtract and Group', 'Converts Formats', 'Fetch Sequences', 'Operate on Genomic Intervals', 'Graph/Display Data', 'Tools', 'BWA Alignments', 'PA manipulation', 'NGS: QC and manipulation', 'Mapping', 'RNA Analysis', 'SAM Tools', 'NGS: Peak Calling', 'NGS: Miscellaneous', 'NGS: ChIP Analysis', 'NGS: BED Tools', 'NGS: Picard', 'Methylation', 'Workflows', and 'All workflows'.

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

Your research institute



World wide



Download and upload of the data



Remote instance of Galaxy



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

Galaxy local server

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



Description of the main features of Galaxy

Galaxy web interface

The screenshot shows the Galaxy France web interface. At the top, there is a navigation bar with the title "Galaxy France" and several menu items: "Workflow", "Visualize", "Données partagées", "Aide", "Utilisateur", and a grid icon. A status indicator on the right shows "Using -197%".

On the left side, there is a "Tools" panel with a search bar and an "Upload Data" button. Below this, there are several categories of 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", "N", "C", "O", "Fetch Align", and "Sequences".

The main content area features a "Welcome to usegalaxy.fr" message with a red Santa hat icon. Below the welcome message, there is a notice about the release 21.05 of Galaxy. At the bottom of the main area, there is a section for "Workflow4Metabol" with a logo and the text "The thematic dedicated to".

On the right side, there is a "History" panel with a search bar and a message indicating that the history is empty. Below this, there is a blue box with the text "Cet historique est vide. You can Charger vos propres données or Charger des données depuis une source externe".

Three orange callout boxes with arrows point to specific parts of the interface: "Top menu" points to the navigation bar, "Tool panel" points to the left sidebar, and "History panel" points to the right sidebar. A fourth orange callout box, "Data display and tools dialog window", points to a blue bar at the bottom of the main content area.

Top menu

The image shows the top navigation bar of the Galaxy France interface. The navigation bar includes the following items from left to right: a home icon, 'Workflow', 'Visualize', 'Données partagées', 'Aide', 'Utilisateur', a graduation cap icon, and a grid icon. A 'Using -197%' indicator is on the far right. Below the navigation bar, several callout boxes with orange arrows point to specific elements:

- Run workflows**: Points to the 'Workflow' menu item.
- Access public data**: Points to the 'Données partagées' menu item.
- Get Help**: Points to the 'Aide' menu item.
- Get access to training materials**: Points to the graduation cap icon.
- Run analyses**: Points to the 'Utilisateur' menu item.
- Log in/out, manage your account**: Points to the graduation cap icon.

The main content area below the navigation bar displays the 'Galaxy France' logo, a search bar for tools, an 'Upload Data' button, and a sidebar with categories like 'GENERAL TEXT TOOLS' and 'GENOMIC FILE MANIPULATION'. The central panel shows a 'Welcome to usegalaxy.fr' message with a 'Term Of Use' link and a notification about the 21.05 release. The right sidebar shows a 'History' section with a search bar and a message indicating an empty history.

Hands On

Exercise 1

History

The screenshot displays the Galaxy France web interface. On the left is a navigation sidebar with categories like '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', and 'Fetch Alignments/Sequences'. The main content area features a 'Welcome to usegalaxy.fr' message with a Santa hat icon, a 'Term Of Use' link, and a release update notice for Galaxy 21.05. Below this is a 'Workflow4Metabolomics' section. On the right, the 'History' panel is highlighted with an orange border. It contains a search bar, an 'Unnamed history (empty)' section, and a message: 'Cet historique est vide. You can Charger vos propres données or Charger des données depuis une source externe'. The bottom right corner of the History panel has an orange background with the text 'History panel Keep track of each job run'.

History

History/Dataset options


View all histories side by side

Create new history

Refresh History

Search datasets

History name



History Actions

- Copy
- Partager et publier
- Montrer la structure
- Extraire un Workflow
- Set Permissions
- Make Private
- Reprendre les processus en pause

Actions sur les jeux de données

- Copier des jeux de données
- Réduire les données étendues
- Afficher les données cachées
- Supprimer les données cachées
- Purger les données supprimées

Télécharger

- Exporter les citations des outils
- Exporter l'Historique dans un fichier

Beta Features

View all histories

The screenshot shows the Galaxy France interface with a grid of history panels. The top navigation bar includes 'Galaxy France', 'Workflow', 'Visualize', 'Données partagées', 'Aide', 'Utilisateur', and 'Using -197%'. Below the navigation bar are search bars for 'search histories' and 'search all datasets'. The main content area is divided into several history panels, each with a 'Switch to' dropdown and a search bar.

Unnamed history (empty)

DNA-seq data analysis - run workflow (17 shown, 147.13 MB)

- 17: SnpEff on data 14 - stats
- 16: SnpEff on data 14
- 15: Haplotype Caller on data 9, data 10, and data 11 (log)
- 14: Haplotype Caller on data 9, data 10, and data 11 (VCF)
- 13: Flagstat on data 10
- 12: Base Recalibrator on data 9, data 10, and data 8 (log)
- 11: Base Recalibrator on data 9, data 10, and data 8 (Covariate File)
- 10: Filter on data 8

DNA-seq data analysis (17 shown, 1 hidden, 782.51 MB)

- 17: SnpEff on data 14 - stats
- 16: SnpEff on data 14
- 15: Haplotype Caller on data 7, data 8, and others (log)
- 14: Haplotype Caller on data 7, data 8, and others (VCF)
- 13: Base Recalibrator on data 7, data 8, and data 11 (log)
- 12: Base Recalibrator on data 7, data 8, and data 11 (Covariate File)
- 11: MarkDuplicates on data 9: MarkDuplicates BAM output
- 10: MarkDuplicates on data

MegaDebug_3 (15 sept 2016) (48 shown, 1 hidden, 69.7 MB)

- 48: SamToFastq run
- 47: SamToFastq run (UNPAIRED_READS)
- 46: SamToFastq run (READ2)
- 45: SamToFastq run (READ1)
- 44: SamToFastq run (UNPAIRED_READS)
- 43: SamToFastq run (READ2)
- 42: SamToFastq run (READ1)
- 41: ReplaceSamHeader on data 37 and data 2: BAM file with replaced header
- 40: MergeSamFiles on data 2 and data 35: Merged BAM

DNA-seq (19 shown, 1 hidden, 800.21 MB)

- 32: SnpEff
- 31: SnpEff
- 30: Haplotype
- 29: Haplotype
- 15: Base Re
- 14: Base Re
- 13: MarkDu

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

Galaxy France

Tools

search tools

Upload Data

Get Data

Send Data

Collection Operations

GENERAL TEXT TOOLS

Download from web or upload from disk

Regular Composite Collection Rule-based

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

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

Type (set all): Auto-detect Genome (set all): ----- Additional ...

Choose local files Choose remote files Paste/Fetch data Start Pause Reset Close

Display the drag and drop utility used to upload local files

Name of the dataset

Size of the dataset

File format

Genome

Import shared data (data libraries)

1. 2.

List of data libraries →

The screenshot shows the Galaxy France interface. At the top, there is a navigation bar with 'Galaxy France' on the left and 'Workflows', 'Visualize', 'Données partagées', 'Aide', 'Utilisateur', and a user profile icon on the right. A dropdown menu is open under 'Données partagées', with 'Bibliothèque de données' selected. Below the navigation bar, there is a search bar with 'Search' and an 'exclude restricted' checkbox. The main content area displays a table of data libraries. The table has columns for 'Name', 'Description', and 'Synopsis'. The 'Name' column is highlighted in blue. The 'Synopsis' column is also highlighted in blue. The table contains five rows of data libraries. At the bottom of the table, there is a pagination control showing '1' of 5 pages, with '10' items per page and '5 total' items. The URL 'https://usegalaxy.fr/libraries' is visible in the bottom left corner.

Name	Description	Synopsis
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)	

« < 1 > » 10 per page, 5 total

https://usegalaxy.fr/libraries

Import shared data (data libraries)

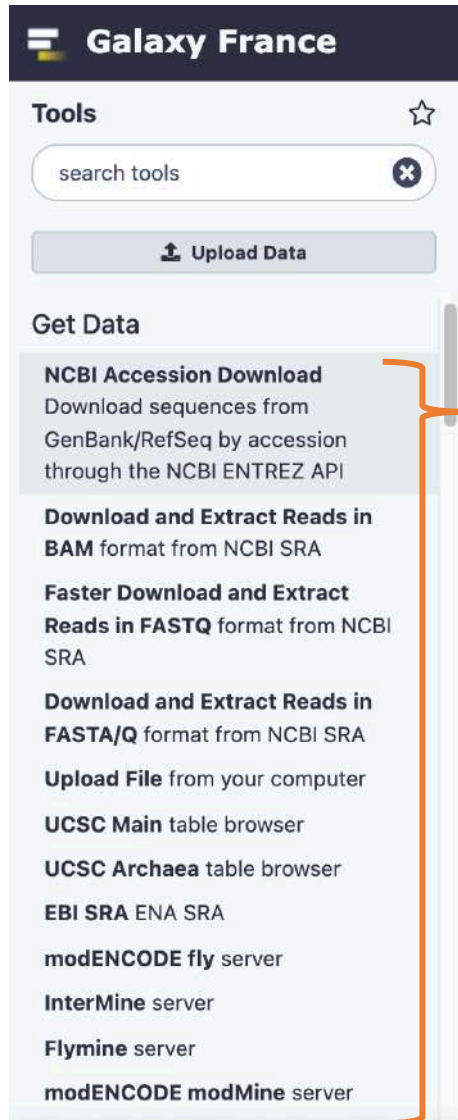
2. Import selected dataset to history

1. Select dataset

The screenshot shows the Galaxy France interface. At the top, there is a navigation bar with the logo 'Galaxy France' and several menu items: 'Workflow', 'Visualize', 'Données partagées', 'Aide', 'Utilisateur', and a grid icon. On the right side of the navigation bar, it says 'Using -197%'. Below the navigation bar, there is a search bar and a toolbar with buttons for 'Export to History', 'Download', 'Delete', and 'Details'. There is also a checkbox for 'include deleted'. The main content area is titled 'Libraries / GTN - Material'. It contains a table with the following columns: 'Name', 'Description', 'Type', 'Size', 'Updated', and 'State'. The table lists several folders, each with a checkbox in the 'Name' column. The folders are: 'Assembly', 'ChIP-Seq data analysis', 'Ecology', 'Epigenetics', 'Genome Annotation', 'Imaging', 'Introduction to Galaxy Analyses', 'Metabolomics', 'Metagenomics', and 'PAPAA PI3K_OG:Pancancer Aberrant Pathway Activity Analysis'. An orange arrow points from the 'Export to History' button in the toolbar to the 'Assembly' folder in the table. Another orange arrow points from the 'Export to History' button in the toolbar to the 'Export to History' button in the toolbar.

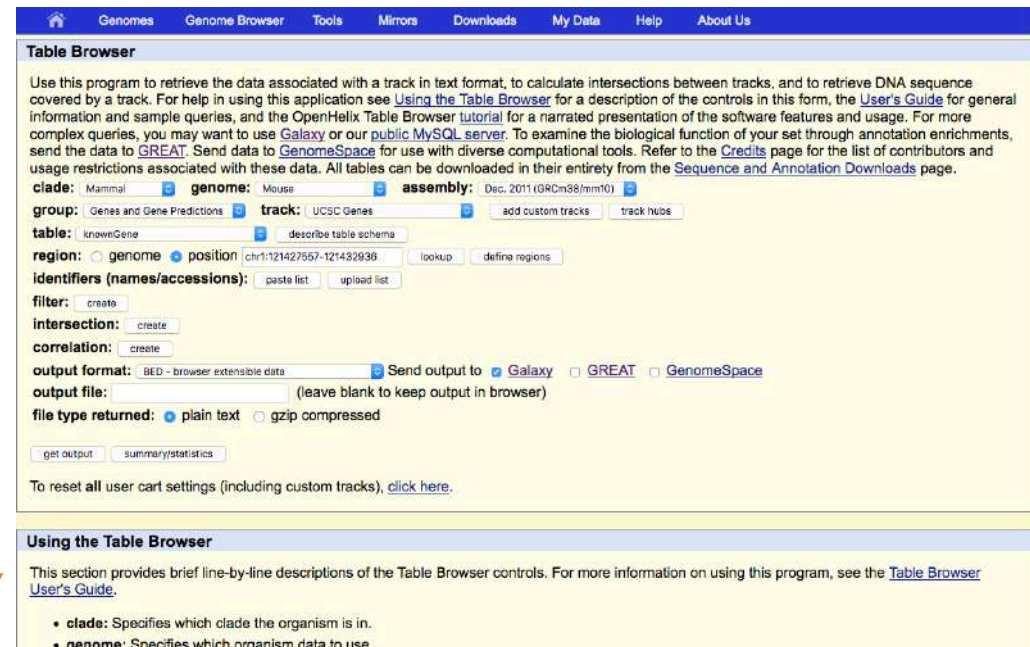
<input type="checkbox"/>	Name	Description	Type	Size	Updated	State
<input type="checkbox"/>	Assembly	DNA sequence data has become an indispen ... (more)	folder		11 days ago	
<input type="checkbox"/>	ChIP-Seq data analysis	ChIP-sequencing is a method used to anal ... (more)	folder		a year ago	
<input type="checkbox"/>	Ecology	Learn to analyse Ecological data through ... (more)	folder		a year ago	
<input type="checkbox"/>	Epigenetics	DNA methylation is an epigenetic mechani ... (more)	folder		a year ago	
<input type="checkbox"/>	Genome Annotation	Genome annotation is a multi-level proce ... (more)	folder		a year ago	
<input type="checkbox"/>	Imaging	Image analysis using Galaxy tools	folder		a year ago	
<input type="checkbox"/>	Introduction to Galaxy Analyses	Galaxy is a scientific workflow, data in ... (more)	folder		4 months ago	
<input type="checkbox"/>	Metabolomics	Training material to analyse Mass spectr ... (more)	folder		a year ago	
<input type="checkbox"/>	Metagenomics	Metagenomics is a discipline that enable ... (more)	folder		a year ago	
<input type="checkbox"/>	PAPAA PI3K_OG:Pancancer Aberrant Pathway Activity Analysis	Summary	folder		5 months ago	

Import public data



The screenshot shows the Galaxy France sidebar menu. At the top is the 'Galaxy France' logo. Below it is a 'Tools' section with a search bar and an 'Upload Data' button. The 'Get Data' section is highlighted with an orange bracket and contains several options: 'NCBI Accession Download', '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', and 'modENCODE modMine server'.

Browse and import external data from public databases



The screenshot shows the UCSC Table Browser interface. The top navigation bar includes 'Genomes', 'Genome Browser', 'Tools', 'Mirrors', 'Downloads', 'My Data', 'Help', and 'About Us'. The main content area is titled 'Table Browser' and contains a detailed description of the tool's purpose. Below the description are several control fields: 'clade' (Mammal), 'genome' (Mouse), 'assembly' (Dec. 2011 (GRCm38/mm10)), 'group' (Genes and Gene Predictions), 'track' (UCSC Genes), 'table' (knownGene), 'region' (genome/position), 'identifiers (names/accessions)', 'filter', 'intersection', 'correlation', 'output format' (BED - browser extensible data), 'output file', and 'file type returned' (plain text). At the bottom, there is a section titled 'Using the Table Browser' with a brief description and a list of parameters: 'clade' and 'genome'.

Hands On

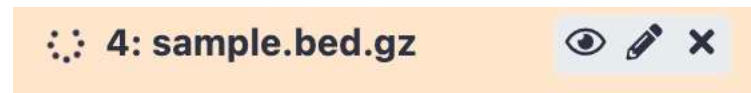
Exercise 3

Datasets/Jobs in the History

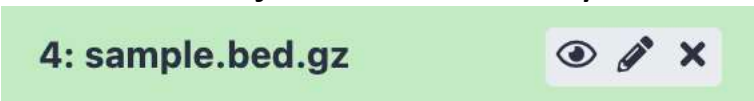
Grey: the job is waiting to run



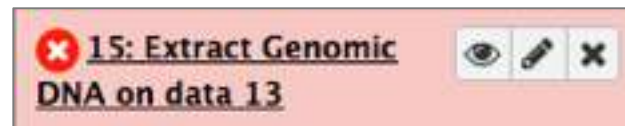
Orange: the job is running



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

Genome

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

The screenshot shows a file upload interface for a BED file. At the top, the file name is "4: sample.bed.gz" with icons for view, edit, and close. Below the name, it displays "32,561 regions" and "format: bed, génome de référence: mm10". A text input field contains "uploaded bed file". Below this are icons for save, share, info, chart, and help, along with a right arrow and a chat icon. Three options are listed: "display in IGB View", "display with IGV local", and "display at UCSC main test". At the bottom, a table shows the first five lines of the 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	chr1-5376854

Datasets/Jobs in the History

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

Download dataset

4: sample.bed.gz

32,561 regions
format: **bed**, génome de référence: **mm10**

uploaded bed file

display in IGB View
display with IGV local
display at UCSC main test

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	chr1-5376854

Delete dataset

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

Size of histories and quota

The screenshot displays a software interface with a dark header bar at the top right containing a green progress indicator labeled "Using 61%". Below this is a "History" section with a search bar containing the text "Rechercher des données" and a refresh icon. The main content area is titled "Unnamed history" and shows "4 shown" items. The first item is highlighted in green and shows a size of "56.26 MB". Below this are four more items, each with a green background and a list of icons (eye, pencil, and X). The items are numbered 1 through 4 from bottom to top.

Using 61%

History

Rechercher des données

Unnamed history

4 shown

56.26 MB

4: sample.bed.gz

3: CRN-107_R1.fastq.gz

2: CRN-107_11-R2.fastq.g
z

1: CRN-107 11-R1.fastq.g

Size of history

Quota

Hands On

Exercise 4

Tool Panel / Run analyses

The screenshot displays the Galaxy France web interface. On the left, a sidebar titled "Tools" is highlighted with an orange border. It contains a search bar, an "Upload Data" button, and a list of tool 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", "BEI", "VC", "Na", and "COMMON C... TOOLS". An orange arrow points from the text "Tool panel" to the sidebar. The main workspace features a header with navigation links (Workflow, Visualize, Données partagées, Aide, Utilisateur) and a user profile icon. Below the header is a Christmas-themed graphic with a red hat and a bar chart. The main content area displays a "Welcome to usegalaxy.fr" message, a "Term Of Use" link, and a news item dated 22/07/2021 about Galaxy release 21.05. At the bottom, there is a "Ask the GalaxyCat" section with a cat icon and a search prompt. On the right, a "History" panel shows an empty "Unnamed history" and a message in French: "Cet historique est vide. You can Charger vos propres données or Charger des données depuis une source externe".

Tool Panel / Run analyses

The image shows a screenshot of a web-based tool panel. The panel is titled "Tools" and includes a search bar labeled "search tools" and an "Upload Data" button. Below these are several tool categories: "Get Data", "Send Data", "Collection Operations", "GENERAL TEXT TOOLS", "Text Manipulation", "Filter and Sort", "Join, Subtract and Group", "GENOMIC FILE MANIPULATION", "Convert Formats", and "FASTA/FASTQ".

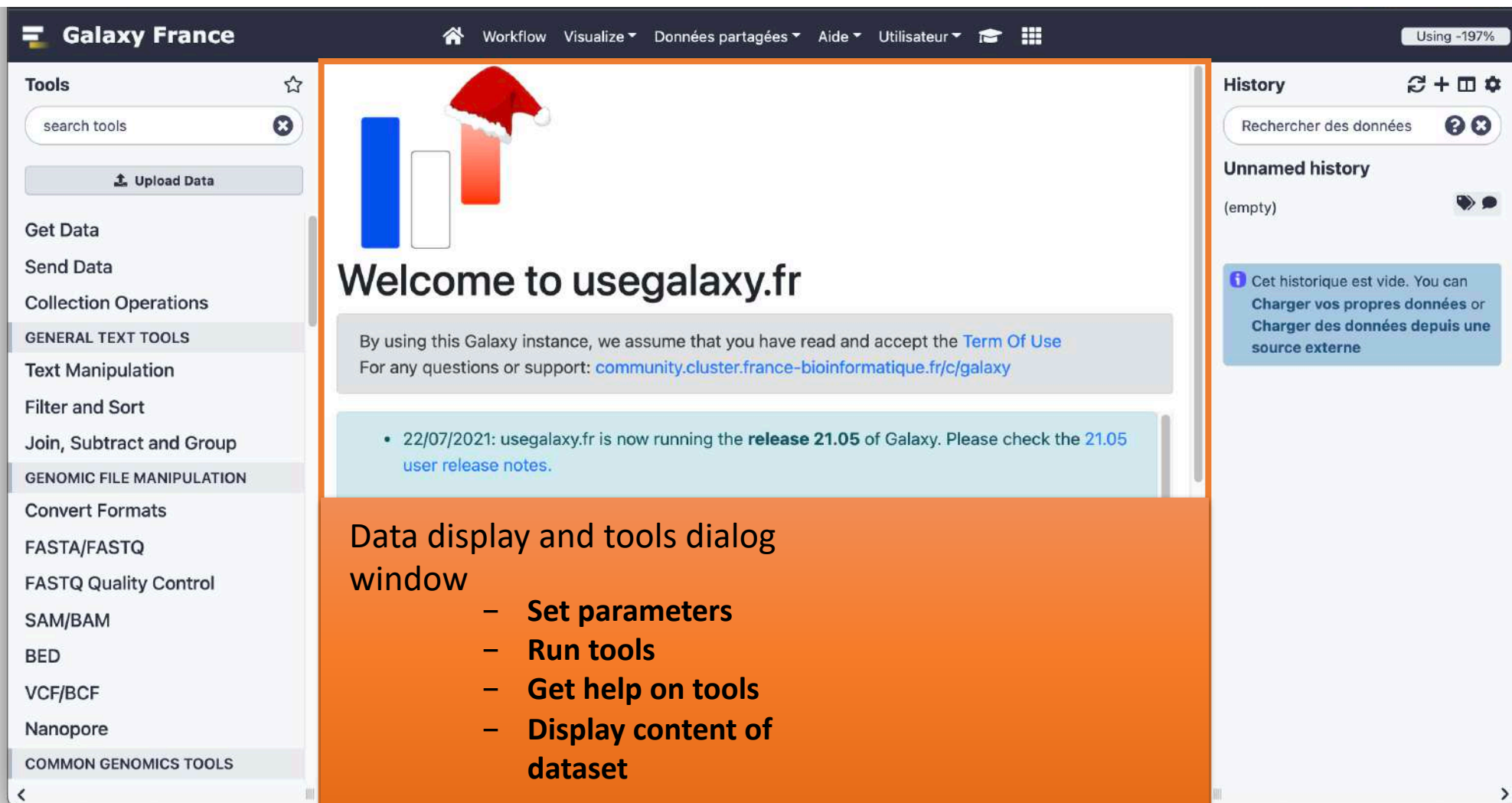
Annotations with orange lines and arrows point to various elements:

- "Search a tool" points to the search bar.
- "Tool category" points to the "Text Manipulation" category in the list.
- "Upload data" points to the "Upload Data" button.
- "Tool" points to the "Query Tabular" tool in the detailed view.

A large orange arrow points from the "Text Manipulation" category in the tool panel to a detailed view of that category on the right. This detailed view lists the following tools:

- 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 displays the Galaxy France web interface. The top navigation bar includes 'Galaxy France', 'Workflow', 'Visualize', 'Données partagées', 'Aide', 'Utilisateur', and a user profile icon. The left sidebar contains a 'Tools' section with a search bar and an 'Upload Data' button, followed by categories like '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'. The main content area features a 'Welcome to usegalaxy.fr' message with a Santa hat icon, a 'Term Of Use' link, and a release note for Galaxy 21.05. The right sidebar shows a 'History' section with a search bar and an 'Unnamed history' section that is currently empty. An orange box highlights the 'Tools dialog window' area, which is used for setting parameters, running tools, getting help, and displaying dataset content.

Tools dialog window

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

Hands On

Exercise 5

Hands On

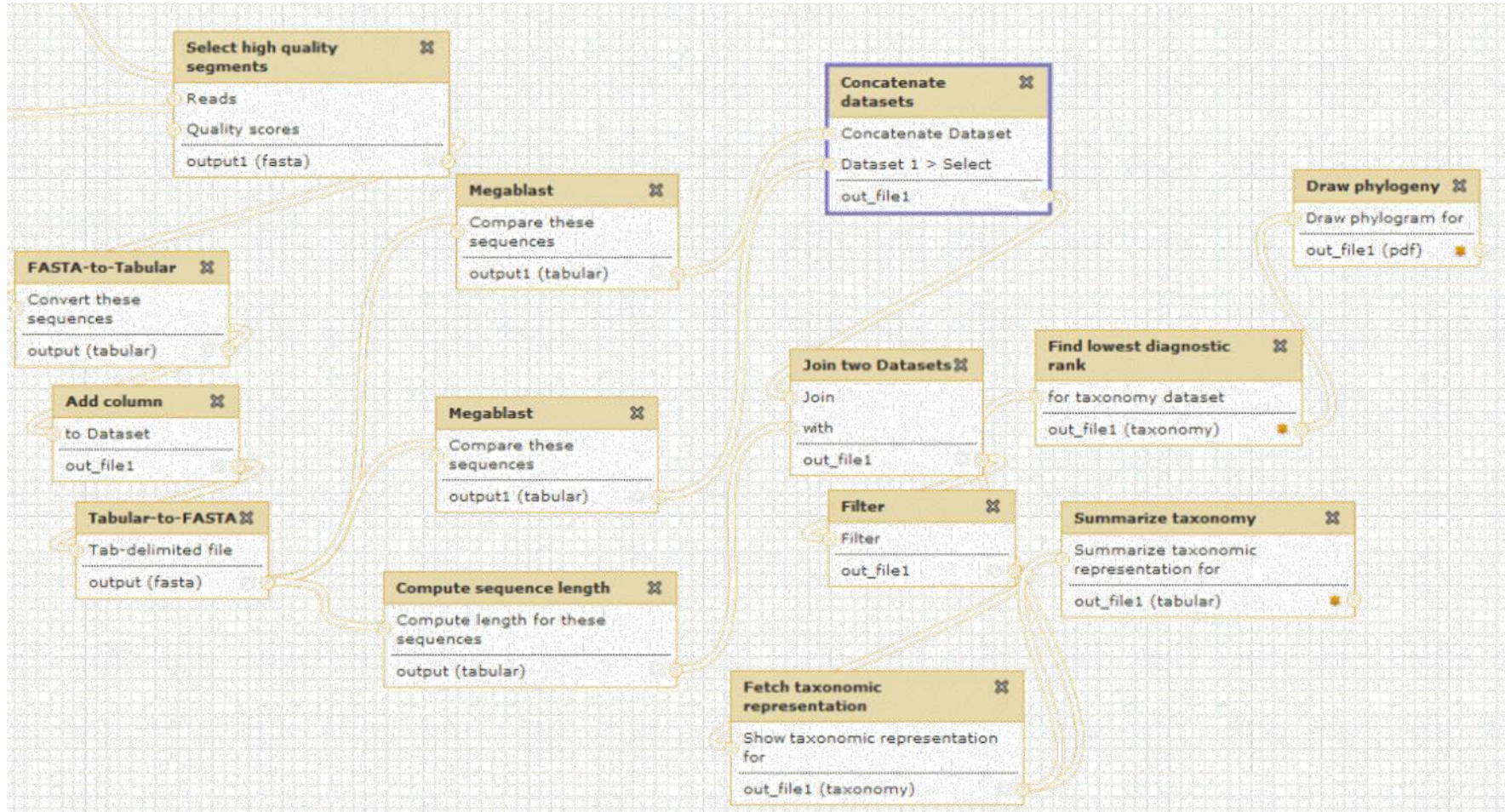
Exercise 6

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 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 displays the Galaxy France web interface. The top navigation bar includes 'Workflow', 'Visualize', 'Données partagées', 'Aide', 'Utilisateur', and a user profile icon. The left sidebar contains a 'Tools' section with a search bar and an 'Upload Data' button, followed by various tool categories like 'Visualization', 'Get Features/Annotations', 'Functional Analysis', 'Pathway Analysis', 'GALAXY-P', 'STATISTICS AND VISUALISATION', 'Statistics', 'Graph/Display Data', 'MISCELLANEOUS TOOLS', 'RNA-seq', 'Get Data', 'Motif', 'Interactive Tools', and 'WORKFLOWS'. The 'WORKFLOWS' category is highlighted with an orange arrow pointing to a box labeled 'Run workflows'. The main content area shows a workflow titled 'Query Tabular using sqlite sql (Galaxy Version: 3.3.0)'. An orange arrow points from this workflow title to a box labeled 'Create, run, edit (...) workflows'. Below the workflow title are options to 'Add tables to an existing database', 'Database Table', 'Modify the database', and 'Save the sqlite database in your history'. The 'Save the sqlite database in your history' option is currently set to 'No'. There is a text input field for 'SQL Query to generate tabular output' and a dropdown menu for 'include query result column headers' set to 'Yes'. The right sidebar shows a 'History' section with a search bar and a message: 'Cet historique est vide. You can Charger vos propres données or Charger des données depuis une source externe'. The bottom right corner of the interface shows the page number '51'.

Run workflows

Create, run, edit (...) workflows

Workflows

Create workflows

The screenshot displays the Galaxy France interface. The top navigation bar includes 'Galaxy France', a home icon, and menu items for 'Workflow', 'Visualize', 'Données partagées', 'Aide', 'Utilisateur', and a user profile icon. The right side of the top bar shows 'Using -197%'. The left sidebar contains a 'Tools' section with a search bar and an 'Upload Data' button, followed by various tool categories like 'Visualization', 'Functional Analysis', and 'WORKFLOWS'. The main area shows a 'List of Workflows' table with columns for Name, Tags, Updated, Sharing, and Bookmarked. The right sidebar shows a 'History' section with a search bar and a message indicating an empty history.

Name	Tags	Updated	Sharing	Bookmarked
▼ workflowAtacseq	▶	a year ago		<input type="checkbox"/>
▼ ChipSeq workflow	▶	a year ago		<input type="checkbox"/>
▼ Workflow constructed from history 'Unnamed history'	▶	a year ago		<input type="checkbox"/>
▼ Demo Galaxy	▶	a year ago		<input type="checkbox"/>
▼ Demo	▶	a year ago		<input type="checkbox"/>
▼ Demo 071113	▶	a year ago		<input type="checkbox"/>
▼ Debug Workflow	▶	a year ago		<input type="checkbox"/>
▼ Motif detection	▶	a year ago		<input type="checkbox"/>
▼ Repeat Analysis for Joanna	▶	a year ago		<input type="checkbox"/>
▼ chIP-seq Workflow	▶	a year ago		<input type="checkbox"/>
▼ chip-seq	▶	a year ago		<input type="checkbox"/>

Workflow creation

The screenshot displays the Galaxy France workflow creation interface. At the top, the navigation bar includes the Galaxy France logo, a home icon, and menu items for 'Workflow', 'Visualize', 'Données partagées', 'Aide', 'Utilisateur', and a user profile icon. A 'Using -197%' indicator is visible in the top right corner.

The main workspace is titled 'Unnamed workflow' and features a large grid for building the workflow. On the left, a 'Tools' sidebar contains a search bar and several categories: 'Inputs', 'Get Data', 'Send Data', 'Collection Operations', 'Expression Tools', 'GENERAL TEXT TOOLS' (highlighted with an orange arrow), 'Text Manipulation', 'Filter and Sort', 'Join, Subtract and Group', 'GENOMIC FILE MANIPULATION', 'Convert Formats', 'FASTA/FASTQ', 'FASTQ Quality Control', 'SAM/BAM', 'BED', and 'VCF/BCF'. An orange callout box at the bottom left contains the text: 'Add tools or input datasets to the workflow'.

On the right, a metadata panel for the workflow includes fields for 'Name' (set to 'Unnamed workflow'), 'Version' (set to '1: Dec 14th 2021, 0 steps'), and an 'Annotation' text area. Below these are sections for 'License' (with a prompt to specify a license), 'Creator' (with a prompt to add a creator), and 'Tags' (with a prompt to apply tags for searchability). A zoom level of '100%' and a '+' button are visible at the bottom of the grid.

Add tools or input datasets to the workflow

Workflow creation

Galaxy France

Workflow Visualize Données partagées Aide Utilisateur Using -197%

Tools

search tools

Inputs

- Input dataset
- Input dataset collection
- Simple inputs used for workflow logic

Get Data

Send Data

Collection Operations

Expression Tools

GENERAL TEXT TOOLS

Text Manipulation

Filter and Sort

- Filter data on any column using simple expressions
- Sort data in ascending or descending order
- Select lines that match an expression

GFF

Unnamed workflow

Input Dataset

output (input)

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

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

Galaxy France

Workflow Visualize Données partagées Aide Utilisateur Using -197%

Tools

search tools

Inputs

- Input dataset
- Input dataset collection
- Simple inputs used for workflow logic

Get Data

Send Data

Collection Operations

Expression Tools

GENERAL TEXT TOOLS

Text Manipulation

Filter and Sort

- Filter data on any column using simple expressions
- Sort data in ascending or descending order
- Select lines that match an expression
- GFF
- Extract features from GFF

Unnamed workflow

Input Dataset

output (input)

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

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

Workflow creation

The screenshot displays the Galaxy France workflow creation interface. The main canvas shows an 'Unnamed workflow' with two tools connected: 'Input Dataset' and 'Filter'. The 'Filter' tool is pre-configured with the condition 'c1=='chr22''. The right sidebar shows the configuration options for the 'Filter' tool, including a search bar, a label field, a step annotation field, and a filter condition field. The 'Filter' tool is selected, and its configuration options are visible in the sidebar.

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

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

Workflow creation

Save workflow

Run

The screenshot displays the Galaxy France workflow editor. The main workspace shows a workflow named "Unnamed workflow" with two steps: "Input Dataset" and "Filter". The "Filter" step is selected, and its configuration panel is open on the right. The configuration panel includes fields for "Label", "Step Annotation", and "Filter" (with a checkbox for "With following condition" and a text input field containing "c1=='chr22"). The top navigation bar includes "Galaxy France", "Workflow", "Visualize", "Données partagées", "Aide", "Utilisateur", and "Using 197%". The left sidebar contains a search bar and various tool categories like "Inputs", "Get Data", "Send Data", "Collection Operations", "Expression Tools", "GENERAL TEXT TOOLS", "Text Manipulation", "Filter and Sort", and "Extract features from GFF data".

Run workflows

Galaxy France

Workflow Visualize Données partagées Aide U

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

Workflow: Unnamed workflow

History Options

Send results to a new history

No

1: Input dataset

4: sample.bed.gz

2: Filter (Galaxy Version 1.1.1)

Filter

Output dataset 'output' from step 1

With following condition

c1==chr22

Number of header lines to skip

0

Run Workflow

History

Rechercher des données

Unnamed history

4 shown

56.26 MB

4: sample.bed.gz

3: CRN-107_R1.fastq.gz

2: CRN-107_11-R2.fastq.gz

1: CRN-107_11-R1.fastq.gz

Using -197%

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

Run workflow

Set 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