Vall d'Hebron
Institut de Recerca

DATA FORMATS IN NGS
INTRODUCTION TO GALAXY

Bioinformàtica per a la Recerca Biomèdica

Mireia Ferrer[1], Álex Sánchez[1,2]
Esther Camacho[1], Angel Blanco[1,2]
1 Unitat d'Estadística i Bioinformàtica (UEB) VHIR
2 Departament de Genètica, Microbiologia i Estadística, UB

Vall d'Hebron Hospital · Generalitat de Catalunya Departament de Salut · Institut Català de la Salut · UAB Universitat Autònoma de Barcelona · Generalitat de Catalunya Departament d'Economia i Coneixement · Institució CERCA Centres de Recerca de Catalunya · Instituto de Salud Carlos III

Vall d'Hebron
Institut de Recerca

1. Data formats used in NGS
2. Introduction to Galaxy

2. Introduction to Galaxy

Vall d'Hebron
Institut de Recerca

- An open, web-based platform integrating many popular tools and resources for intensive biomedical research.
- **What can be done?**
  - Obtain data from many data sources like UCSC Table Browser, Biomart, WormBase, or your own data
  - Prepare data for further analysis by rearranging or cutting data columns, filtering data and many other options
  - Analyze data by finding overlapping regions, determining statistics, preprocessing NGS data and much more
  - Share data and workflows

2. Introduction to Galaxy

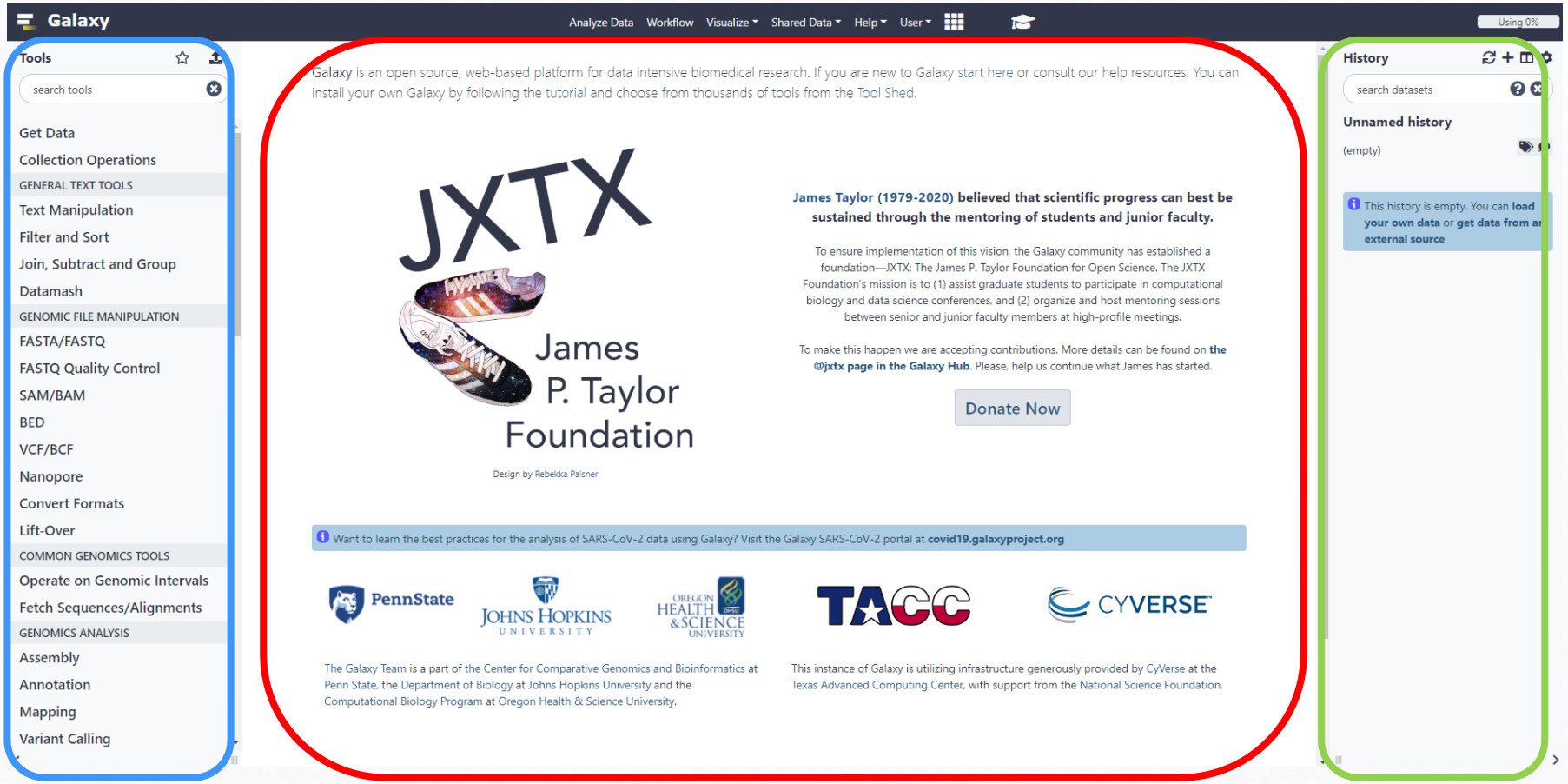
Vall d'Hebron Institut de Recerca

The Galaxy page is divided into three panels:

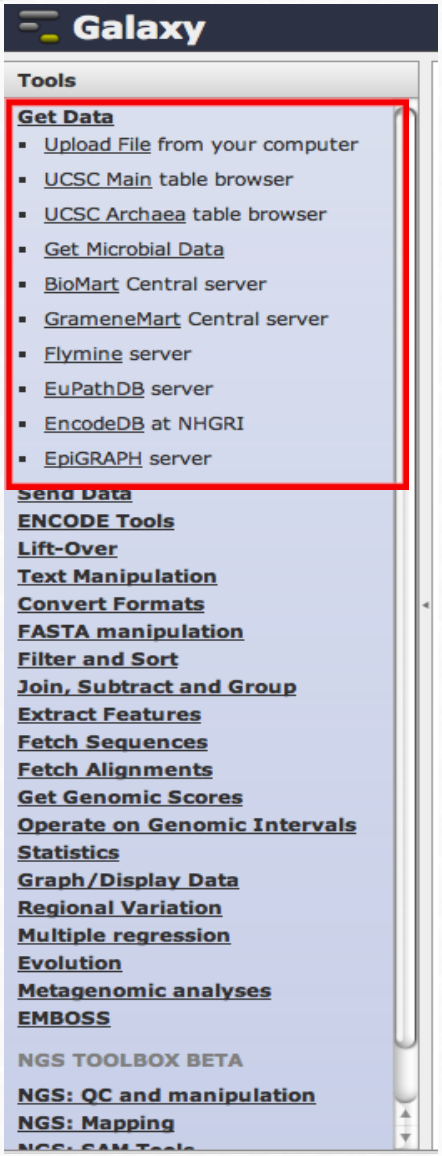
Tools for uploading, processing and analysis

Viewing panel (menus, data, results)

History of analysis steps and datasets



2. Introduction to Galaxy



Tools for data analysis

Get Data

- From databases (UCSC Table Browser, ...)
- From uploaded files
- From urls

Text manipulation

Filter and Sort

Operate on Genomic Intervals

FASTA manipulation

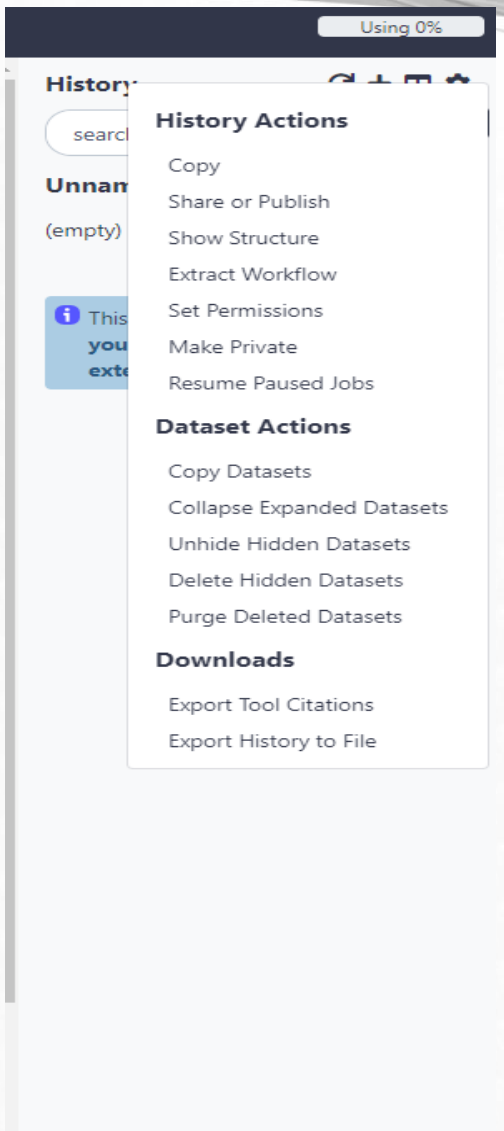
NGS analysis

- QC
- Fastq file pre-processing
- Read Alignment / Mapping
- SAM tools

2. Introduction to Galaxy

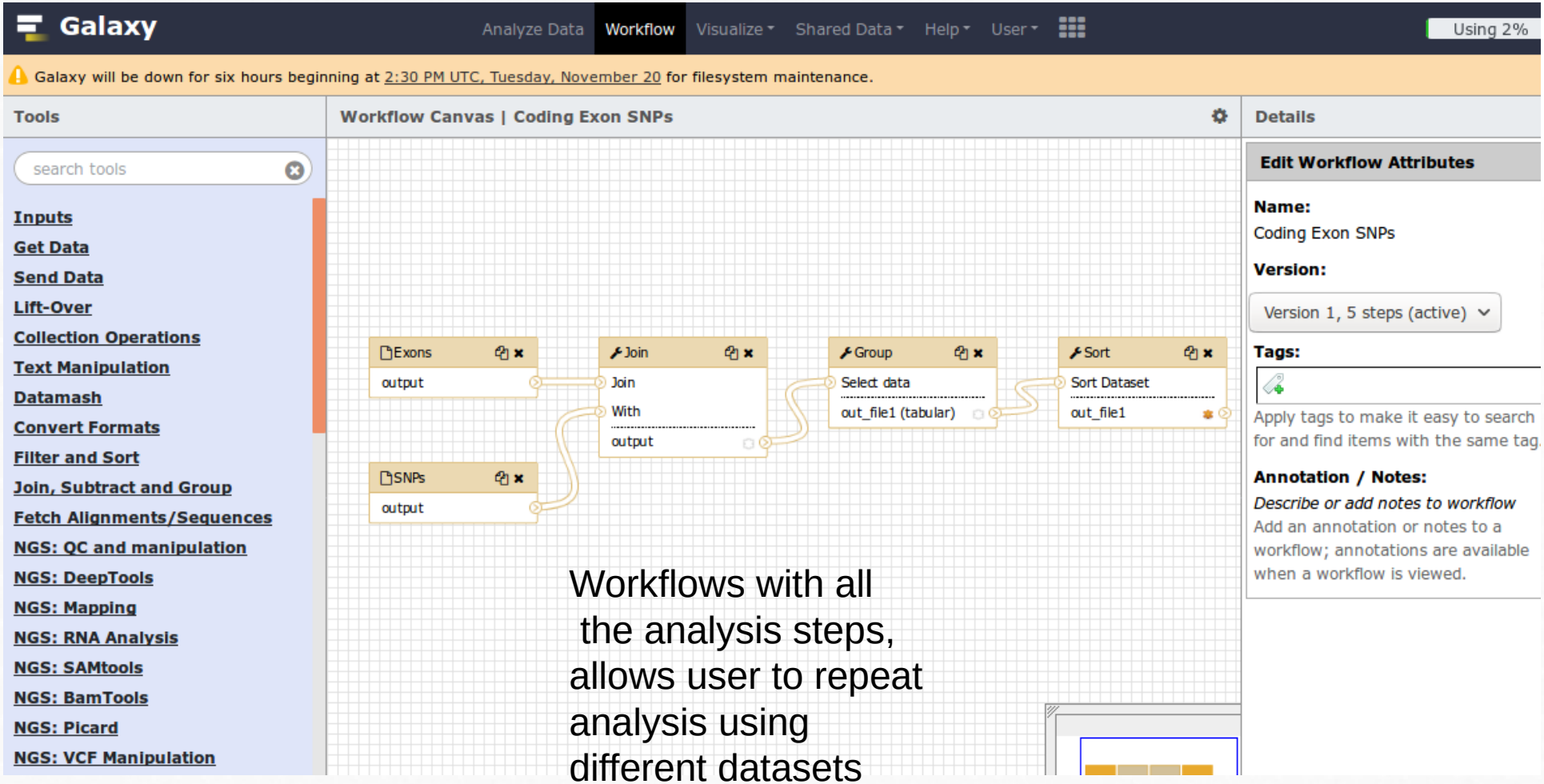
Histories

List saved histories and shared histories.
Work on Current History, create new, clone, share, create workflow, set permissions, show deleted datasets or delete history.



Vall d'Hebron Institut de Recerca

Workflows



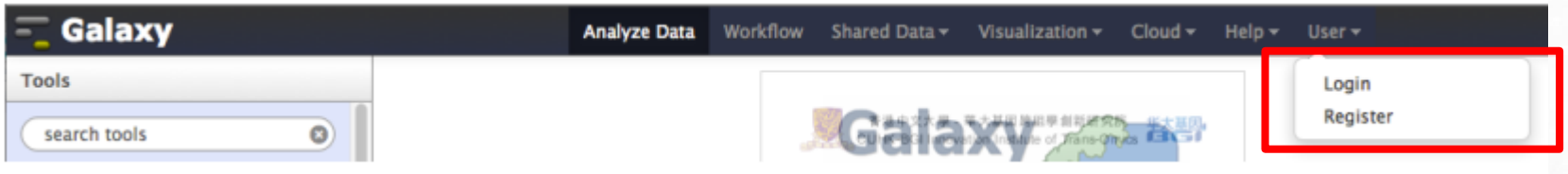
Workflows with all the analysis steps, allows user to repeat analysis using different datasets

2. Introduction to Galaxy

Vall d'Hebron
Institut de Recerca

Register for a Galaxy account

This will only take a moment, and will allow all the work that you do to persist between sessions and allow you to name, save, share, and publish Galaxy histories, workflows, datasets and pages. It allows you to store up to 250GB of data on this public server.



https://usegalaxy.eu/

2. Introduction to Galaxy

Vall d'Hebron
Institut de Recerca

Training Infrastructure as a Service

We want to help you conduct your training seminars. You provide the training, we provide you training infrastructure *at no cost*.

Why use UseGalaxy.eu training infrastructure?

- Free
- Private queue, no wait times
- No Galaxy Maintenance
- No Galaxy Administration
- Official Galaxy Training Materials guaranteed to work

TIaaS

Simply fill out the infrastructure request form and we'll get back to you shortly.

Find out more

After registration in European Galaxy server

https://usegalaxy.eu/join-training/ueb_bi2021

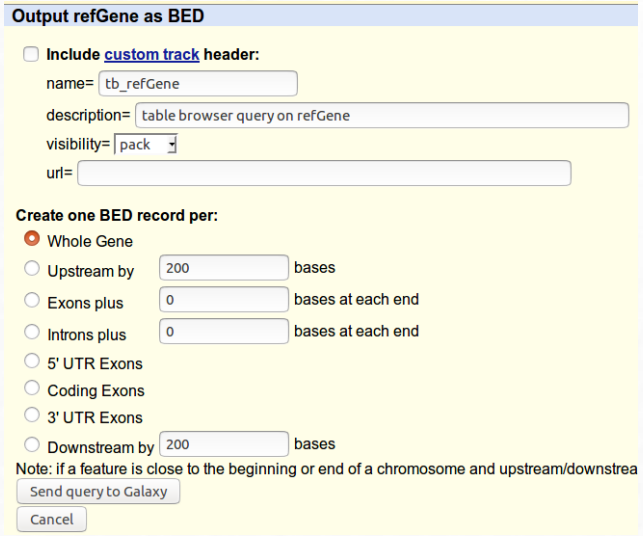
Vall d'Hebron Institut de Recerca

Importing data into Galaxy

1. From database queries (eg. UCSC): obtain a BED-formatted dataset of all RefSeq genes from platypus.

   Get Data > UCSC Main – Table Browser tool

   Set genome, RefSeg Genes, and BED output format (send to Galaxy)

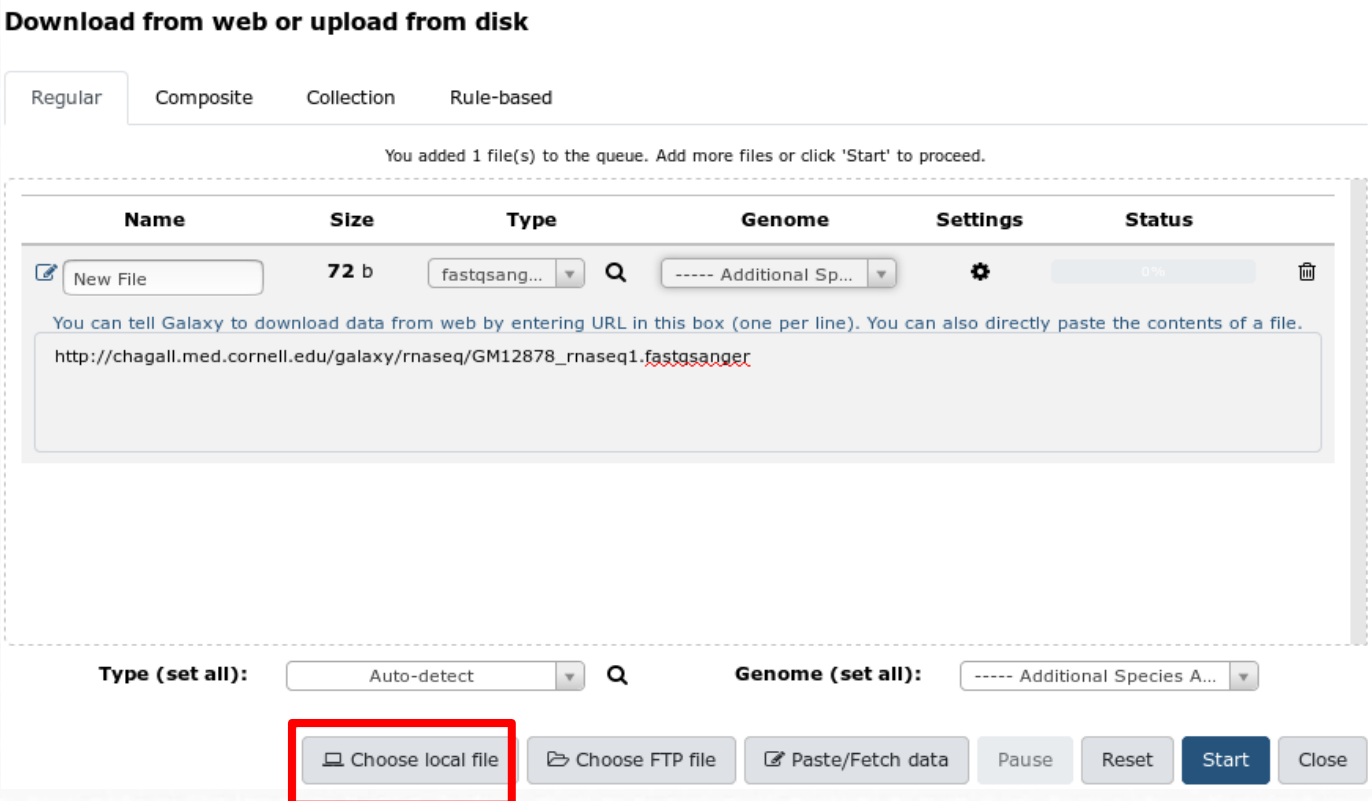


Vall d'Hebron
Institut de Recerca

Importing data into Galaxy

2. From a File on your computer / FTP file:

Get Data > Upload File



Vall d'Hebron
Institut de Recerca

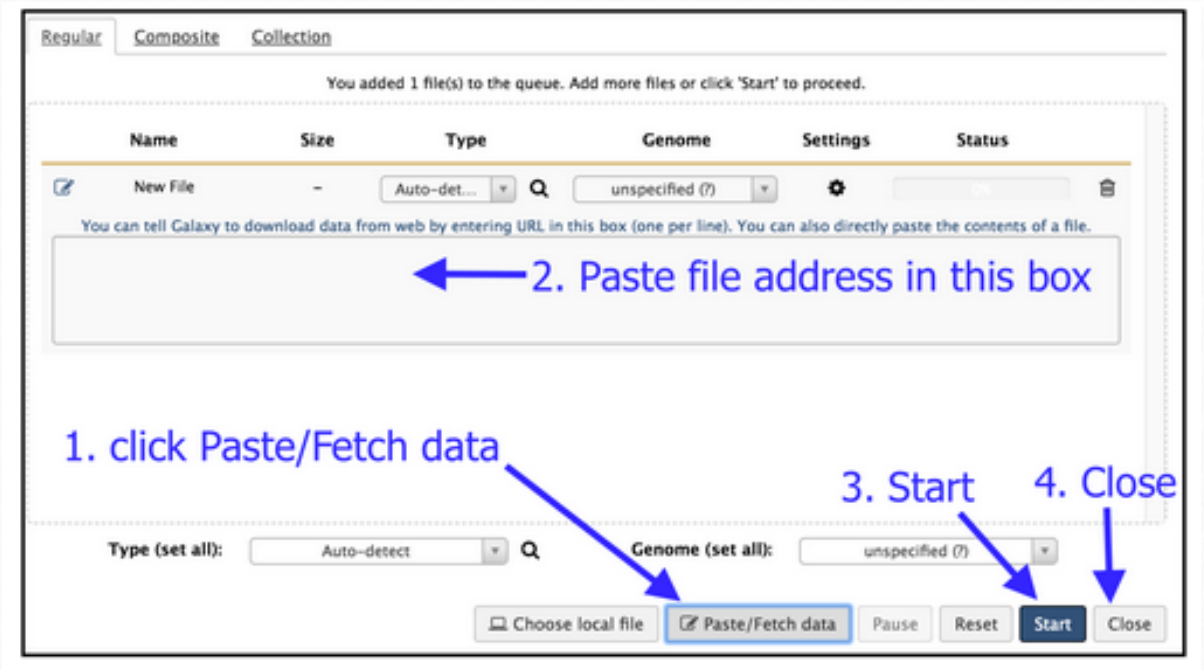
Importing data into Galaxy

3. From a website:

Get Data > Upload File

Copy this URL into the text-entry box:

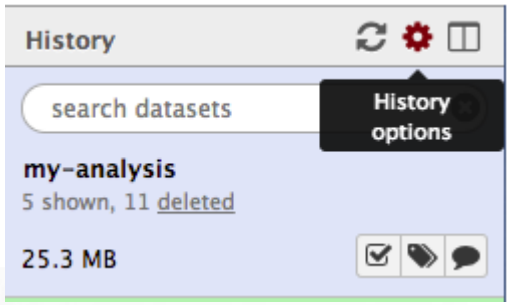
url: https://zenodo.org/record/582600/files/mutant_R1.fastq



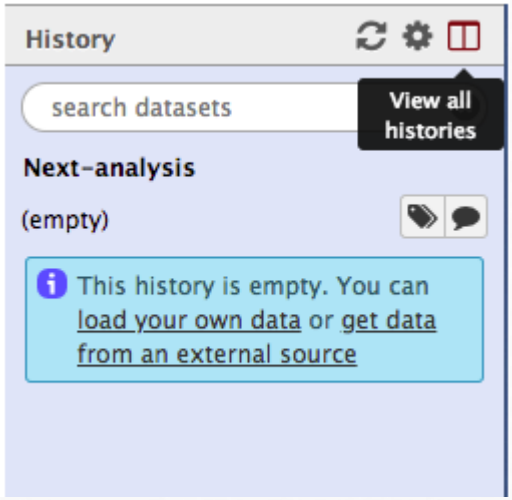
Vall d'Hebron Institut de Recerca

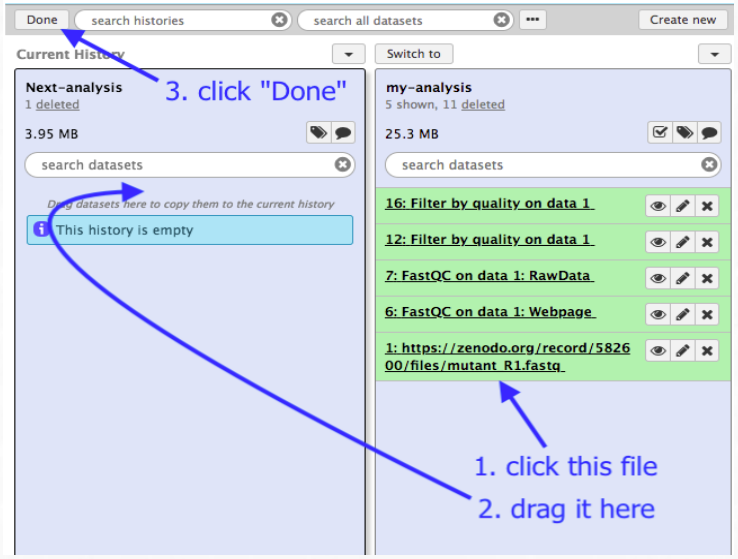
Managing histories

- Name your current history
- Create new history and rename it



- Manage datasets and histories:
- View all histories
- Drag files between histories (new history must be set to current)





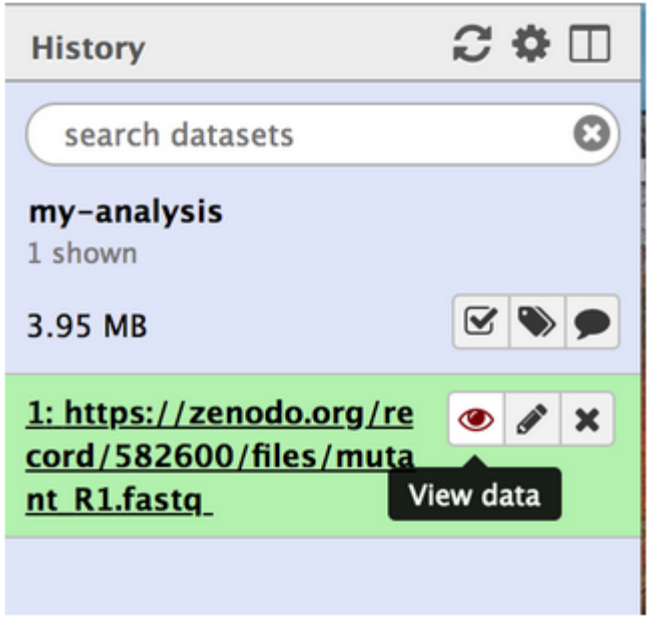
Vall d'Hebron Institut de Recerca

Visualizing the dataset

- You can view file content clicking the eye icon in history.

The mutant_R1.fastq file contains DNA sequencing reads from a bacteria, in FASTQ format:

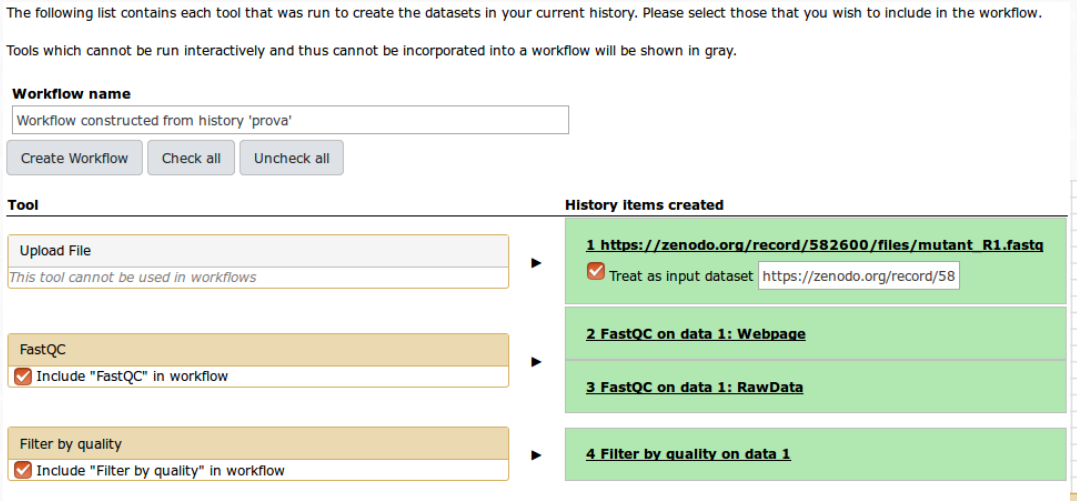
```
@mutant-no_snps.gff-24960/1                                read 1 sequence
AATGTTGTCACTTGGATTCAAATGACATTTTAAATCTAATTATTCATGAATCGAACTAGTACGAAATGCAATGAG
+                                                          read 1 quality scores
5??A9?BBBDDDBEDDBFF+FGHHIIHHHEIHIIHIIAHDHIIHIG#IIHIFHHHFGIII*IHHHIHFIIHGICI
@mutant-no_snps.gff-24958/1
CAAAGTCGTTGGTCATATAAAAAACCGCGTACAGTCAACTATAGATACAATCAAGATAAACTCATGCACAGATTG
+
?A????@?DDDABDE9FGGGFGICFHIIIBGHIIIGICHHIFH=IHAFIHIHHHHIFCIIEIHAIFGIHIDDIHE
@mutant-no_snps.gff-24956/1
TATAAATTCAACTTTGCAACAGAACCATCTAATCTTCAACAAACTGGCCCGTTTGTTGAACTACTCTTTAATAAA
+
?????BBADD5DDDDDGFGCFFEECFBBICIII,IIHIICHIHIIFHHHHHIHIIIIIIIAHHIHHH5FHDHHHH
```

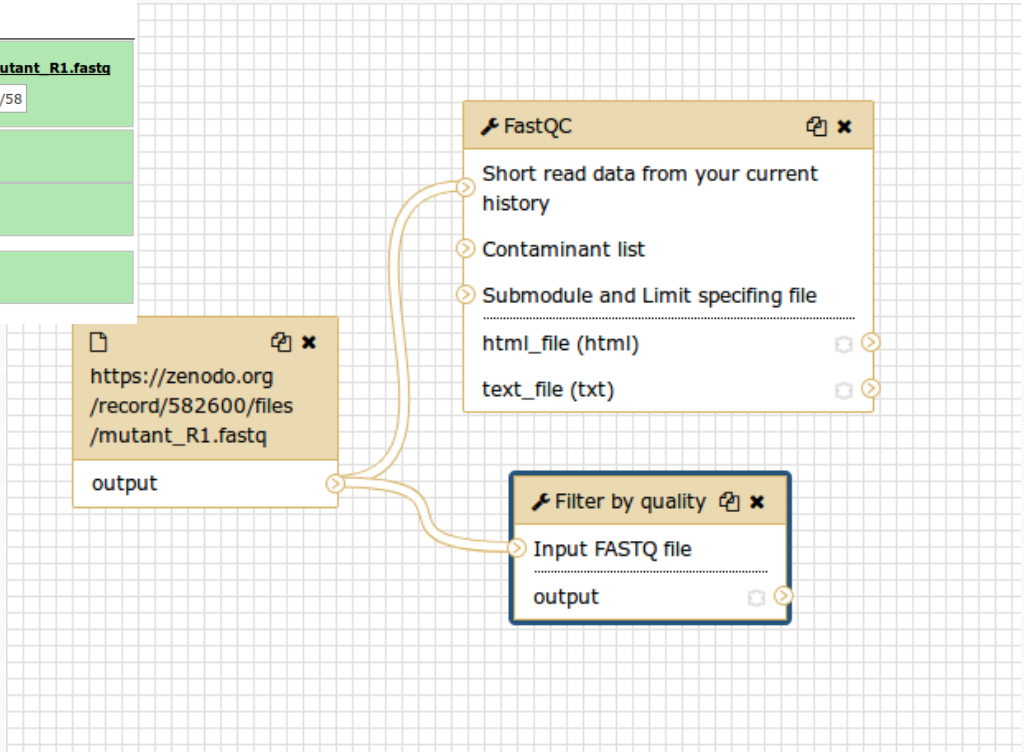


Vall d'Hebron Institut de Recerca

Create workflow from history

- From history options: Export workflow





2. Introduction to Galaxy

Vall d'Hebron
Institut de Recerca

- https://galaxyproject.org/learn/

Learn Galaxy

There are many approaches to learning how to *use* Galaxy. The most popular is probably to just dive in and use it. Galaxy is simple enough to use that you can do many analyses just by exploring the interface. However, you may miss much of the power this way.

Have you created or know of a resource that is useful for teaching with Galaxy? Then please share it! This will help others and also help get the word out about your resource. Use this Google form to describe your resource. **Also**: consider joining Galaxy Training Network and contributing your tutorial as described here!

Tutorials by Galaxy Training Network

Thanks to a large group of wonderful contributors there is a constantly growing set of tutorials maintained by the Galaxy Training Network. These include:

Introductory Tutorials

- Introduction to Galaxy Analyses
- Data Manipulation
- User Interface and Features

Scientific Analyses

- Assembly
- Computational chemistry
- Ecology
- Epigenetics
- Genome Annotation
- Imaging
- Metabolomics
- Metagenomics
- Proteomics
- Sequence analysis
- Statistics and machine learning
- Transcriptomics
- Variant Analysis

Material

Search

Lesson	Slides	Hands-on	Input dataset	Workflows	Galaxy tour	Galaxy instances
Introduction to metagenomics						
16S Microbial Analysis with mothur (extended)						
16S Microbial Analysis with mothur (short)						
Analyses of metagenomics data - The global picture						
Antibiotic resistance detection nanopore plasmids						
Metatranscriptomics analysis using microbiome RNA-seq data metatranscriptomics						
Metatranscriptomics analysis using microbiome RNA-seq data (short) metatranscriptomics						