

GABIOMATIC: Bioinformatics in Galaxy

Lambaréné, 2024-04-19

Fabian Amman

Institute for Hygiene and Applied Immunology
Center for Pathophysiology, Infectiology and Immunology
Medical University of Vienna

What is Galaxy

- Data Intensive analysis for everyone without the need to master the command line
- Web-based (galaxyproject.org)
- Easy to use
- Free and Open Source
- Many tools (~9400 in the Galaxy Tool Shed)
- Popular (>11.900 publications)

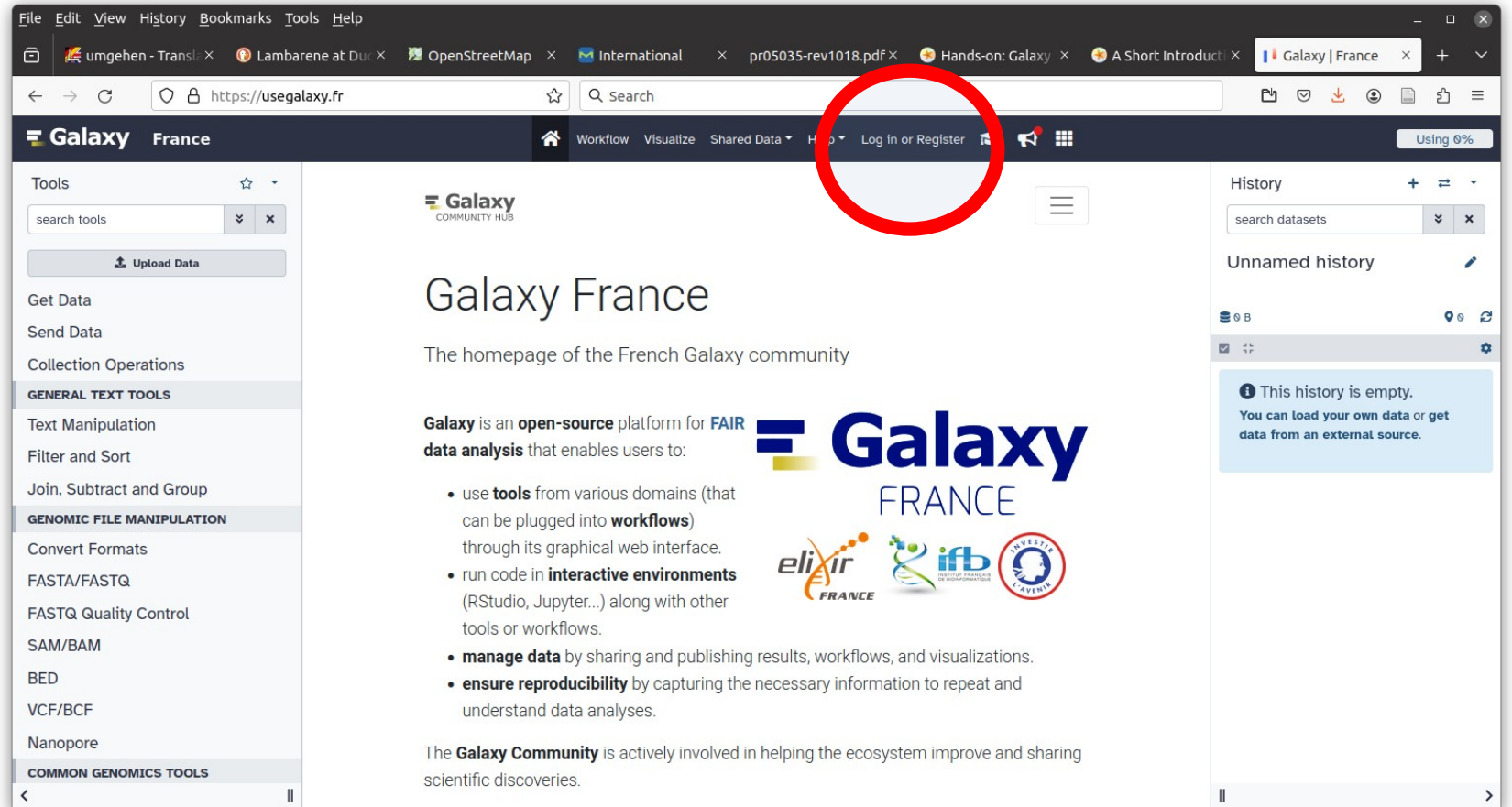
Find a Galaxy Server

- Several large open server available:
 - Galaxy Europe (UseGalaxy.eu)
 - Galaxy Main (UseGalaxy.org)
 - Galaxy Australia (UseGalaxy.org.au)
 - Galaxy France (UseGalaxy.fr)
- Many other smaller, often domain-specific Galaxies available
 - e.g., <https://ecology.usegalaxy.eu/>
- List of all public Galaxies (135+)
 - galaxyproject.org/use



Create an account

- Go to any of the large server provider, e.g.,
 - Galaxy France (UseGalaxy.fr)
 - Galaxy Europe (UseGalaxy.eu)
- Create an account
- Confirm via specified e-mail
- Log into your preferred server



The screenshot shows the Galaxy France homepage. The browser address bar displays <https://usegalaxy.fr>. The navigation bar includes 'Workflow', 'Visualize', 'Shared Data', and 'Log in or Register'. The 'Log in or Register' button is circled in red. The main content area features the 'Galaxy France' logo and the text 'The homepage of the French Galaxy community'. Below this, it states 'Galaxy is an open-source platform for FAIR data analysis that enables users to:' followed by a list of capabilities: using tools from various domains, running code in interactive environments, managing data, and ensuring reproducibility. Logos for Elixir France, IFB, and Université Clermont Auvergne are displayed. A sidebar on the left lists various tool categories like 'GENERAL TEXT TOOLS' and 'GENOMIC FILE MANIPULATION'. A right sidebar shows an empty history section with a message: 'This history is empty. You can load your own data or get data from an external source.'


The Workspace

- Go to any of the large server provider, e.g.,
 - Galaxy France (UseGalaxy.fr)
 - Galaxy Europe (UseGalaxy.eu)
- Create an account
- Confirm via specified e-mail
- Log into your preferred server

The screenshot displays the Galaxy web interface at usegalaxy.org. The interface is divided into three main sections:

- Tools Panel (left, green border):** Contains a search bar, an 'Upload Data' button, and a list of tool categories including 'Get Data', 'Send Data', 'Collection Operations', 'GENERAL TEXT TOOLS' (with sub-items like Text Manipulation, Filter and Sort, etc.), 'GENOMIC FILE MANIPULATION' (with sub-items like FASTA/FASTQ, FASTQ Quality Control, etc.), and 'Convert Formats'. The word 'TOOLS' is overlaid in large green text at the bottom of this panel.
- Center Pane (middle, purple border):** Displays a welcome message: 'Galaxy is an open source, web-based platform for data intensive biomedical research. If you are new to Galaxy start here or consult our help resources. You can install your own Galaxy by following the tutorial and choose from thousands of tools from the Tool Shed.' Below the text is a video player showing a woman speaking, with a quote: 'The most important job of senior faculty is to mentor junior faculty and students.' — @jxtx. The word 'CENTER PANE' is overlaid in large purple text at the bottom of this section.
- History Panel (right, red border):** Shows a list of recent datasets and workflows. The top entry is 'Galaxy 101' (8.94 MB). Below it is a list of workflows: '7: Compare two Datasets on data 6 and data 1', '6: Select first on data 5', '5: Sort on data 4', '4: Datamash on data 3', '3: bedtools Intersect intervals on data 2 and data 1', '2: SNPs', and '1: Exons'. The word 'HISTORY' is overlaid in large red text at the bottom of this panel.

Let's use Galaxy.eu

- Create and name new History
 - Title: "GABIOMATIC SC2 WORKFLOW"
 - Annotation: "Example of Galaxy workflow used in the GABIOMATIC Bioinformatics Workshop in Lambarene/Gabon"
- Go to "Upload Data"
 - upload barcode02.fastq (specifiy that it is a fastq file)
 - upload MN908947.3_reference_genome.fasta
- Add tool minimap2
 - run on reference fasta and read fastq files
- Add tool ivar consensus
 - run on minimap2 bam file
- Add tool pangolin
 - run on ivar consensus fasta file
- Go to  right to History
 - Export Tools citation

Tool	History items created
Data Fetch <i>This tool cannot be used in workflows</i>	2 MN908947.3_reference_genome.fasta <input checked="" type="checkbox"/> Treat as input dataset MN908947.3_reference_genome
Data Fetch <i>This tool cannot be used in workflows</i>	5 barcode02.fastq <input checked="" type="checkbox"/> Treat as input dataset barcode02.fastq
Map with minimap2 <input checked="" type="checkbox"/> Include "Map with minimap2" in workflow	6 Map with minimap2 on data 5 and data 2 (mapped reads in BAM format)
ivar consensus <input checked="" type="checkbox"/> Include "ivar consensus" in workflow	7 ivar consensus on data 6 Consensus
Pangolin <input checked="" type="checkbox"/> Include "Pangolin" in workflow	8 pangolin on data 7

Summary

- Galaxy lets you perform complex data analysis right from your browser
 - It enables access to computing infrastructure
 - Circumvent the need to maintain local computing infrastructure
- Galaxy allows for public, collaborative, and private data and workflow handling
- These analyses are reproducible
 - Galaxy keeps track of all the details of your analysis
 - Galaxy allows for easy referencing of used tools
- Galaxy has a very large and active user community
 - <https://help.galaxyproject.org/>
- There are many training materials available to learn Galaxy, facilitating making first relevant analysis
 - <https://training.galaxyproject.org/training-material/>