Galaxy

Galaxy is an open source, web-based platform for data intensive biomedical research. A universal **GUI** and **workflow manager** for lots of tools.
Official website: https://galaxyproject.org/main/

There are approximately **120 public servers**, **dozens** of private **academic servers** and even several **commercial servers**. Lots of which are highly specialised or modified as Galaxy project is completely **open access**.

Anyone can use one of public servers, with or without an account, but **Galaxy user accounts** are simple to create (email, password and go!). Advantages are: **increased data quotas, more resources** available**, parallel jobs** and **extended functionality** across sessions (i.e., **Histories**), such as **naming, saving, sharing,** and **publishing**.

Usegalaxy.org

The main Galaxy server at http://usegalaxy.org

- combines many **common tools** with **data sources**;
- is available since 2007 for **anyone** to analyze their data **free of charge**;
- provides **substantial CPU and disk space**, making it possible to analyze large datasets;
- supports **thousands of users** and **hundreds of thousands of jobs** per month (see Statistics);
- sustained by TACC hardware using allocation generously provided by the CyVerse project.

Notation: **History, Dataset, Tool panel, Job, Workflow, Visualization, Library**

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.

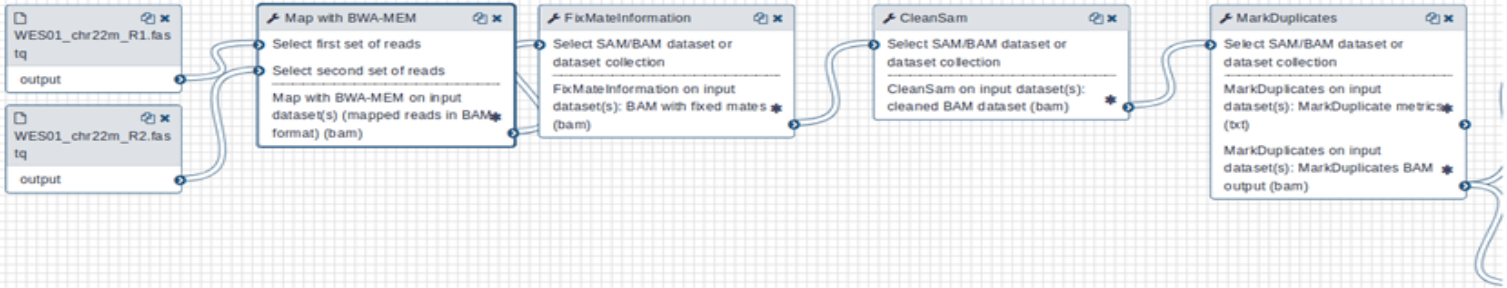
The Galaxy Team is a part of the Center for Comparative Genomics and Bioinformatics at Penn State, the Department of Biology and at Johns Hopkins University and the Computational Biology Program at Oregon Health & Science University.

This instance of Galaxy is utilizing infrastructure generously provided by the CyVerse at the Texas Advanced Computing Center, with support from the National Science Foundation.

The Galaxy Project is supported in part by NSF, NHGRI, The Huck Institutes of the Life Sciences, The Institute for CyberScience at Penn State, and Johns Hopkins University.

Galaxy version 17.09.rc1, commit 288dc48c386f41c858ae3bbf9971979296b5abbc

imported: 01-ngs-bwa-mem-markdupe-filter



90%

Details

Map with BWA-MEM - map
medium and long reads (> 100 bp) against reference genome (Galaxy Version 0.7.12.1)

Label

Add a step label.

Annotation

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

Will you select a reference genome from your history or use a built-in index?

Use a built-in genome index

Built-ins were indexed using default options. See `Indexes` section of help below

Using reference genome

Human (Homo sapiens) (b3...

Select genome from the list

Single or Paired-end reads

Paired

Select between paired and single end data

Select first set of reads

Data input 'fastq_input1' (fastqsanger)

Specify dataset with forward reads

Select second set of reads

Data input 'fastq_input2' (fastqsanger)

Specify dataset with reverse reads

CEITEC Galaxy server

Public Galaxy is quite **versatile** and **flexible** (but not that much).
You can always install your **own Galaxy instance**. Advantages:

- (simple) installation of **additional tools, visualizations,** etc.
- own computational **resources** and **personalised quotas** in data storages
- easy to **share datasets**, **workflows**, etc.
- adjustable **access security level** (from no authentication to fully restricted access)

Our Galaxy server can be found here:

https://galaxy.ceitec.muni.cz/

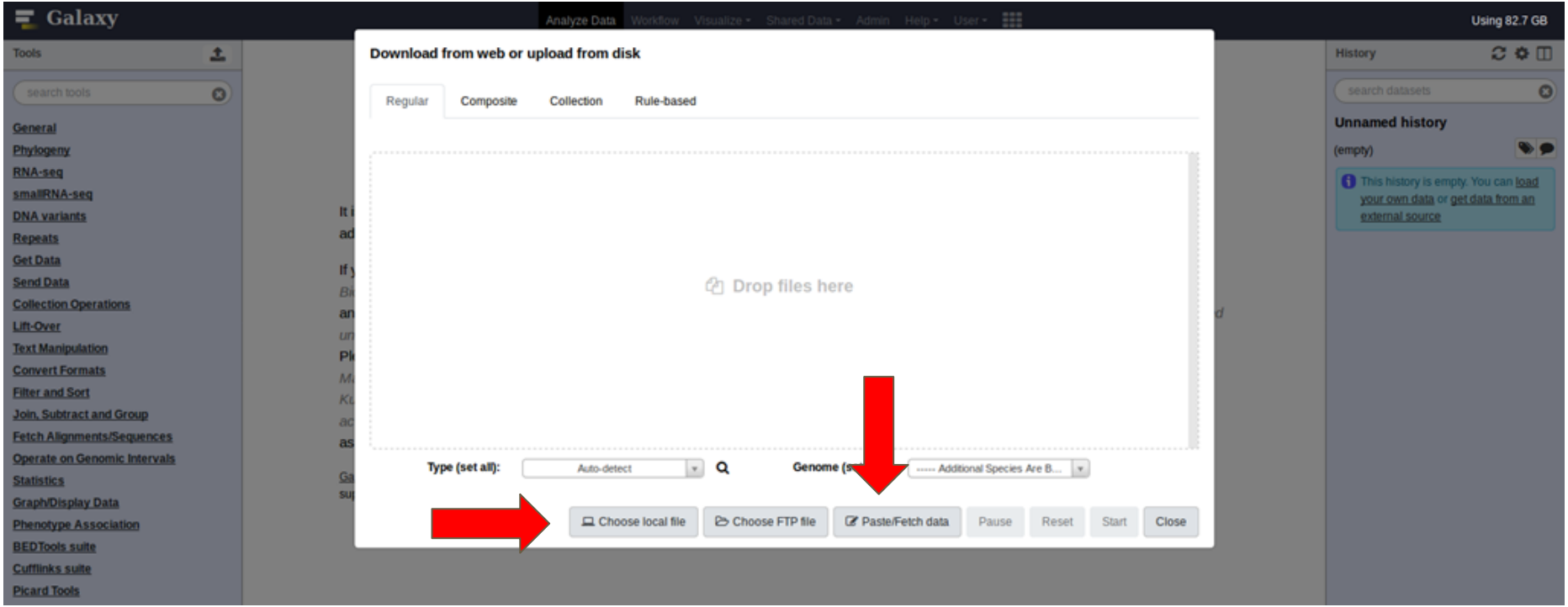
To login please use your **university UCO** and **secondary password**

CEITEC Galaxy server

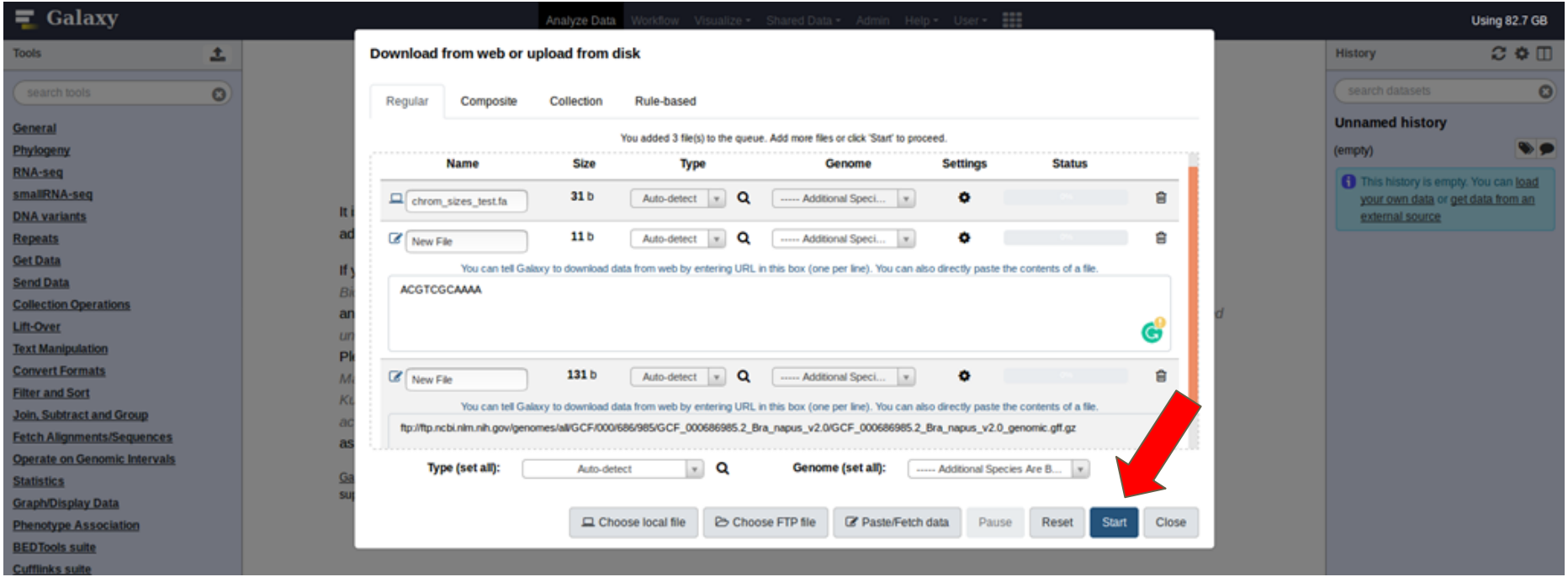
Upload data


Galaxy
Analyze Data
Workflow
Visualize
Shared Data
Admin
Help
User
Using 82.7 GB
Tools
search tools
General
Phylogeny
RNA-seq
smallRNA-seq
DNA variants
Repeats
Get Data
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
MouseMine server
Ratmine server
YeastMine server
modENCODE worm server
WormBase server
ZebrafishMine server
EuPathDB server
HbVar Human Hemoglobin Variants and Thalassemias
GenomeSpace Importer - receive data from GenomeSpace
Send Data
Collection Operations
Lift-Over
CEITEC
Welcome to the CEITEC MU private Galaxy server
It is maintained by Bioinformatics Core Facility. In the case of a problem or with any question about Galaxy do not hesitate to contact our administrator, Martin Demko, at 325073@mail.muni.cz.
If you used the services of this Galaxy server for your published research, we expect you to acknowledge these contributions: Core Facility Bioinformatics of CEITEC Masaryk University is gratefully acknowledged for the obtaining of the scientific data presented in this paper. and MetaCentrum Computational resources were provided by the CESNET LM2015042 and the CERIT Scientific Cloud LM2015085, provided under the programme "Projects of Large Research, Development, and Innovations Infrastructures".
Please, do not forget to cite the Galaxy project itself Enis Afgan, Dannon Baker, Marius van den Beek, Daniel Blankenberg, Dave Bouvier, Martin Čech, John Chilton, Dave Clements, Nate Coraor, Carl Eberhard, Björn Grüning, Aysam Guerler, Jennifer Hillman-Jackson, Greg Von Kuster, Eric Rasche, Nicola Soranzo, Nitesh Turaga, James Taylor, Anton Nekrutenko, and Jeremy Goecks. The Galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2016 update. Nucleic Acids Research (2016) doi: 10.1093/nar/gkw343
as well as the individual tools you used in your analysis. For more information about how to cite Galaxy please follow this link.
Galaxy is an open platform for supporting data intensive research. It is developed by The Galaxy Team with the support of many contributors. The Galaxy Project is supported in part by NHGRI, NSF, The Huck Institutes of the Life Sciences, The Institute for CyberScience at Penn State, and Johns Hopkins University.
History
search datasets
Unnamed History
(empty)

Upload data - sources

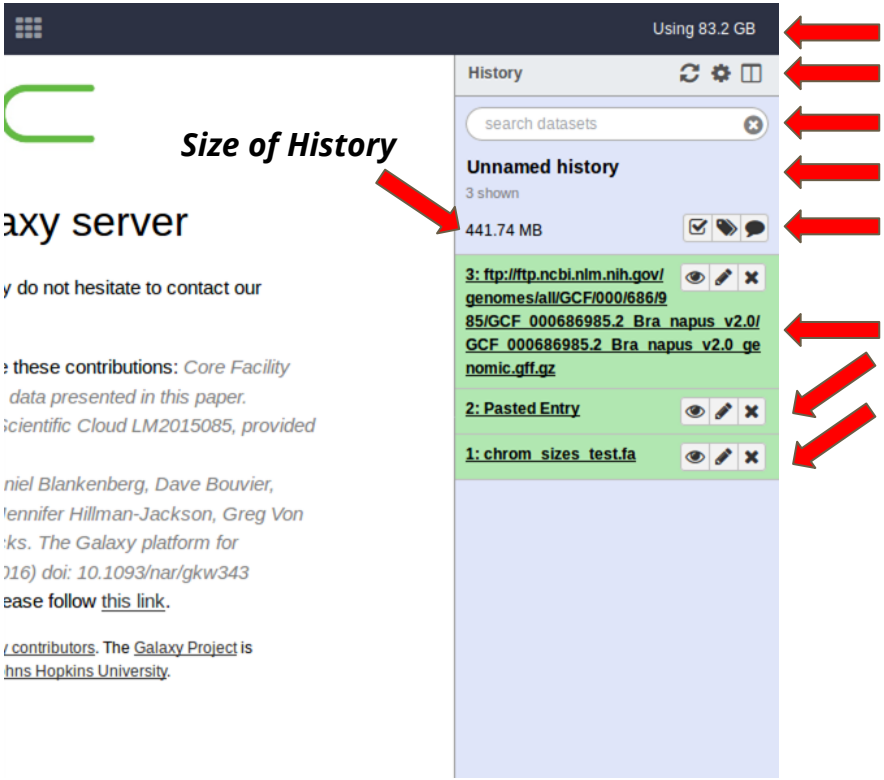


Upload data into History



Upload data into History

History



Total usage *across the Histories*

Reload, ***Advanced options***, ***Multy-history panel***

Search field *(looks in names, tags, comments, metadata)*

History name *(adjustable anytime, recommended to use)*

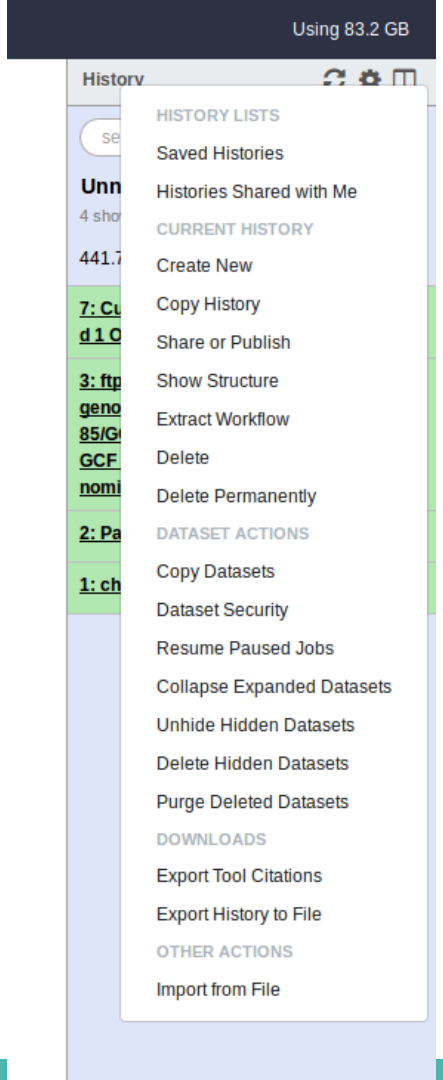
Multi-dataset operations, ***History tags***, ***History annotation***

Dataset *with* ***ID*** *and* ***name***

History - advanced options

Lot of useful features:

- **Copy** whole History
- **Share** or **Publish** History
- **Extract Workflow**
- **Permanent delete** of History or Datasets
- **Export citations** from used Tools



Dataset

HID *- History ID of Dataset (increasing)*
Name *of Dataset (usually name of uploaded file or description of job result, strongly recommended to change)*

Information about ***length***, ***format*** *and associated* ***genome database***

Download *Dataset*
Show ***Details*** *of Dataset or Job*
Show ***available Visualizations***
Show ***Help page***

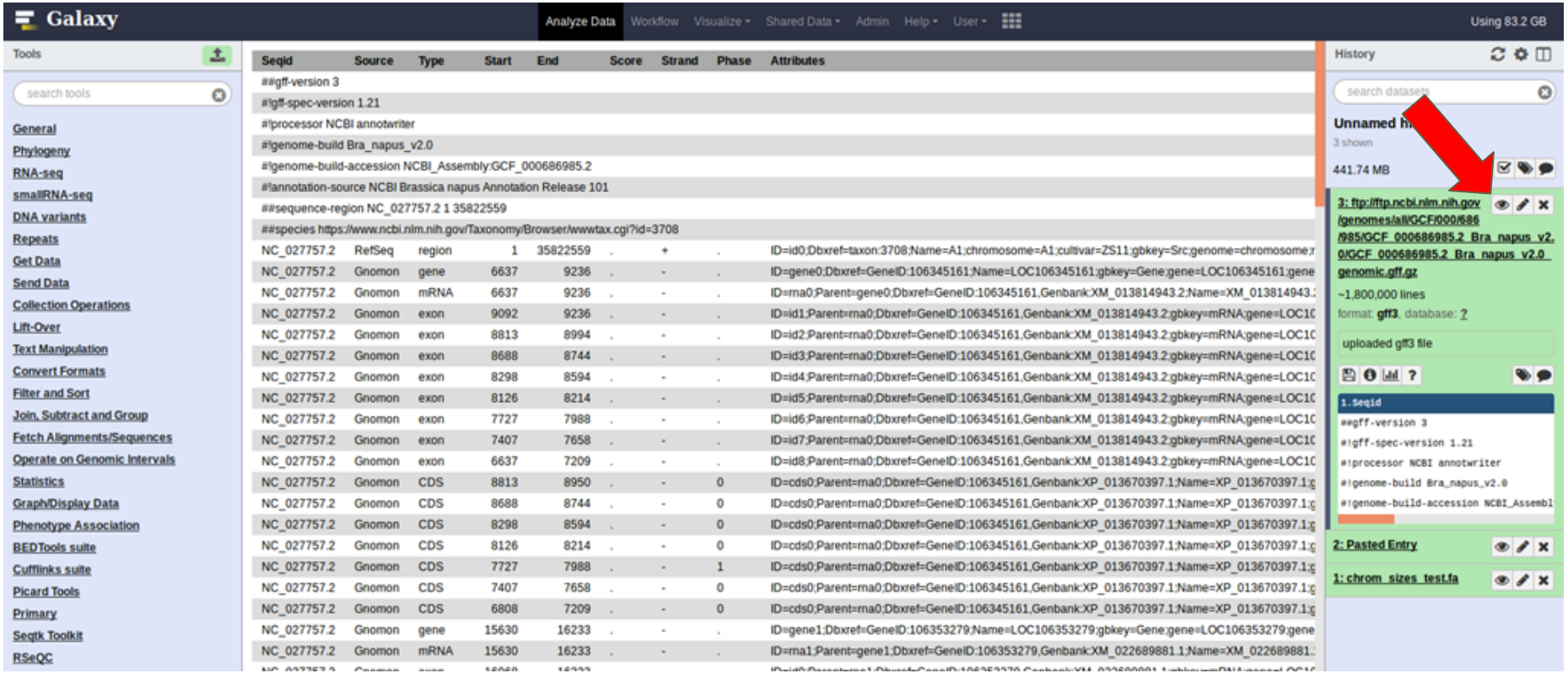
Peek *into Dataset (first several lines)*

View content *of Dataset*
Edit attributes *(name, datatype, access)*
Delete *Dataset (not permanently)*

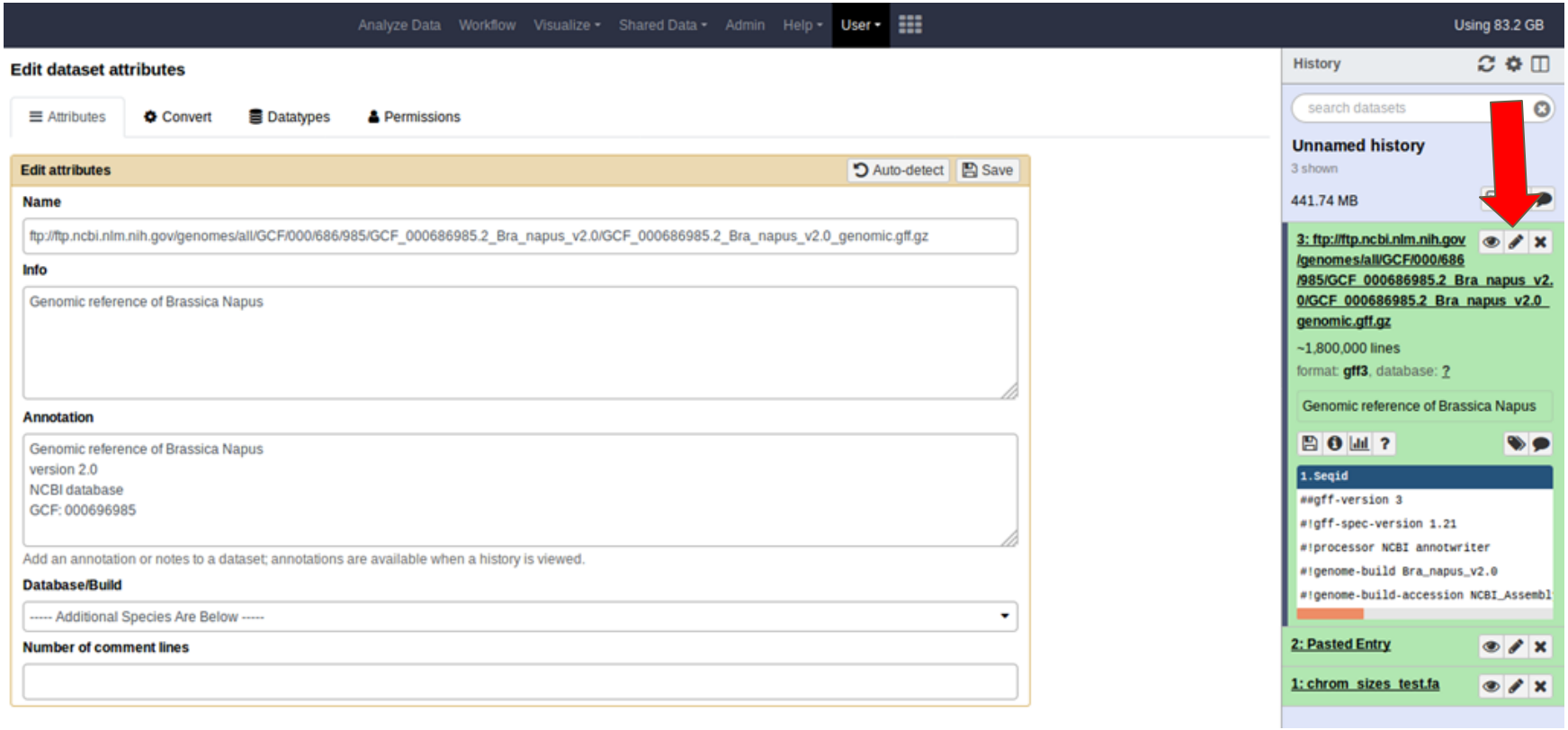
Adjustable ***Info panel***

Edit tags *of Dataset (very usefull)*
Edit attributes *(name, datatype, access)*
Delete *Dataset (not permanently)*

Dataset content

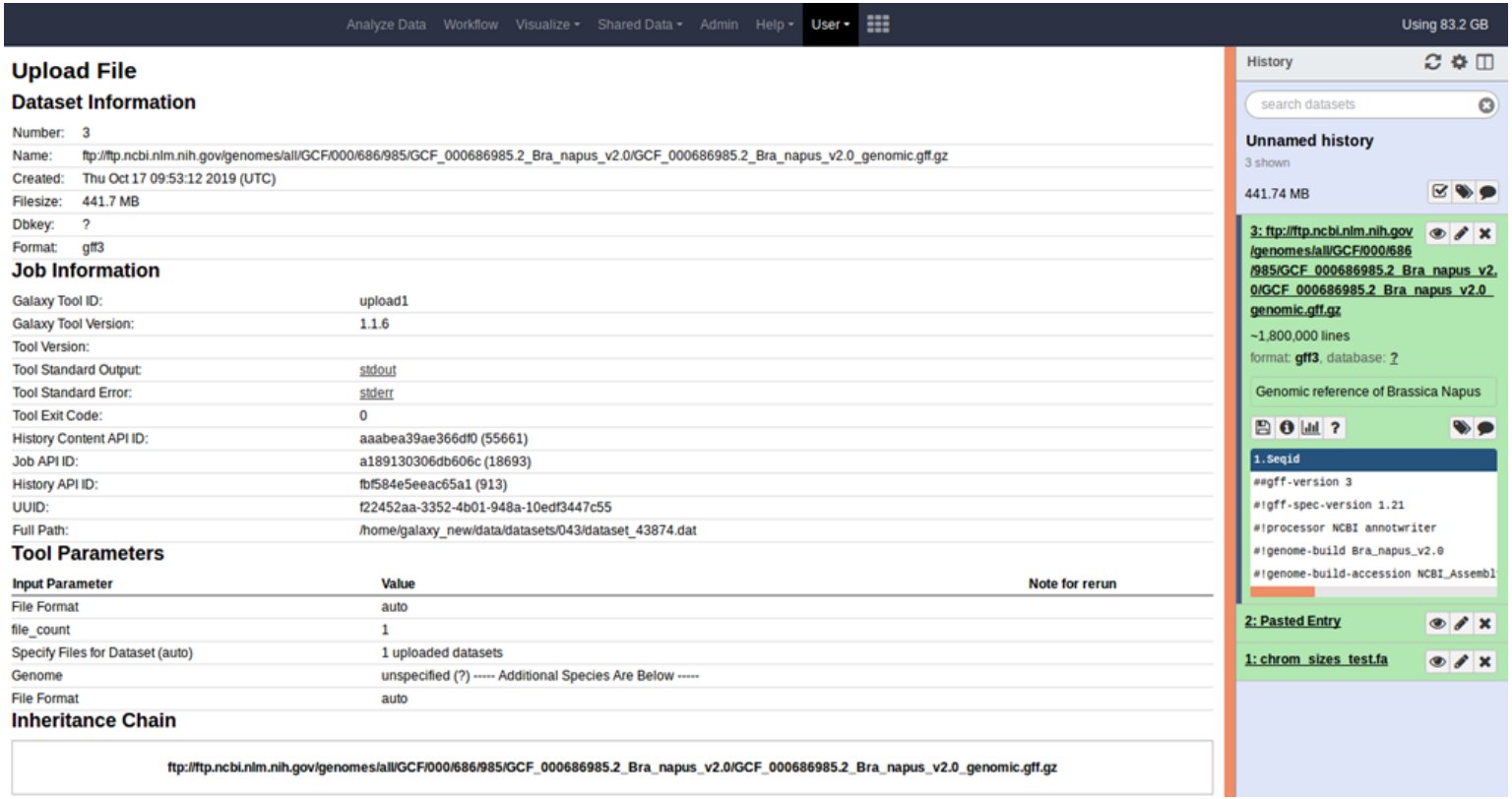


Dataset attributes

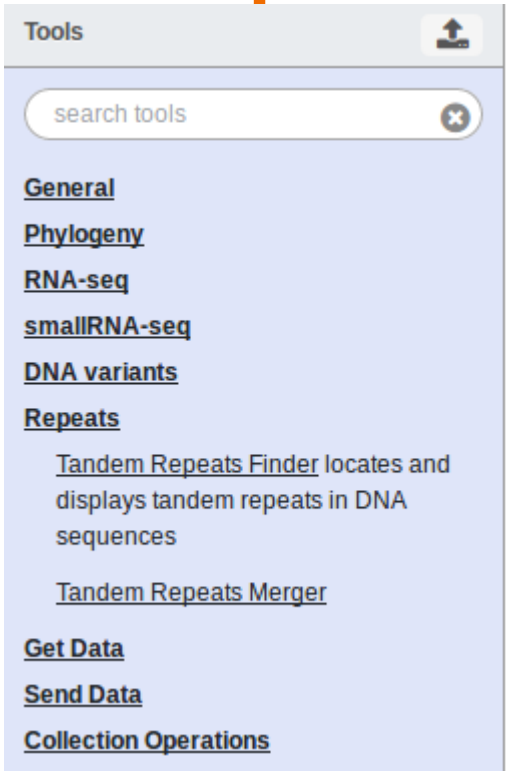


Dataset details

Each **Dataset** is a result of a **Job** using some **Tool** from **Tool panel** (here are useful information about performed Job)



Tool panel

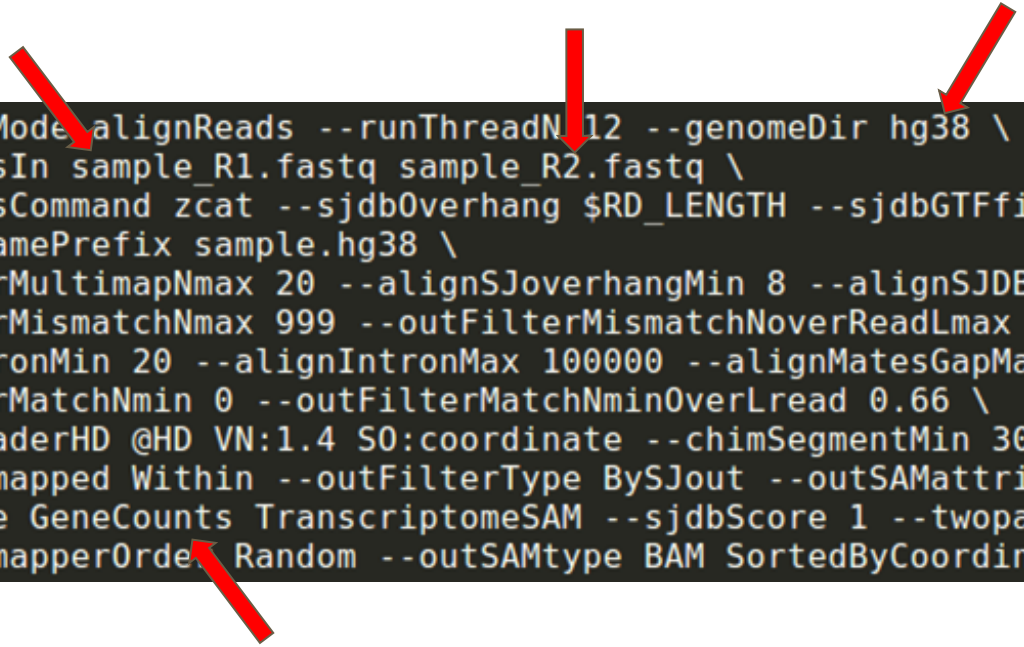


Search field *of Tool panel (lloks into* ***names***, ***descriptions***, ***metadata****)*

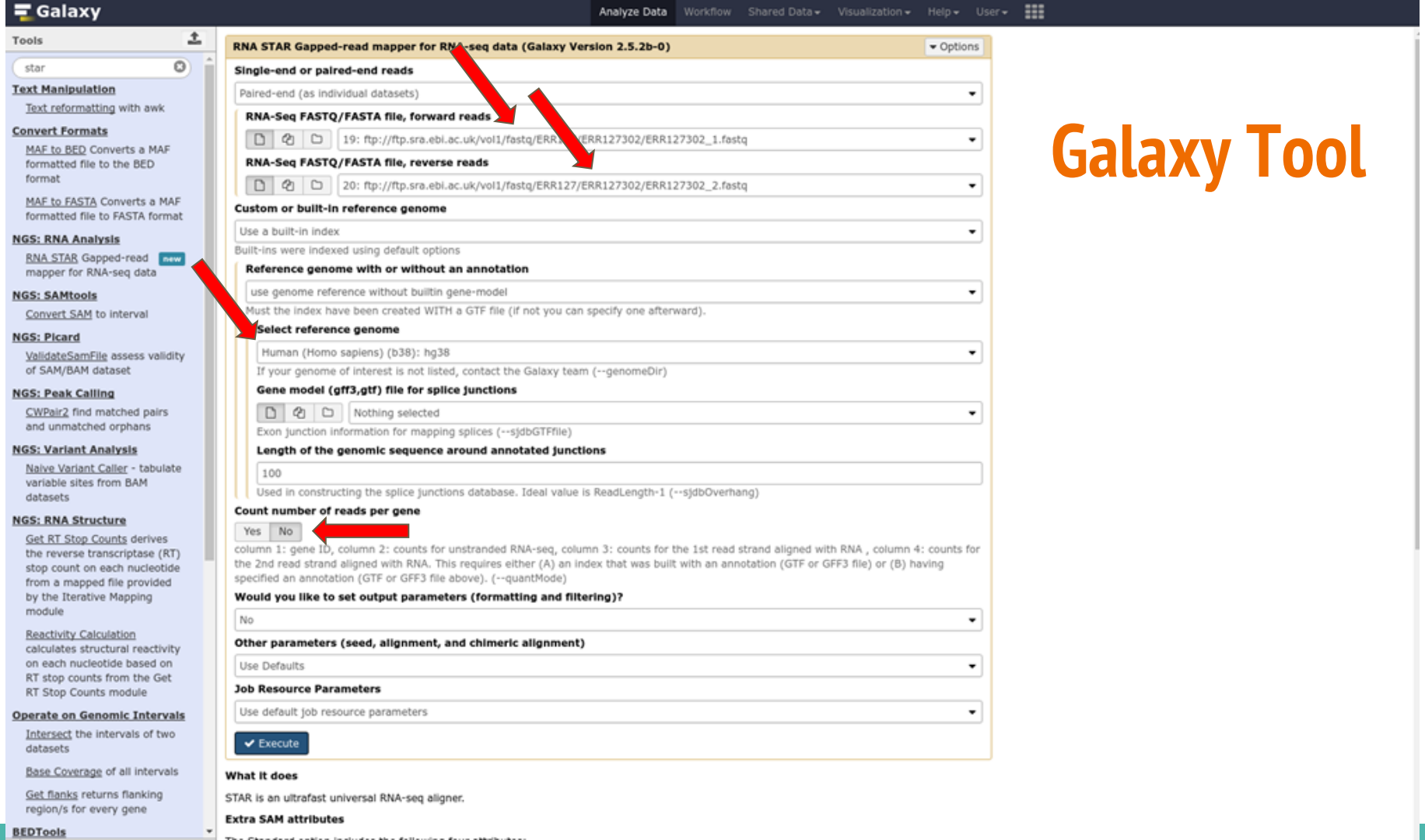
Tool panel *with many (named)* ***groups of tools****.*

Each group can be ***unrolled*** *into a* ***list of tools***

Command-line Tool



```
STAR --runMode alignReads --runThreadN 12 --genomeDir hg38 \
--readFilesIn sample_R1.fastq sample_R2.fastq \
--readFilesCommand zcat --sjdbOverhang $RD_LENGTH --sjdbGTFfile hg38.gtf \
--outFileNamePrefix sample.hg38 \
--outFilterMultimapNmax 20 --alignSJoverhangMin 8 --alignSJDBoverhangMin 1 \
--outFilterMismatchNmax 999 --outFilterMismatchNoverReadLmax 0.04 \
--alignIntronMin 20 --alignIntronMax 100000 --alignMatesGapMax 1000000 \
--outFilterMatchNmin 0 --outFilterMatchNminOverLread 0.66 \
--outSAMheaderHD @HD VN:1.4 SO:coordinate --chimSegmentMin 30 --chimOutType SeparateSAMold \
--outSAMunmapped Within --outFilterType BySJout --outSAMattributes All \
--quantMode GeneCounts TranscriptomeSAM --sjdbScore 1 --twopassMode Basic \
--outMultimapperOrder Random --outSAMtype BAM SortedByCoordinate
```


Galaxy
Analyze Data
Workflow
Shared Data
Visualization
Help
User
Tools
star
Text Manipulation
Text reformatting with awk
Convert Formats
MAF to BED Converts a MAF formatted file to the BED format
MAF to FASTA Converts a MAF formatted file to FASTA format
NGS: RNA Analysis
RNA STAR Gapped-read mapper for RNA-seq data
new
NGS: SAMtools
Convert SAM to interval
NGS: Picard
ValidateSamFile assess validity of SAM/BAM dataset
NGS: Peak Calling
CWPair2 find matched pairs and unmatched orphans
NGS: Variant Analysis
Naive Variant Caller - tabulate variable sites from BAM datasets
NGS: RNA Structure
Get RT Stop Counts derives the reverse transcriptase (RT) stop count on each nucleotide from a mapped file provided by the Iterative Mapping module
Reactivity Calculation calculates structural reactivity on each nucleotide based on RT stop counts from the Get RT Stop Counts module
Operate on Genomic Intervals
Intersect the intervals of two datasets
Base Coverage of all intervals
Get flanks returns flanking region/s for every gene
BEDTools
RNA STAR Gapped-read mapper for RNA-seq data (Galaxy Version 2.5.2b-0)
Options
Single-end or paired-end reads
Paired-end (as individual datasets)
RNA-Seq FASTQ/FASTA file, forward reads
19: ftp://ftp.sra.ebi.ac.uk/vol1/fastq/ERR127302/ERR127302_1.fastq
RNA-Seq FASTQ/FASTA file, reverse reads
20: ftp://ftp.sra.ebi.ac.uk/vol1/fastq/ERR127/ERR127302/ERR127302_2.fastq
Custom or built-in reference genome
Use a built-in index
Built-ins were indexed using default options
Reference genome with or without an annotation
use genome reference without builtin gene-model
Must the index have been created WITH a GTF file (if not you can specify one afterward).
Select reference genome
Human (Homo sapiens) (b38): hg38
If your genome of interest is not listed, contact the Galaxy team (--genomeDir)
Gene model (gff3,gtf) file for splice junctions
Nothing selected
Exon junction information for mapping splices (--sjdbGTFfile)
Length of the genomic sequence around annotated junctions
100
Used in constructing the splice junctions database. Ideal value is ReadLength-1 (--sjdbOverhang)
Count number of reads per gene
Yes
No
column 1: gene ID, column 2: counts for unstranded RNA-seq, column 3: counts for the 1st read strand aligned with RNA , column 4: counts for the 2nd read strand aligned with RNA. This requires either (A) an index that was built with an annotation (GTF or GFF3 file) or (B) having specified an annotation (GTF or GFF3 file above). (--quantMode)
Would you like to set output parameters (formatting and filtering)?
No
Other parameters (seed, alignment, and chimeric alignment)
Use Defaults
Job Resource Parameters
Use default job resource parameters
Execute
What it does
STAR is an ultrafast universal RNA-seq aligner.
Extra SAM attributes

Galaxy Tool

Command-line vs Galaxy

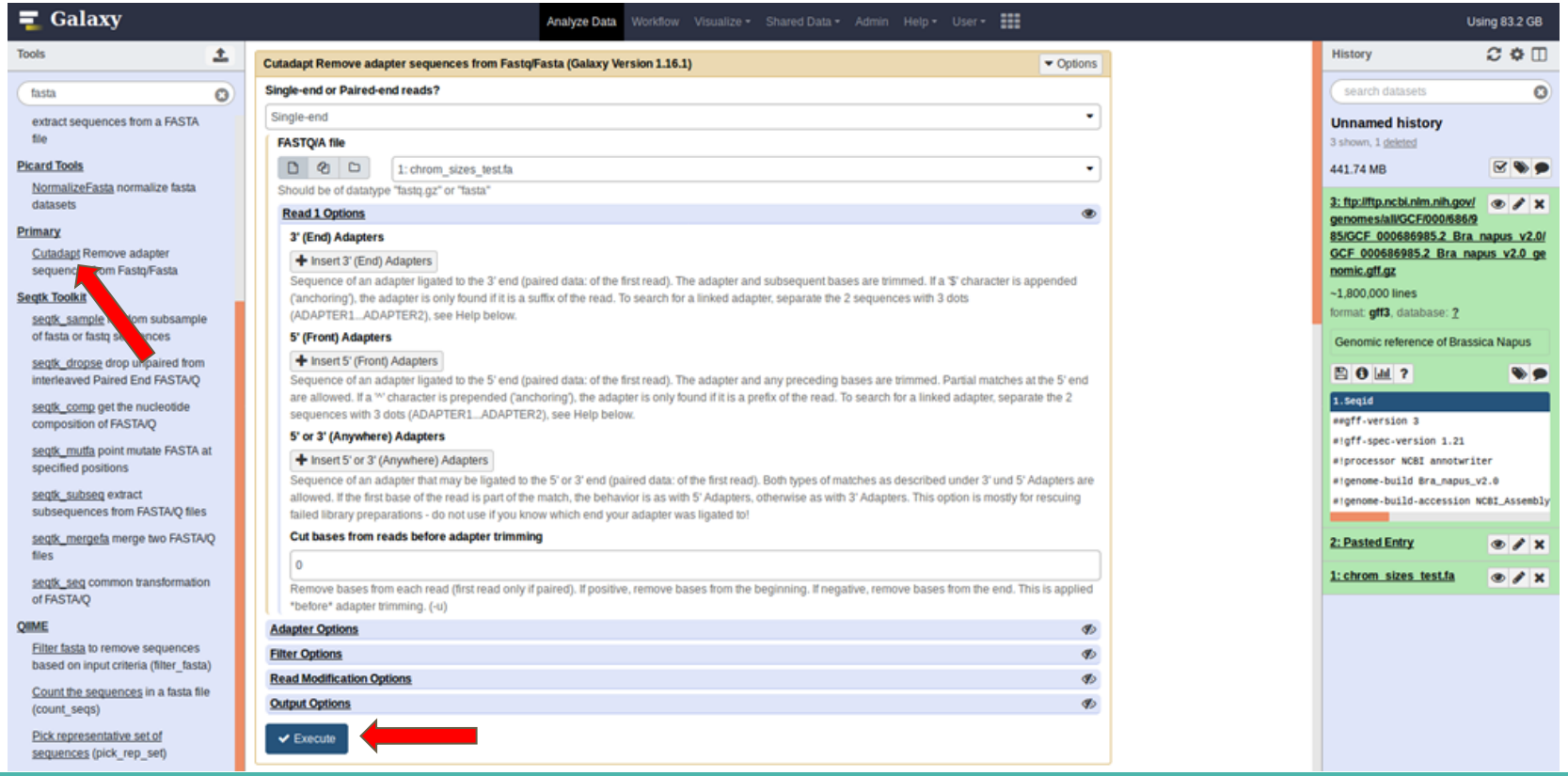
Galaxy

- is build **on top of the bioinformatic** (and not only) **tools** providing non-bioinformaticians an access to the tools
- puts a **graphic interface** on the top of the command-line making it much more user-friendly
- allows **sharing** of the data, **workflows**, **results**, **visualizations**, etc.
- allows to **repeat** and **edit Job settings** easily
- Is great for **small-scale analyses**
- does **not** support all types of tools (e.g., **online tools**)

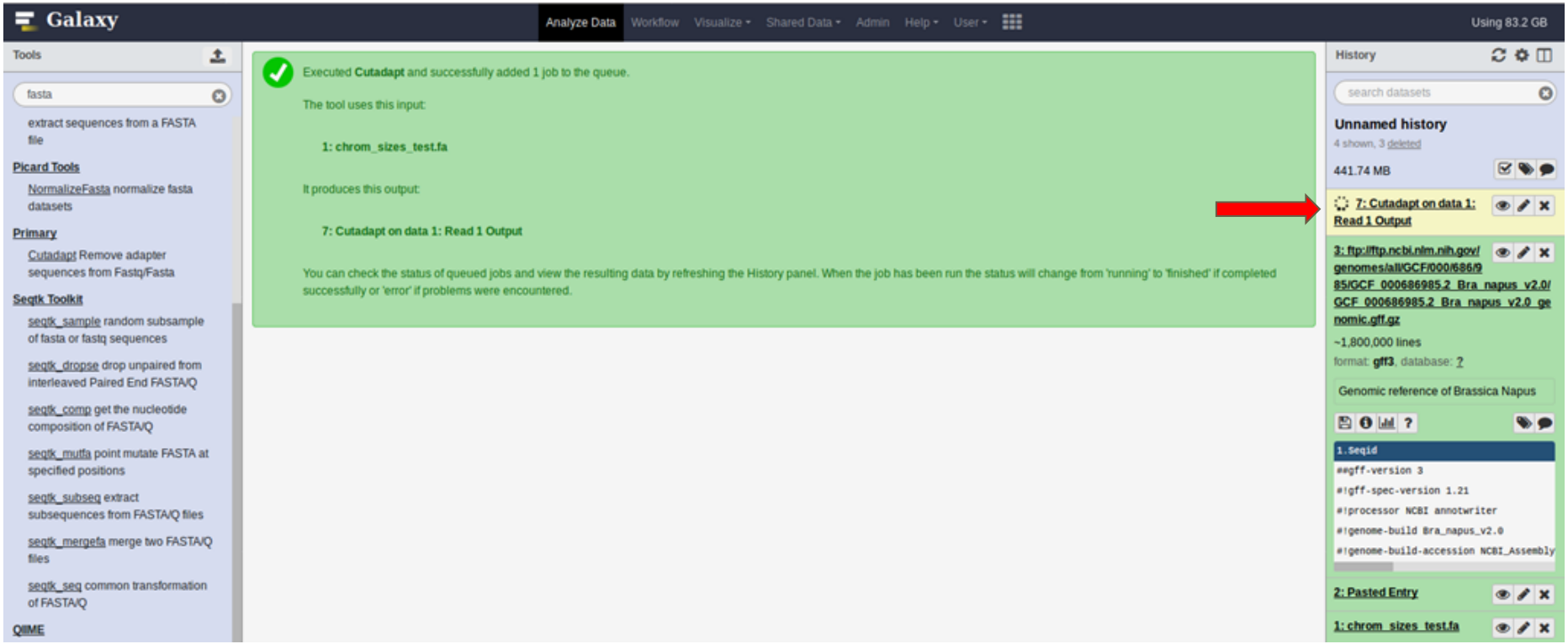
Command-line is more **flexible**

For **large amount of data** or **unsupported tools** it is still better to use **command-line**

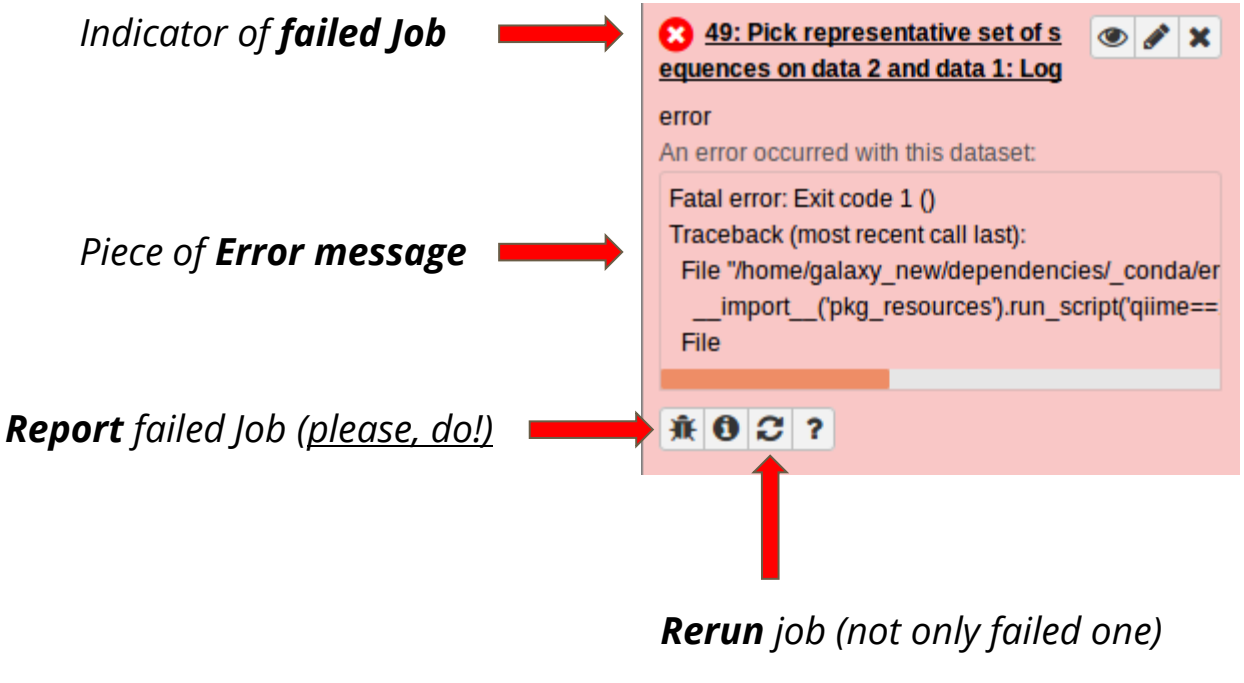
Job - setup


Galaxy
Analyze Data
Workflow
Visualize
Shared Data
Admin
Help
User
Using 83.2 GB
Tools
fasta
extract sequences from a FASTA file
Picard Tools
NormalizeFasta normalize fasta datasets
Primary
Cutadapt Remove adapter sequences from Fastq/Fasta
Seqtk Toolkit
seqtk_sample
seqtk_dropse drop unpaired from interleaved Paired End FASTA/Q
seqtk_comp get the nucleotide composition of FASTA/Q
seqtk_mutfa point mutate FASTA at specified positions
seqtk_subseq extract subsequences from FASTA/Q files
seqtk_mergefa merge two FASTA/Q files
seqtk_seq common transformation of FASTA/Q
QIIME
Filter fasta to remove sequences based on input criteria (filter_fasta)
Count the sequences in a fasta file (count_seqs)
Pick representative set of sequences (pick_rep_set)
Cutadapt Remove adapter sequences from Fastq/Fasta (Galaxy Version 1.16.1)
Options
Single-end or Paired-end reads?
Single-end
FASTQ/A file
1: chrom_sizes_test.fa
Should be of datatype "fastq.gz" or "fasta"
Read 1 Options
3' (End) Adapters
Insert 3' (End) Adapters
Sequence of an adapter ligated to the 3' end (paired data: of the first read). The adapter and subsequent bases are trimmed. If a '$' character is appended ('anchoring'), the adapter is only found if it is a suffix of the read. To search for a linked adapter, separate the 2 sequences with 3 dots (ADAPTER1...ADAPTER2), see Help below.
5' (Front) Adapters
Insert 5' (Front) Adapters
Sequence of an adapter ligated to the 5' end (paired data: of the first read). The adapter and any preceding bases are trimmed. Partial matches at the 5' end are allowed. If a '^' character is prepended ('anchoring'), the adapter is only found if it is a prefix of the read. To search for a linked adapter, separate the 2 sequences with 3 dots (ADAPTER1...ADAPTER2), see Help below.
5' or 3' (Anywhere) Adapters
Insert 5' or 3' (Anywhere) Adapters
Sequence of an adapter that may be ligated to the 5' or 3' end (paired data: of the first read). Both types of matches as described under 3' und 5' Adapters are allowed. If the first base of the read is part of the match, the behavior is as with 5' Adapters, otherwise as with 3' Adapters. This option is mostly for rescuing failed library preparations - do not use if you know which end your adapter was ligated to!
Cut bases from reads before adapter trimming
0
Remove bases from each read (first read only if paired). If positive, remove bases from the beginning. If negative, remove bases from the end. This is applied *before* adapter trimming. (-u)
Adapter Options
Filter Options
Read Modification Options
Output Options
Execute
History
search datasets
Unnamed history
3 shown, 1 deleted
441.74 MB
3: ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/686/985/GCF_000686985.2_Bra_napus_v2.0/GCF_000686985.2_Bra_napus_v2.0_genomic.gff.gz
~1,800,000 lines
format: gff3, database: ?
Genomic reference of Brassica Napus
1.Seqid
##gff-version 3
#!gff-spec-version 1.21
#!processor NCBI annotwriter
#!genome-build Bra_napus_v2.0
#!genome-build-accession NCBI_Assembly
2: Pasted Entry
1: chrom_sizes_test.fa

Job - execution



Job - error


Indicator of **failed Job**
Piece of **Error message**
Report failed Job (please, do!)
49: Pick representative set of sequences on data 2 and data 1: Log
error
An error occurred with this dataset:
Fatal error: Exit code 1 ()
Traceback (most recent call last):
File "/home/galaxy_new/dependencies/_conda/er
__import__('pkg_resources').run_script('qiime==
File
Rerun job (not only failed one)

Report Job error

Dataset Error

An error occured while running the tool **toolshed.g2.bx.psu.edu/repos/iuc/qiime_pick_rep_set/qiime_pick_rep_set/1.9.1.0**.

Tool execution generated the following messages:

```
Fatal error: Exit code 1 ()
Traceback (most recent call last):
  File "/home/galaxy_new/dependencies/_conda/envs/__qiime@1.9.1/bin/pick_rep_set.py", line 4, in <module>
    __import__('pkg_resources').run_script('qiime==1.9.1', 'pick_rep_set.py')
  File "/home/galaxy_new/dependencies/_conda/envs/__qiime@1.9.1/lib/python2.7/site-packages/pkg_resources/__init__.py", line 666, in run_script
    self.require(requires)[0].run_script(script_name, ns)
  File "/home/galaxy_new/dependencies/_conda/envs/__qiime@1.9.1/lib/python2.7/site-packages/pkg_resources/__init__.py", line 1462, in run_script
    exec(code, namespace, namespace)
  File "/home/galaxy_new/dependencies/_conda/envs/__qiime@1.9.1/lib/python2.7/site-packages/qiime-1.9.1-py2.7.egg-info/scripts/pick_rep_set.py", line 125,
in <module>
    main()
  File "/home/galaxy_new/dependencies/_conda/envs/__qiime@1.9.1/lib/python2.7/site-packages/qiime-1.9.1-py2.7.egg-info/scripts/pick_rep_set.py", line 121,
in main
    sort_by=opts.sort_by)
  File "/home/galaxy_new/dependencies/_conda/envs/__qiime@1.9.1/lib/python2.7/site-packages/qiime/pick_rep_set.py", line 190, in __call__
    of.write('>%s %s\n%s\n' % (cluster, id_, seqs[id_]))
KeyError: 'BarcodeSequence'
```

Troubleshoot This Error

There are a number of help resources to self diagnose and correct problems. Start here: My job ended with an error. What can I do?

Report This Error

Usually the local Galaxy administrators regularly review errors that occur on the server However, if you would like to provide additional information (such as what you were trying to do when the error occurred) and a contact e-mail address, we will be better able to investigate your problem and get back to you.

Error Report

Your email

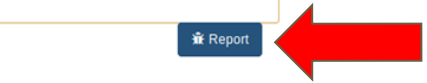
325073@muni.cz

Your email address

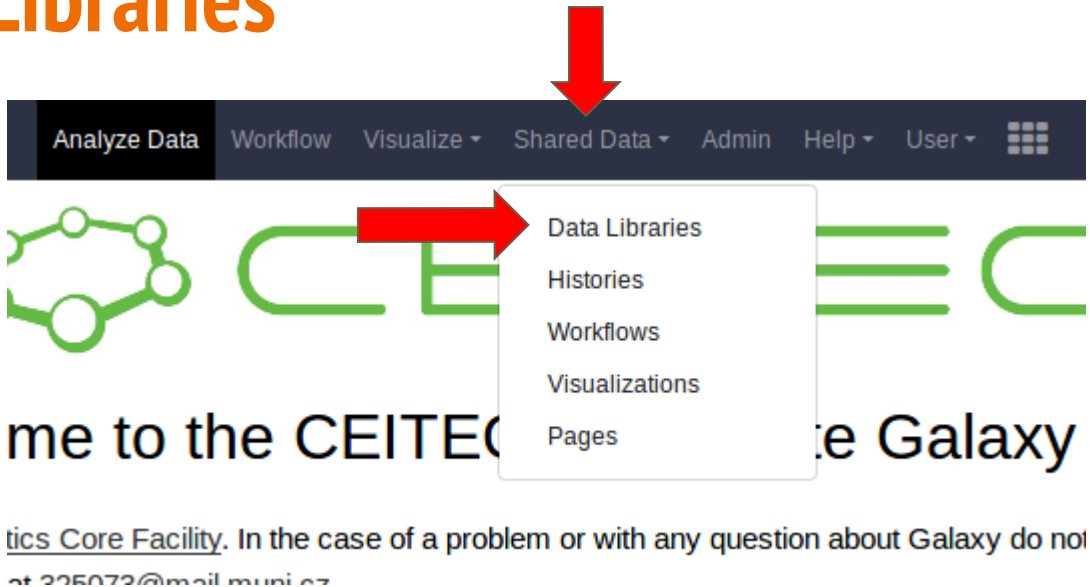
Message

Any additional comments you can provide regarding what you were doing at the time of the bug.

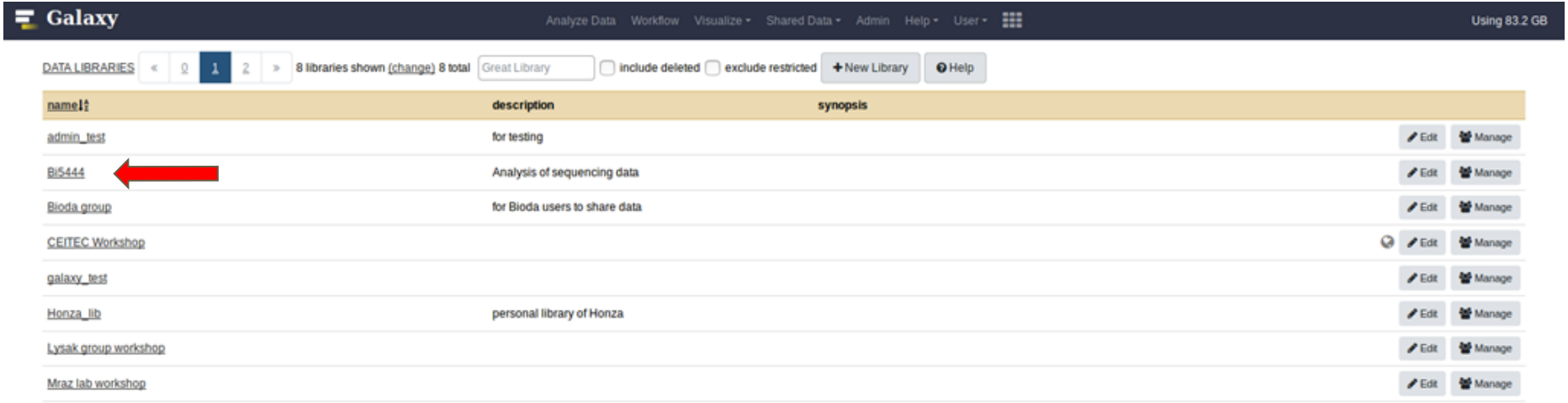
Report



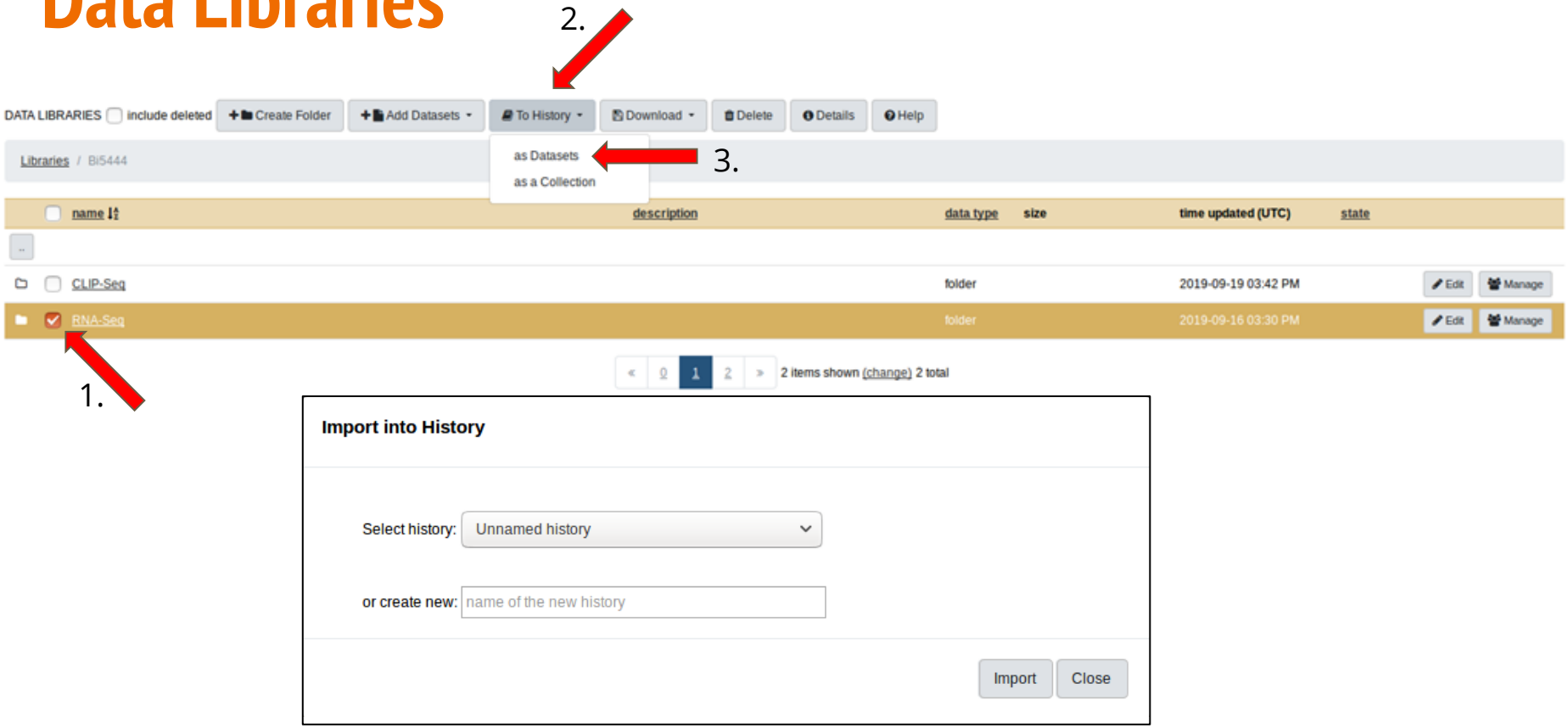
Data Libraries



Data Libraries



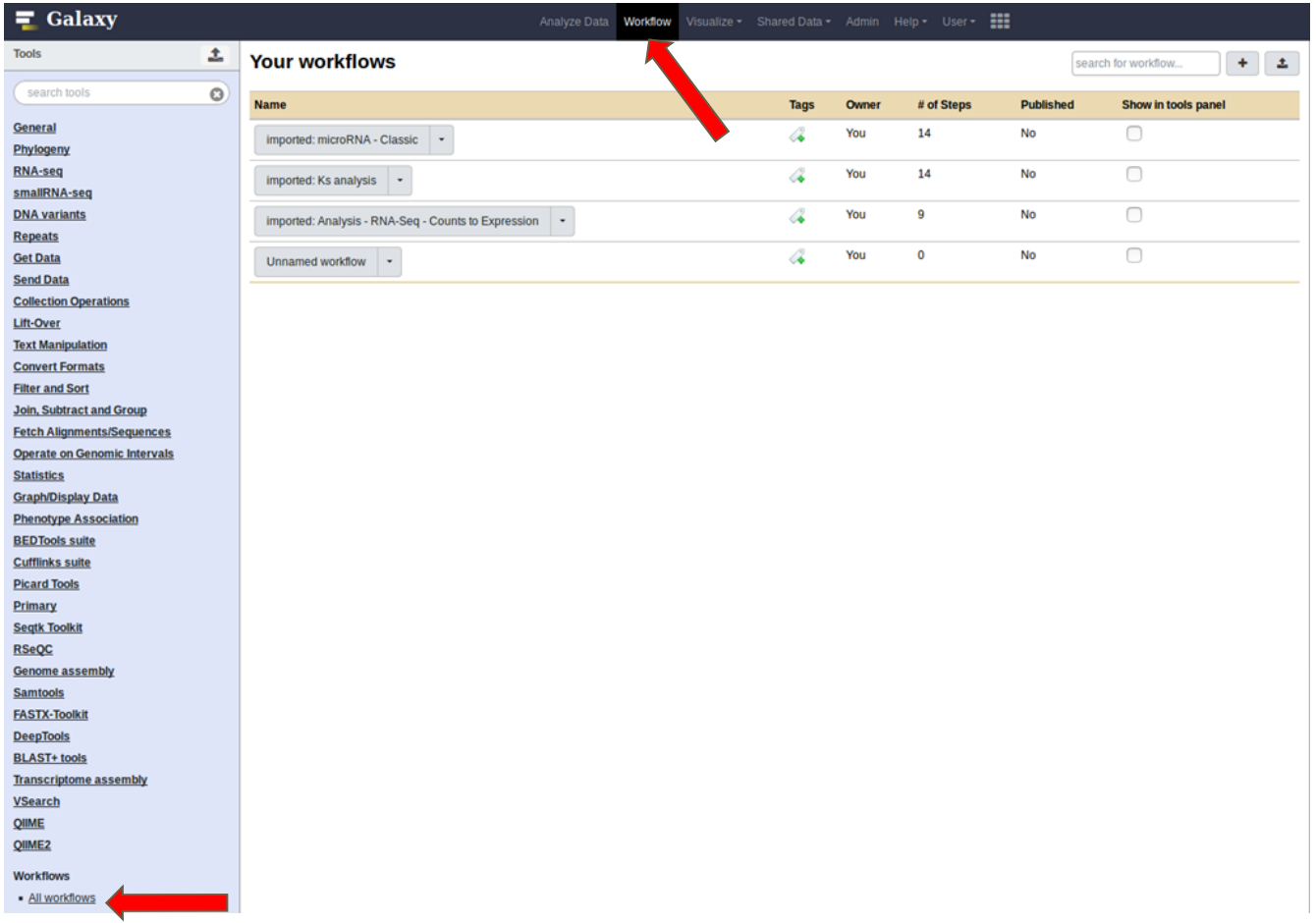
Data Libraries



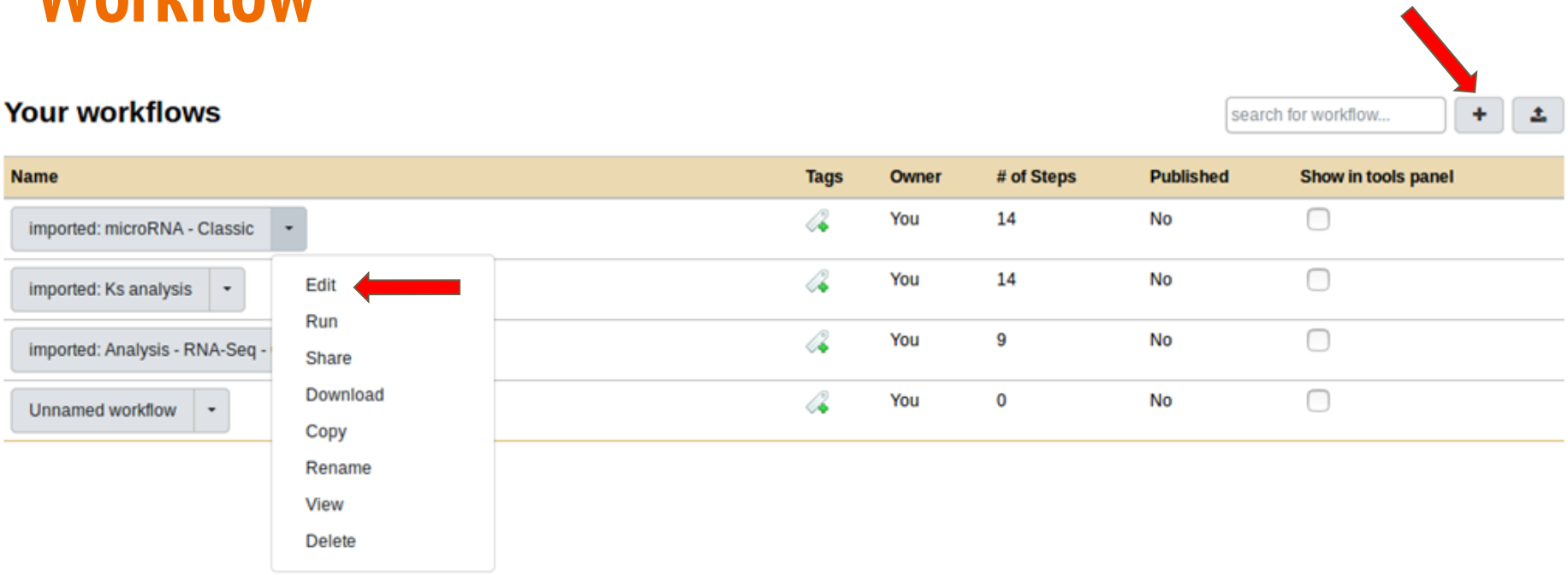
Workflow

Workflow or **pipeline** is an **automatisation** of multi-step **analysis**

Set (or tree) of **tools** taking the **input** from the **output** of another tool (except the very first)

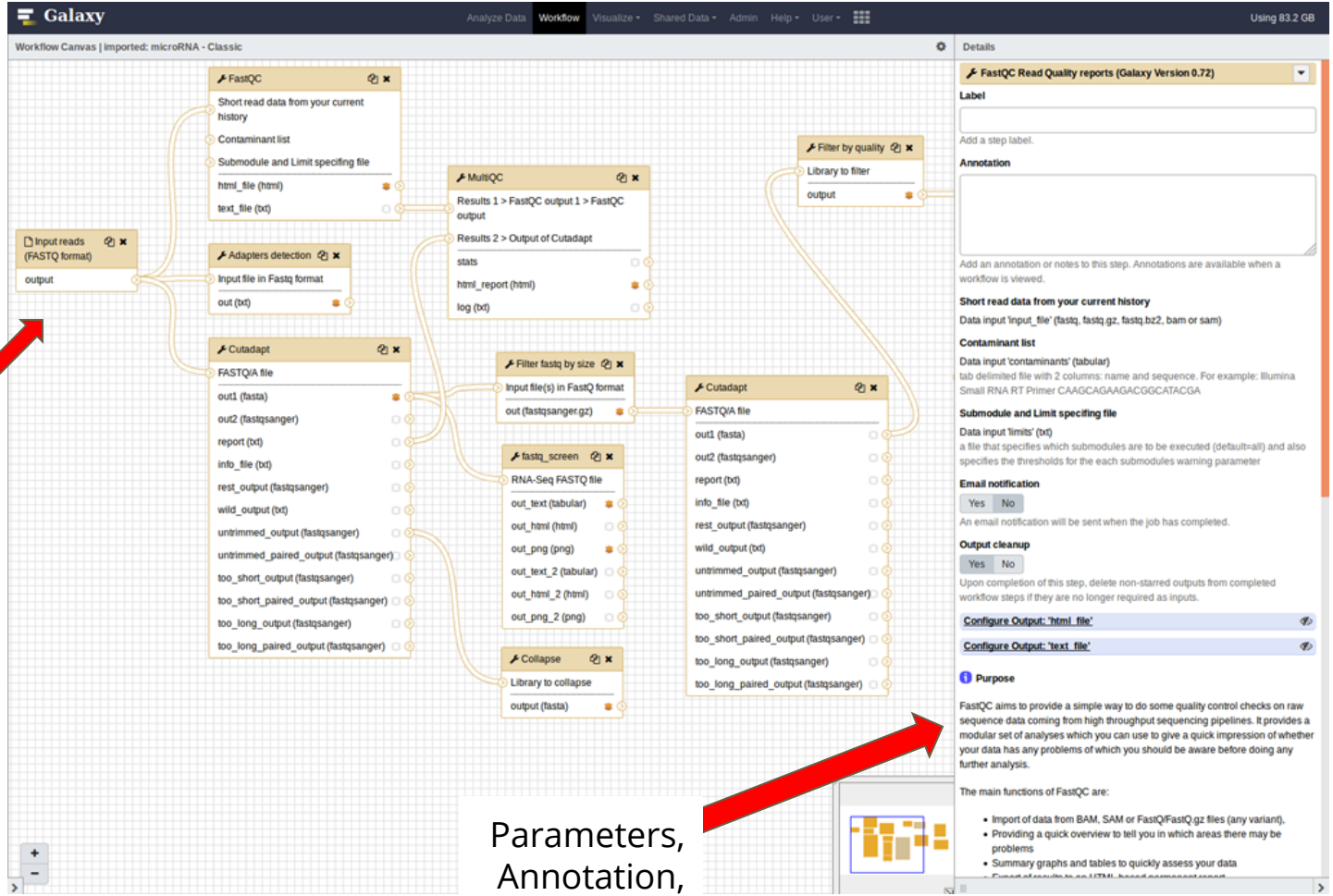


Workflow


Your workflows
search for workflow...
Name
Tags
Owner
# of Steps
Published
Show in tools panel
imported: microRNA - Classic
You
14
No
imported: Ks analysis
You
14
No
imported: Analysis - RNA-Seq -
You
9
No
Unnamed workflow
You
0
No
Edit
Run
Share
Download
Copy
Rename
View
Delete

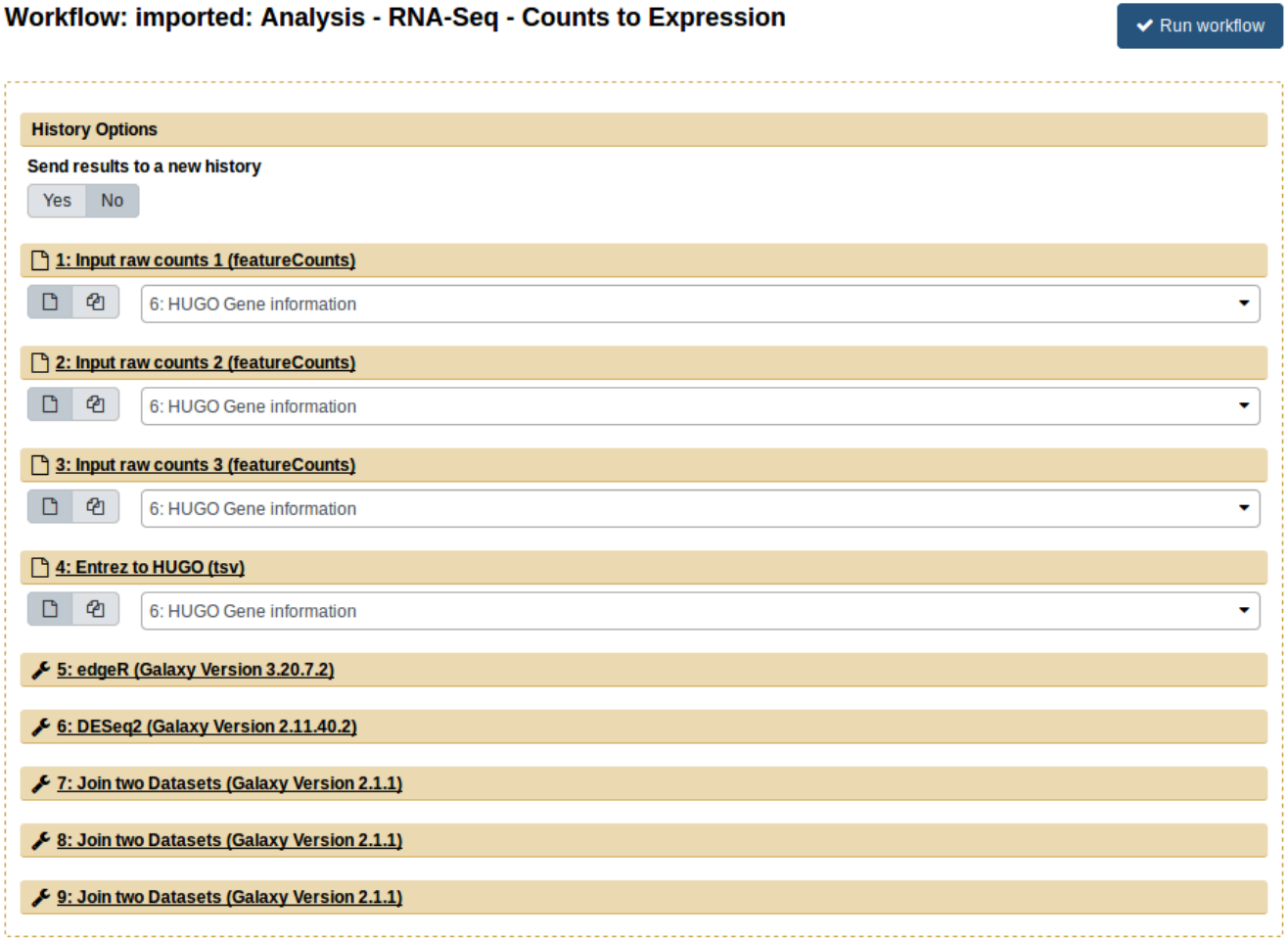
Workflow - build

Tools can be **drag&dropped** from Tool panel and **connected** together only if **output** and **input** Dataset **format matches**

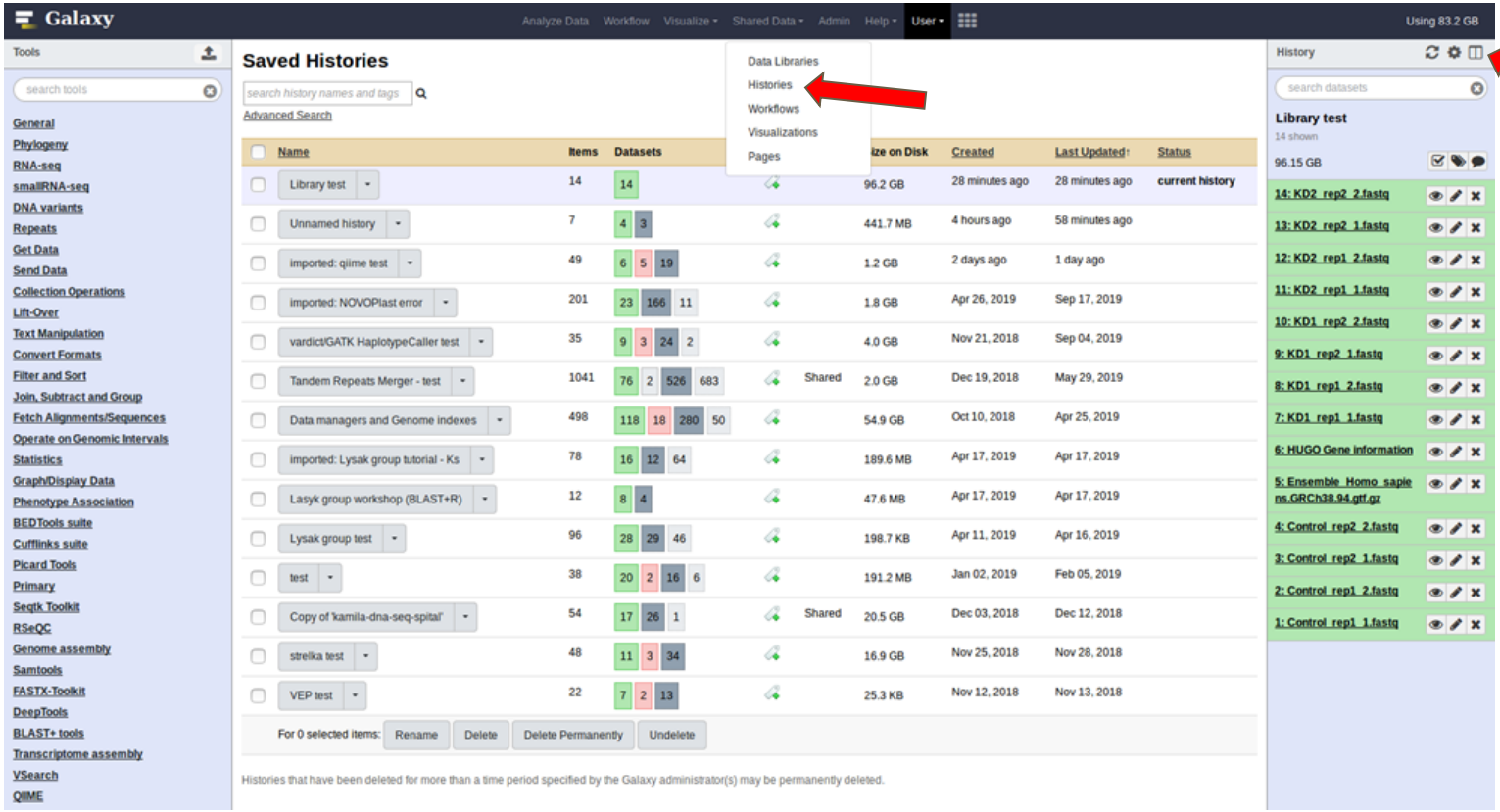


Parameters, Annotation, etc.

Workflow - execution

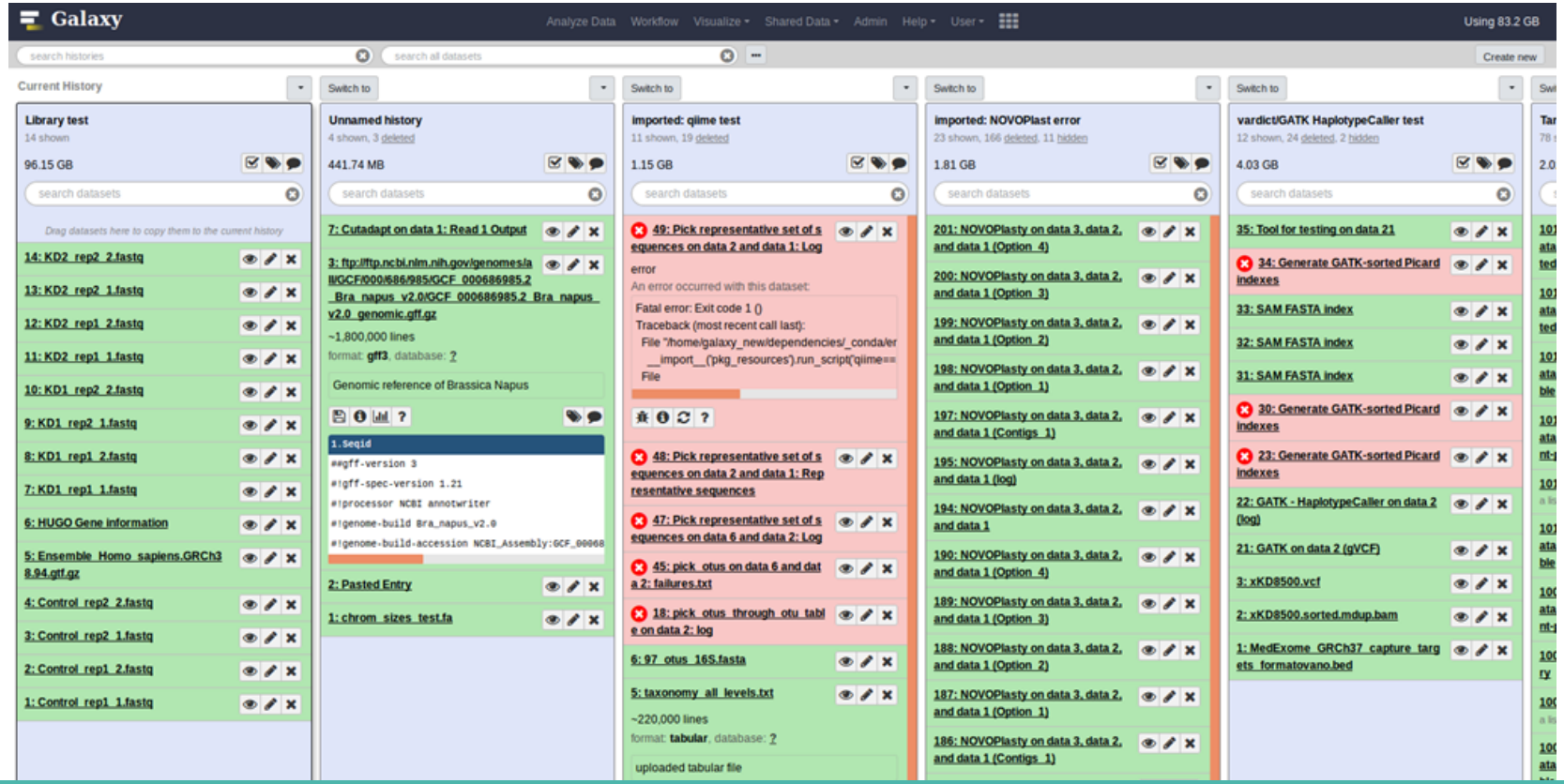


Saved Histories

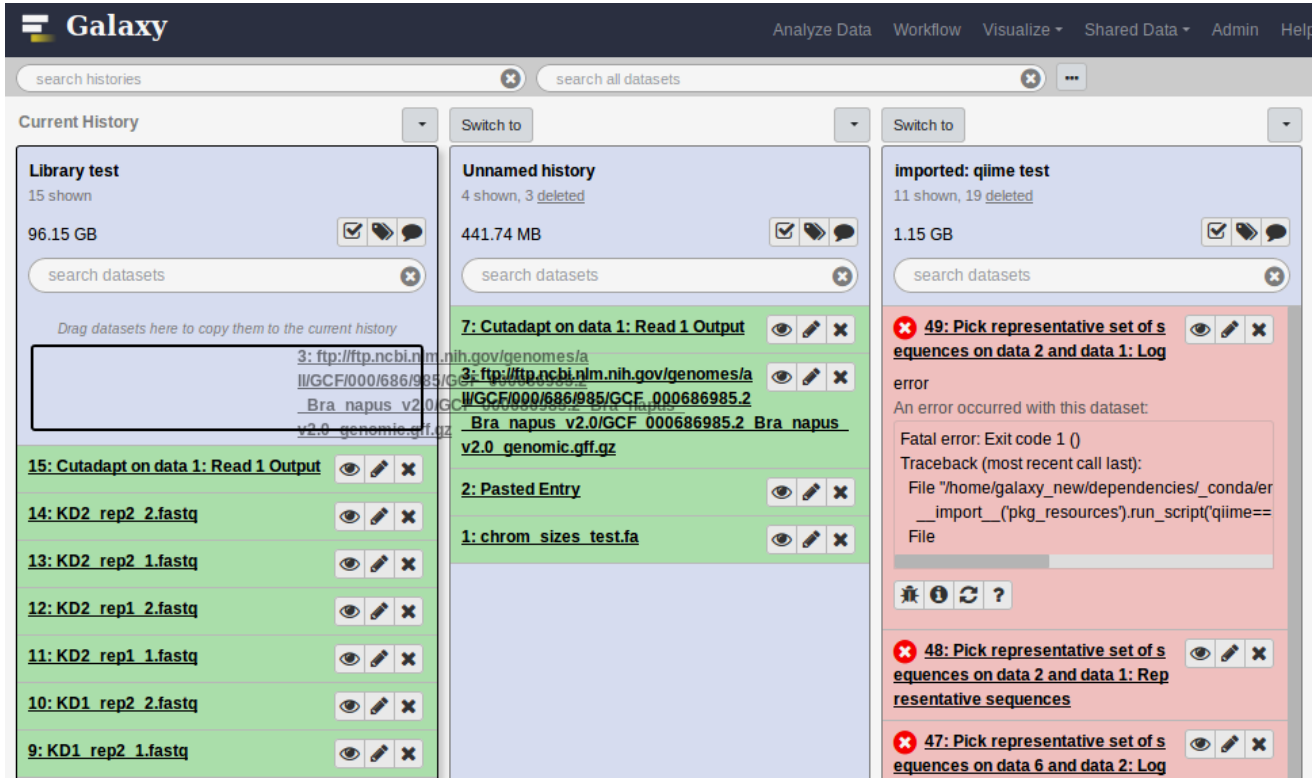


Special panel (next slide)

Multi-history panel



Multi-history panel - drag&drop copy Datasets



Visualization

Depends entirely
on Dataset **format**