Biogenouest

IFB
INSTITUT FRANÇAIS DE BIOINFORMATIQUE

EMBRC FRANCE
CENTRE NATIONAL DE RESSOURCES BIOLOGIQUES MARINES

OCEANOMICS

IDEALG
seaweed for the future

ABiMS

17/01/2022

Galaxy

Initiation

Loraine Guéguen
Erwan Corre

Credits to Gildas Le Corguillé, Galaxy Training Network

v3.0

SORBONNE UNIVERSITÉ
CRÉATEURS DE FUTURS DEPUIS 1257

cnrs

Setup

- Local computer:
  - login: guest
  - pwd: guest29
- Slides:
  - google slides: https://tinyurl.com/glx-init-slides
  - pdf: https://tinyurl.com/glx-init-pdf-slides
- Datasets: https://tinyurl.com/glx-init-datasets
- Galaxy France https://usegalaxy.fr/
  - registration ok ?

Schedule

- Schedule:
  - 09:00 - 12:00
  - 13:30 - 17:30

- Short round table

Learning objectives

- Learning objectives:
  - Familiarize yourself with the basics of Galaxy
  - Learn how to import data
  - Learn how to run tools
  - Learn how histories work
  - Learn how to create a workflow
  - Learn how to share your work
  - Understand and master dataset collections

Program

- Introduction
- Data import
- Tools
- Dataset
- History
- Workflow
- Share

Program

- Introduction
- Data import
- Tools
- Dataset
- History
- Workflow
- Share

Sequence files manipulation

Welcome to Galaxy Training!

Collection of tutorials developed and maintained by the worldwide Galaxy community

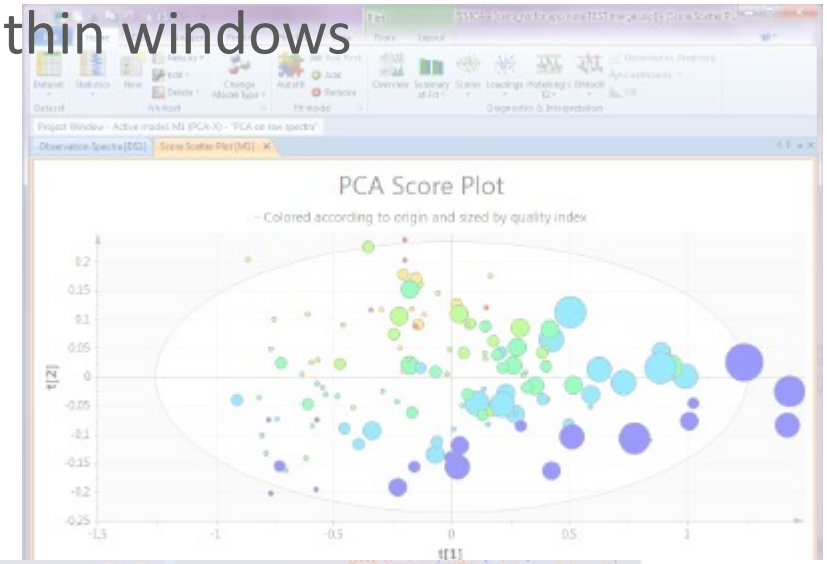
INTRODUCTION / PROBLEMATIC

Introduction / problematic

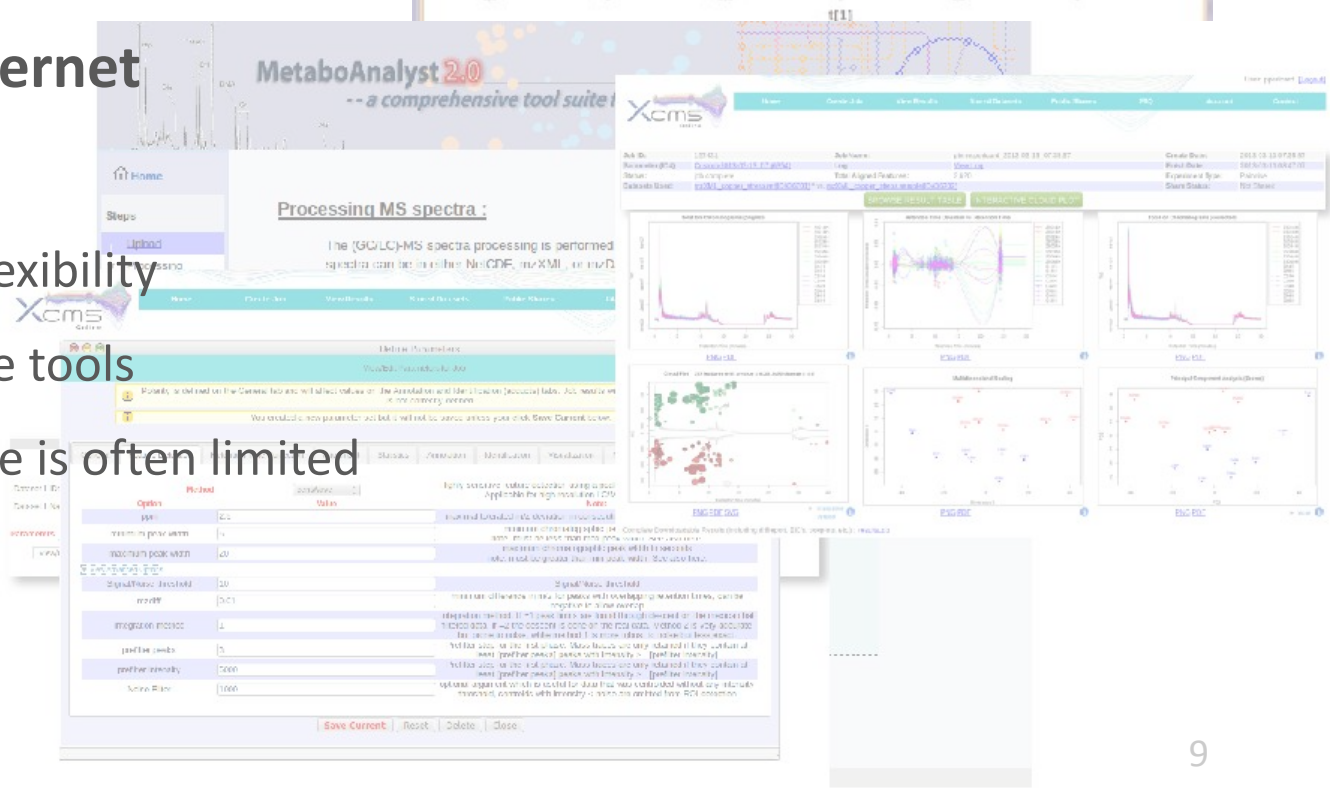
- In biomedical research, high-throughput technologies produce **large datasets**.
- How to perform analyses of these data **without bioinformatics skills** ?
    - Assemble transcript sequences de novo
    - Determine the gene expression
    - Build a phylogenetic tree
    - Predict subcellular targeting for proteins
    - Identify and quantify metabolites detected by LC-MS
    - …

Introduction / problematic

- **Graphical interface** click-button tools within windows
  - + very ergonomic
  - – too ergonomic → lack of flexibility
  - – few
  - – paying for it!



- Tools available on the **internet**
  - + very ergonomic
  - – too ergonomic → lack of flexibility
  - – A small part of the available tools
  - – the submission size /storage is often limited
  - – must not be paranoid



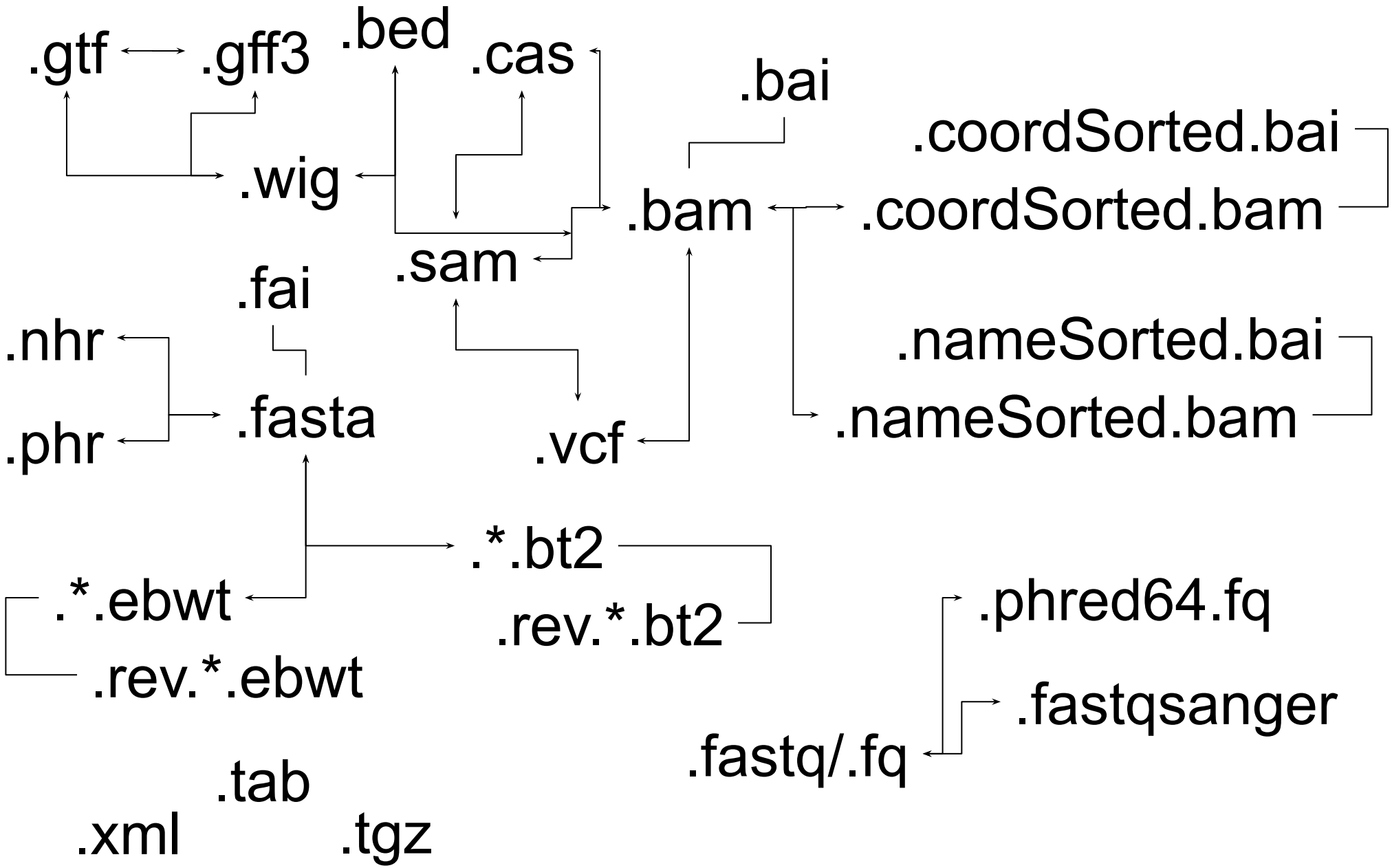
Introduction / problematic

- **Command line** tools
  - \+ represent almost the majority of scientific tools
  - \+ good parameters completeness
  - \+ can be executed on high performance computers
  - \+ automatable, workflowsable, ...
  - – minimum linux knowledge is required
  - – cruel lack of ergonomics

```
login@sbr4-1042:~$ ssh -Y login@bioinfo.sb-roscoff.fr
[...]
[login@n0 ~]$ cdprojet
[login@n0 login]$ cd 13-07-29-panda/tmp/mapping
[login@n0 mapping]$ cat tophat.qsub
#!/bin/bash
#$ -S /bin/bash
#$ -M login@sb-roscoff.fr
#$ -m bea
#$ -V
#$ -cwd
#$ -o qsub.out
#$ -e qsub.err

tophat2 panda_v121029 ../input/IllR1-1.fq ../input/IllR1-2.fq
-GTF ../input/panda_v121029.gtf --b2-sensitive -r 100
-num-threads 8

[login@n0 mapping]$ qsub -q long.q -pe thread 8 tophat.qsub
Your job 5338969 ("tophat.qsub") has been submitted
[login@n0 mapping]$ ls
accepted_hits.bam  junctions.bed    qsub.err  unmapped.bam
deletions.bed      logs             qsub.out
insertions.bed     prep_reads.info  tmp
[login@n0 mapping]$ cd
```

```
login@sbr4-1042:~$ ssh -Y login@bioinfo.sb-roscoff.fr
[...]
[login@n0 ~]$ cdprojet
[login@n0 login]$ cd 13-07-29-panda/tmp/mapping
[login@n0 mapping]$ cat tophat.qsub
#!/bin/bash
#$ -S /bin/bash
#$ -M login@sb-roscoff
#$ -m bea
#$ -V
#$ -cwd
#$ -o qsub.out
#$ -e qsub.err

tophat2 panda_v1[illegible]                    ut/Ill1R1-2.fq
-GTF ../input[illegible]                       0
-num-threads 8

[login@n0 mapping]$ qsub -q long.q -pe thread 8 tophat.qsub
Your job 5338969 ("tophat.qsub") has been submitted
[login@n0 mapping]$ ls
accepted_hits.bam  junctions.bed    qsub.err  unmapped.bam
deletions.bed      logs             qsub.out
insertions.bed     prep_reads.info  tmp
[login@n0 mapping]$ cd
```

Introduction / problematic



Introduction / problematic

Introduction / problematic

Level 1

« I want to know the gene expression »

Introduction / problematic

Level 2

« I want to map my reads on a reference genome and count them »

Introduction / problematic

Level 3

« I want to launch the tools tophat2 and cufflinks.

I have fastq files and my genome in fasta and gtf. »

Introduction / problematic

Level 4

« I want 1TB for my project. I will launch tophat2 through SSH on the cluster in multi-thread mode.

Next I want to submit the bam file to my genome with cufflinks.”

Introduction / problematic

Level 5

« I have a bunch of cool tools!
But I'm the only one who can launch them.

Comments? »

INTRODUCTION / GALAXY

Introduction / Galaxy

Introduction to Galaxy

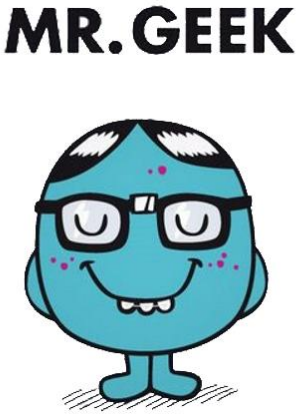
Andrea Bagnacani | Bérénice Batut | Saskia Hiltemann | Anne Pajon
Nicola Soranzo | Helena Rasche | Christopher Barnett | Michele Maroni
Anne Fouilloux | Nadia Goué | Olha Nahorna | Dave Clements

Add Contributions!

Updated: Oct 25, 2021
Or view the plain-text slides without JS
Tip: press P to view the presenter notes

https://training.galaxyproject.org/training-material/topics/introduction/slides/introduction.html

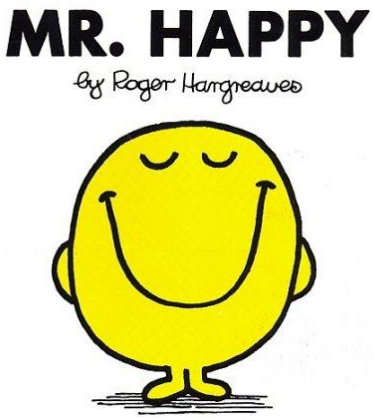
Introduction / Galaxy




```
[login@n0 ~]$ cdprojet
[login@n0 login]$ cd 13-07-29-panda/tmp/mapping
[login@n0 mapping]$ cat tophat.qsub
#!/bin/bash
#$ -S /bin/bash
#$ -M login@sb-roscoff.fr
#$ -m bea
#$ -V
#$ -cwd
#$ -o qsub.out
#$ -e qsub.err

tophat2 panda_v121029 ../input/IllR1-1.fq ../input/IllR1-2.fq
-GTF ../input/panda_v121029.gtf --b2-sensitive -r 100
-num-threads 8

[login@n0 mapping]$ qsub -q long.q -pe thread 8 tophat.qsub
Your job 5338969 ("tophat.qsub") has been submitted
```

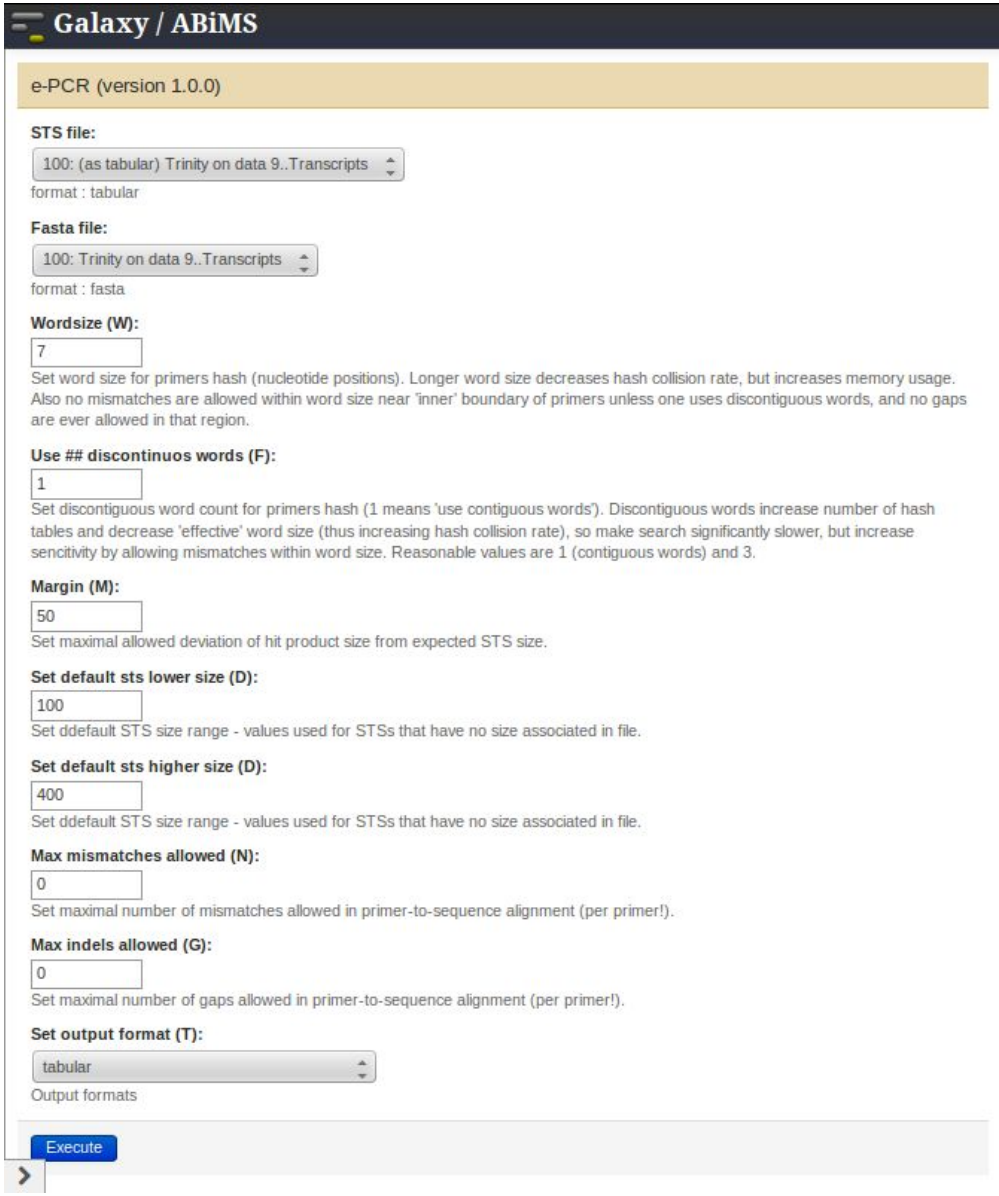
Introduction / Galaxy

```
[login@n0 ~]$ cdprojet
[login@n0 login]$ cd 13-07-29-panda/tmp/mapping
[login@n0 mapping]$ cat tophat.qsub
#!/bin/bash
#$ -S /bin/bash
#$ -M login@sb-roscoff.fr
#$ -m bea
#$ -V
#$ -cwd
#$ -o qsub.out
#$ -e qsub.err

tophat2 panda_v121029 ../input/Ill1R1-1.fq ../input/Ill1R1-2.fq
-GTF ../input/panda_v121029.gtf --b2-sensitive -r 100
-num-threads 8

[login@n0 mapping]$ qsub -q long.q -pe thread 8 tophat.qsub
Your job 5338969 ("tophat.qsub") has been submitted
```



Introduction / Galaxy

```
[lecorguille@n0 ~]$ e-PCR --help
e-PCR: invalid option -- -
usage: [-hV] [posix-options] stsfile [fasta ...]
[compat-options]
where posix-options are:
     -m ##     Margin (default 50)
     -w ##     Wordsize  (default 7)
     -n ##     Max mismatches allowed (default 0)
     -g ##     Max indels allowed (default 0)
     -f ##     Use ## discontiguos words, slow if
                ##>1
     -o ##     Set output file
     -t ##     Set output format:
          1 - classic, range (pos1..pos2)
          2 - classic, midpoint
          3 - tabular
          4 - tabular with alignment in comments
               (slow)
     -d##-##   Set default size range
               (default 100-350)
     -p +-     Turn hits postprocess on/off
     -v ##     Verbosity flags
     -a a|f    Use presize alignmens (only if
               gaps>0), slow
           a - Allways or f - as Fallback
     -x +-     Use 5'-end lowercase masking of
               primers (default -)
     -u +-     Uppercase all primers (default -)

[...]
```

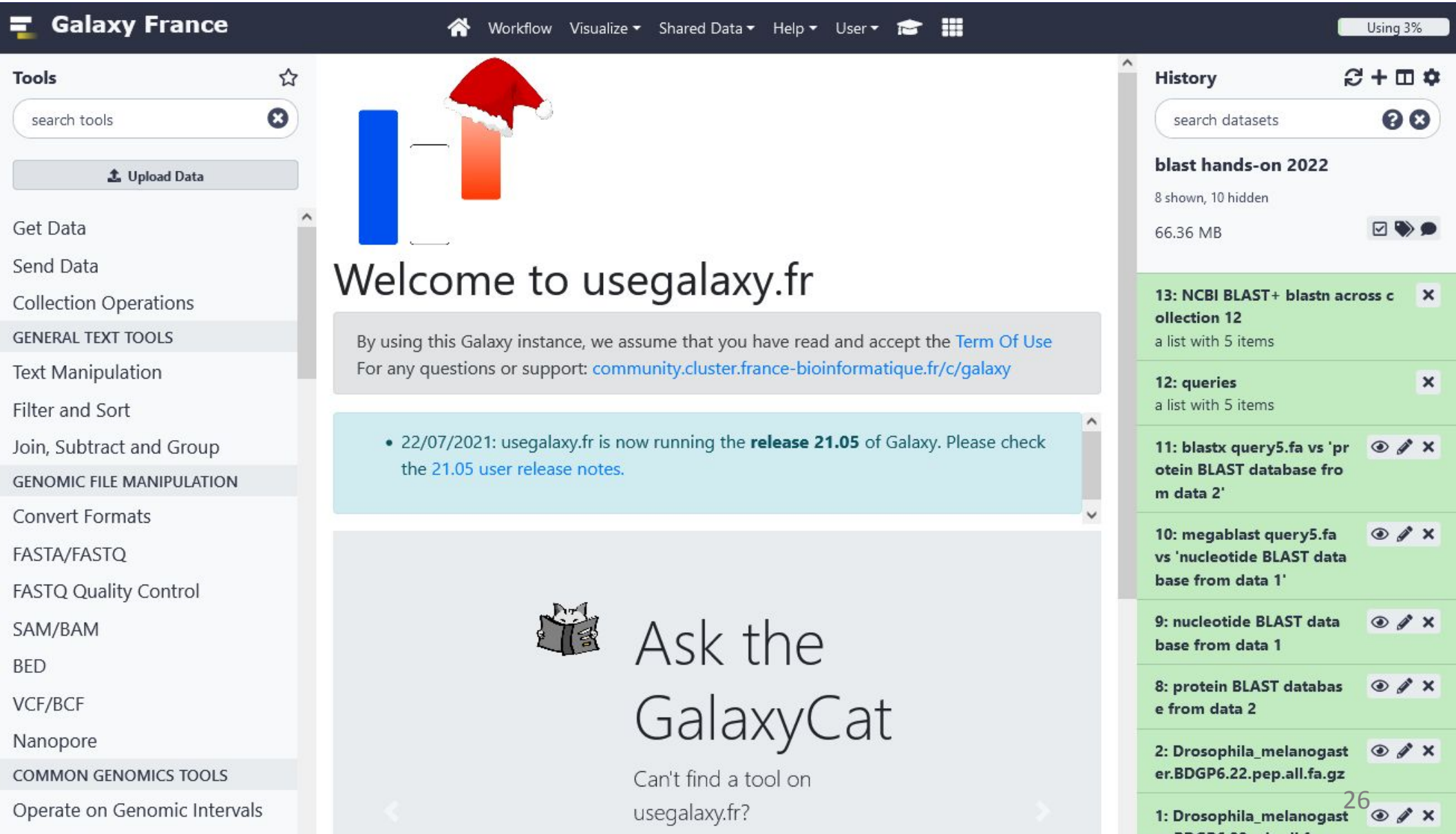


usegalaxy.fr

GALAXY FRANCE

Galaxy France

https://usegalaxy.fr

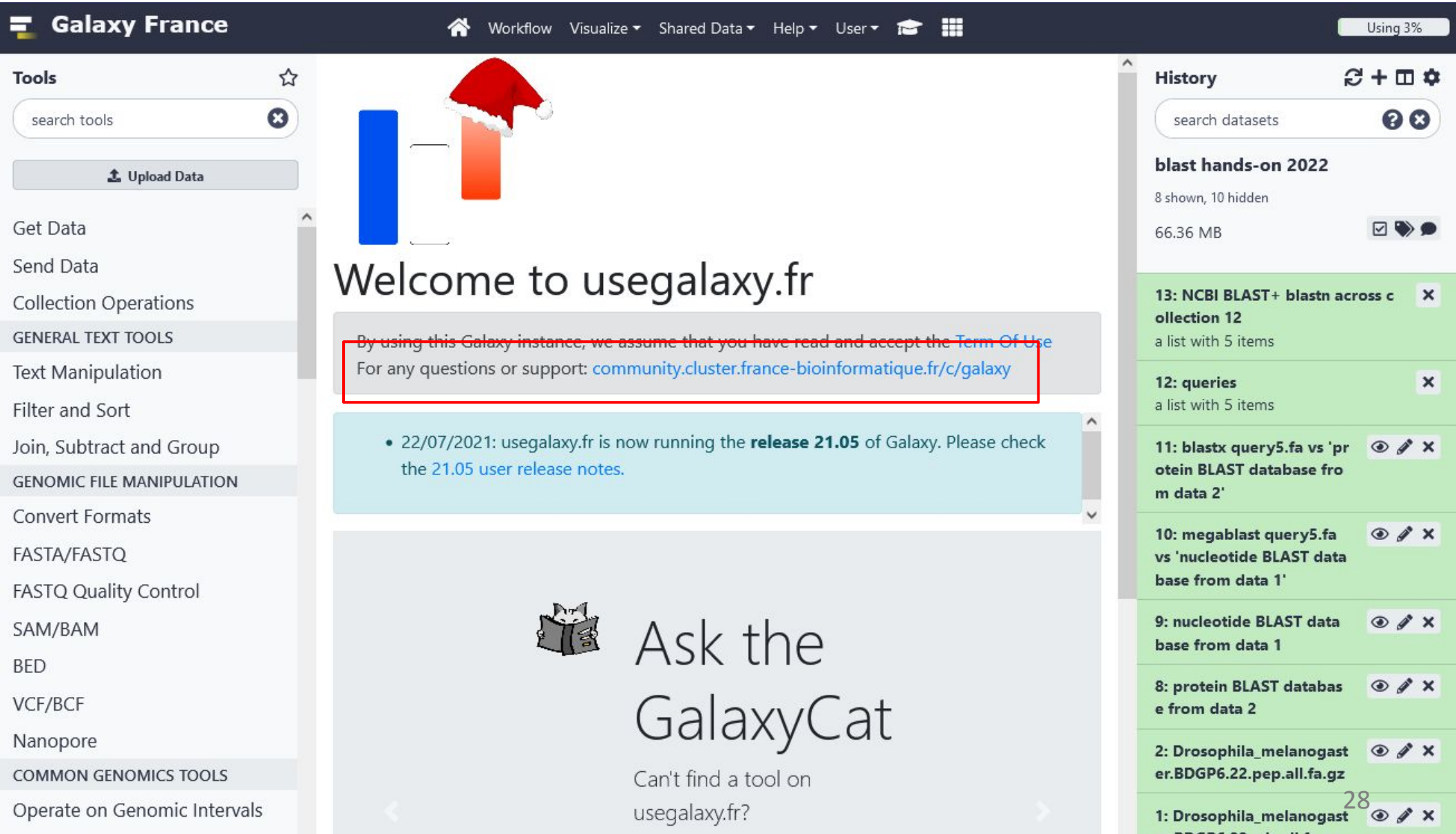


Galaxy ABiMS --> Galaxy France

https://galaxy.sb-roscoff.fr/

https://usegalaxy.fr/

Get help


Galaxy France
Workflow
Visualize
Shared Data
Help
User
Using 3%
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
Welcome to usegalaxy.fr
By using this Galaxy instance, we assume that you have read and accept the Term Of Use
For any questions or support: community.cluster.france-bioinformatique.fr/c/galaxy
22/07/2021: usegalaxy.fr is now running the release 21.05 of Galaxy. Please check the 21.05 user release notes.
Ask the GalaxyCat
Can't find a tool on usegalaxy.fr?
History
search datasets
blast hands-on 2022
8 shown, 10 hidden
66.36 MB
13: NCBI BLAST+ blastn across collection 12
a list with 5 items
12: queries
a list with 5 items
11: blastx query5.fa vs 'protein BLAST database from data 2'
10: megablast query5.fa vs 'nucleotide BLAST database from data 1'
9: nucleotide BLAST database from data 1
8: protein BLAST database from data 2
2: Drosophila_melanogaster.BDGP6.22.pep.all.fa.gz
1: Drosophila_melanogast

Get help: https://community.france-bioinformatique.fr/c/galaxy/8

DATA IMPORT

Data import

Getting data into Galaxy

Anne Pajon Dave Clements Maria Doyle Saskia Hiltemann
Add Contributions!

Updated: Jul 9, 2021
Or view the plain-text slides without JS
Tip: press P to view the presenter notes

https://training.galaxyproject.org/training-material/topics/galaxy-interface/tutorials/get-data/slides.html

Step 1: Choose a FTP Client

DATA IMPORT USING FTP

Data import using FTP

STEP 1: CHOOSE A FTP CLIENT

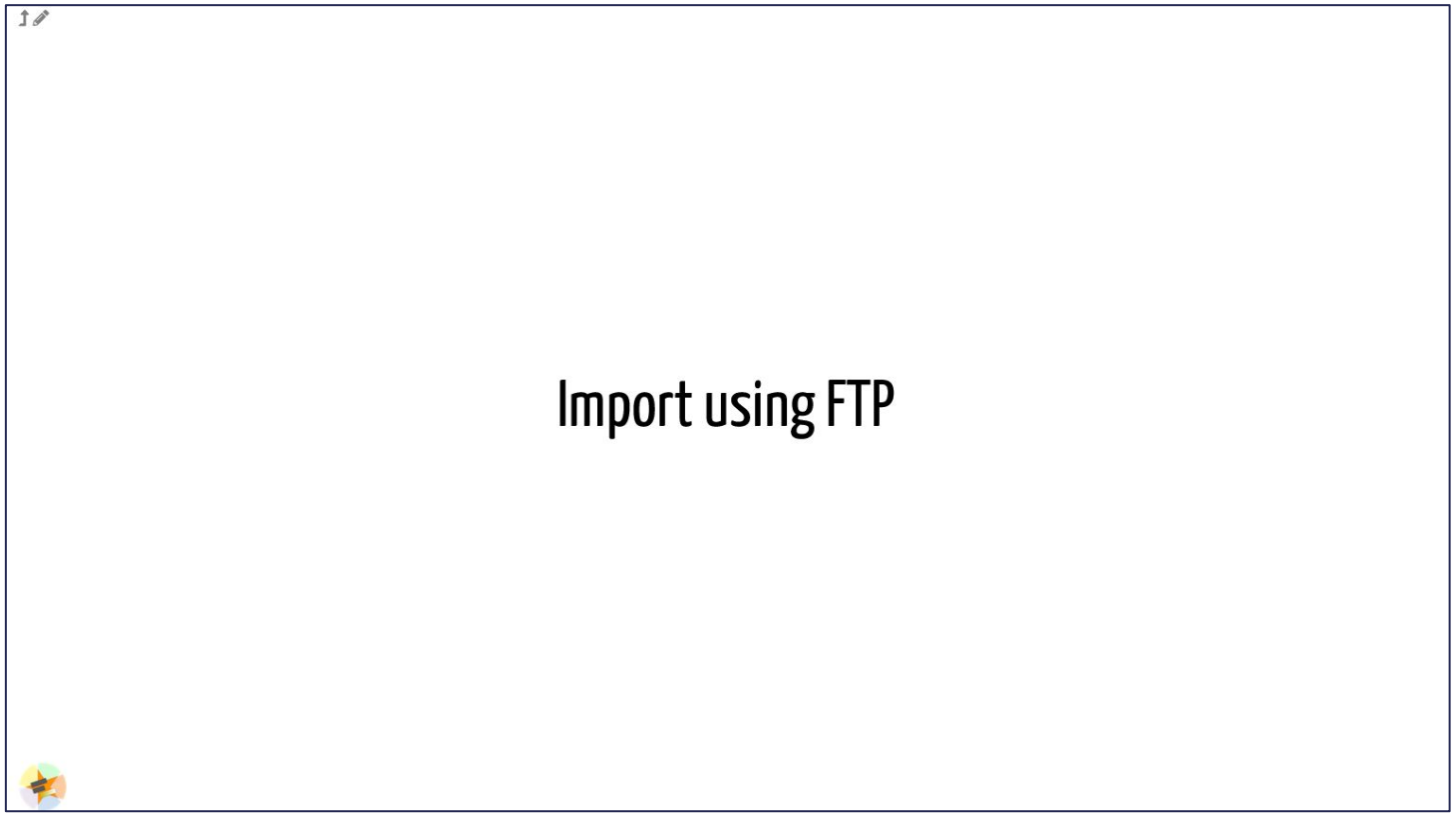
Data import using FTP

STEP 1: CHOOSE A FTP CLIENT

Step 2: Easy!

DATA IMPORT USING FTP

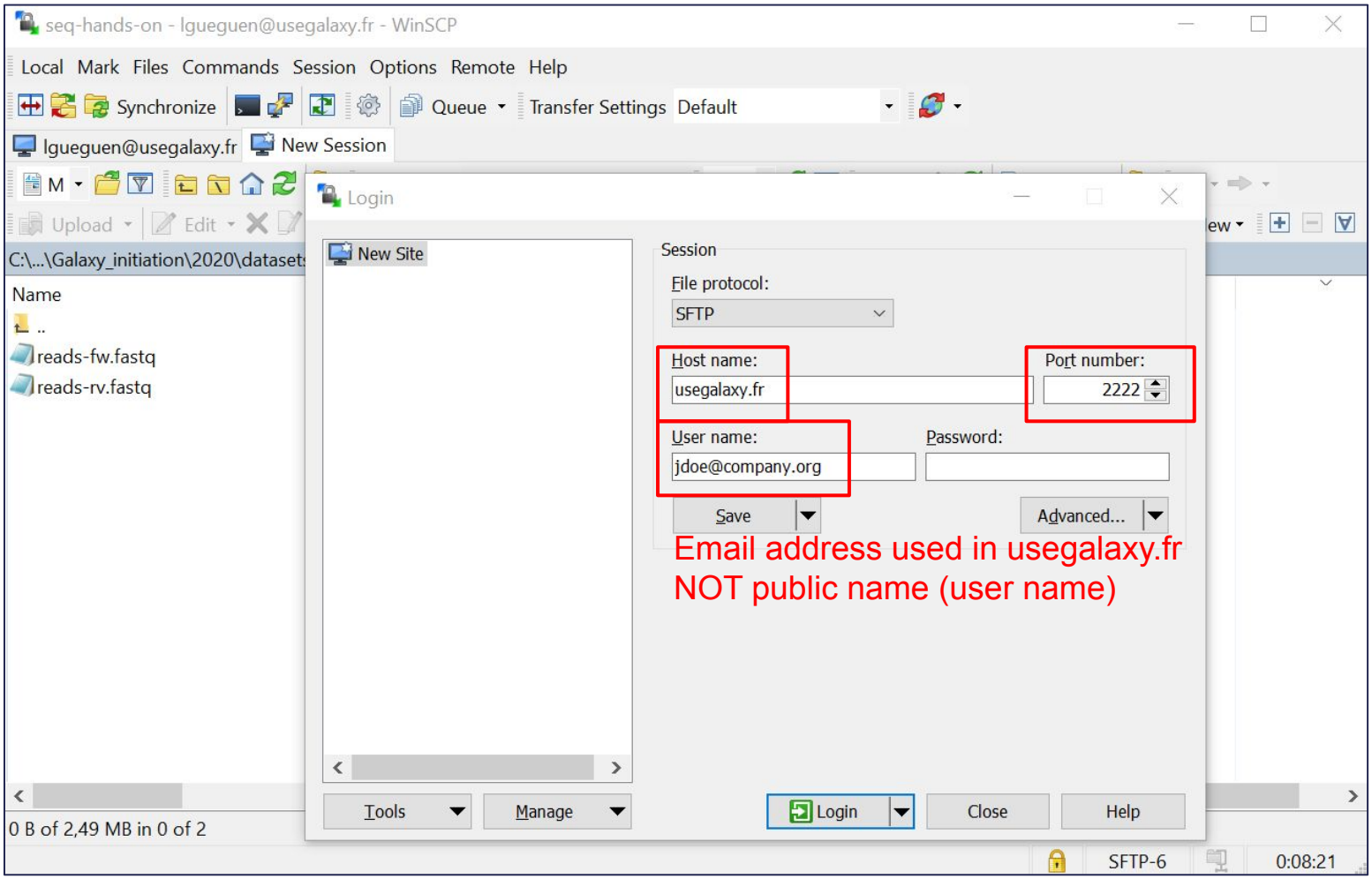
Data import using FTP



https://training.galaxyproject.org/training-material/topics/galaxy-interface/tutorials/get-data/slides.html#28

Data import using FTP

Example with WinSCP for usegalaxy.fr



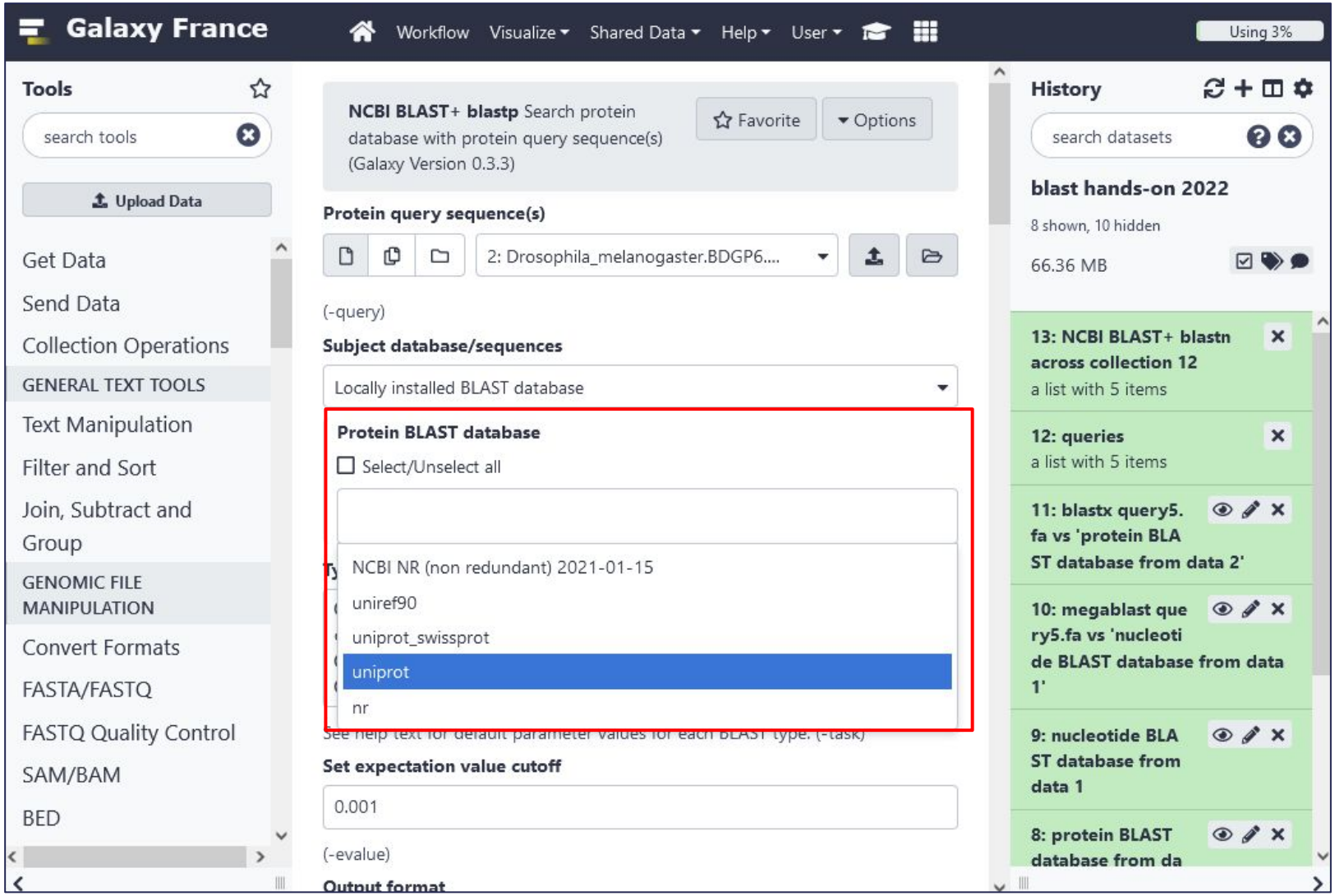
Command line: https://community.france-bioinformatique.fr/t/ftp-usegalaxy-fr/1453

~~DATA IMPORT~~

Data import

For HUGE public resources: genome, databank …

--> Make a request to the support team



Hands-on

DATA IMPORT

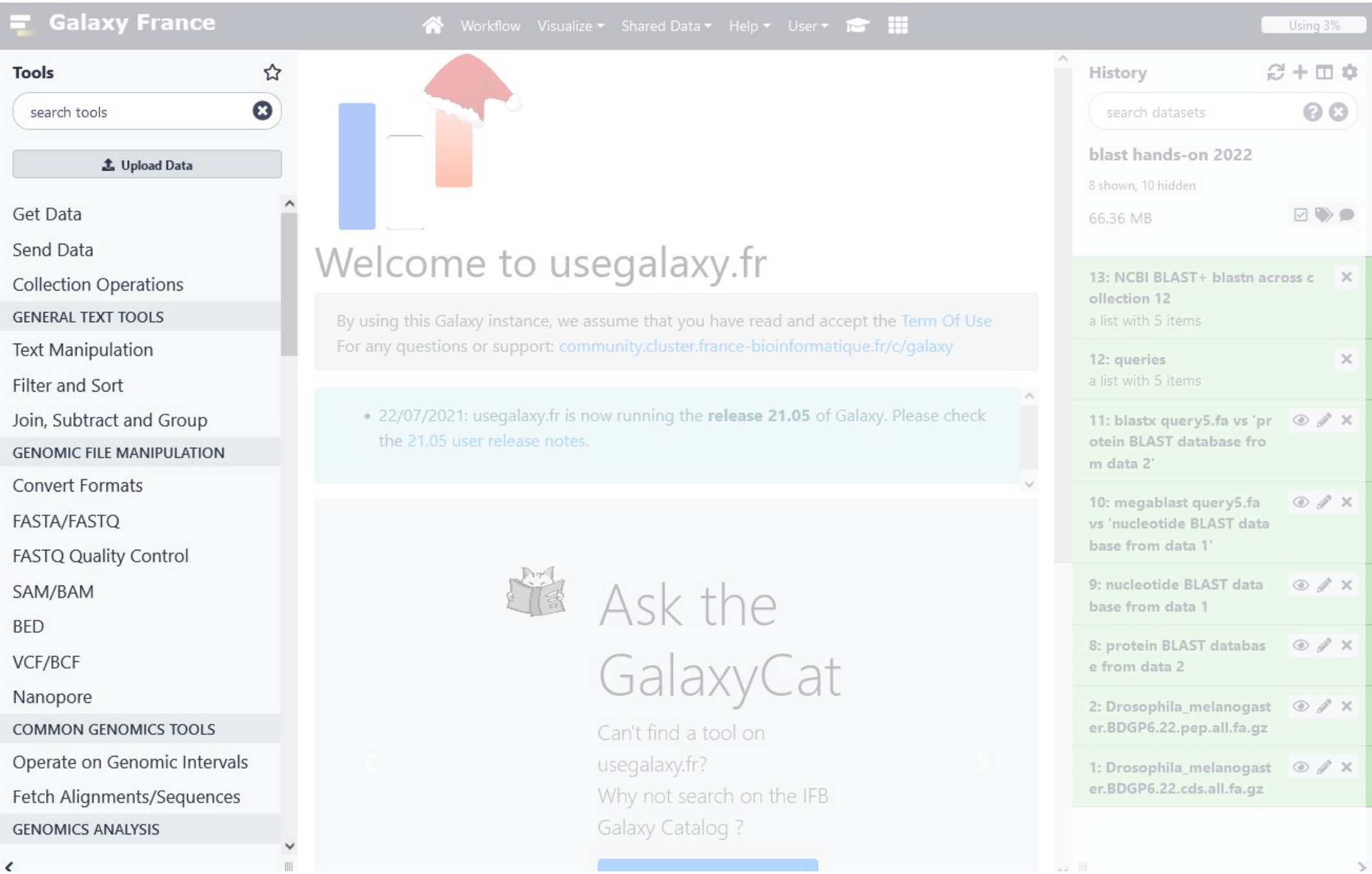
Data import – Hands-on

1. Fetch the files with your internet browser: https://tinyurl.com/glx-init-datasets

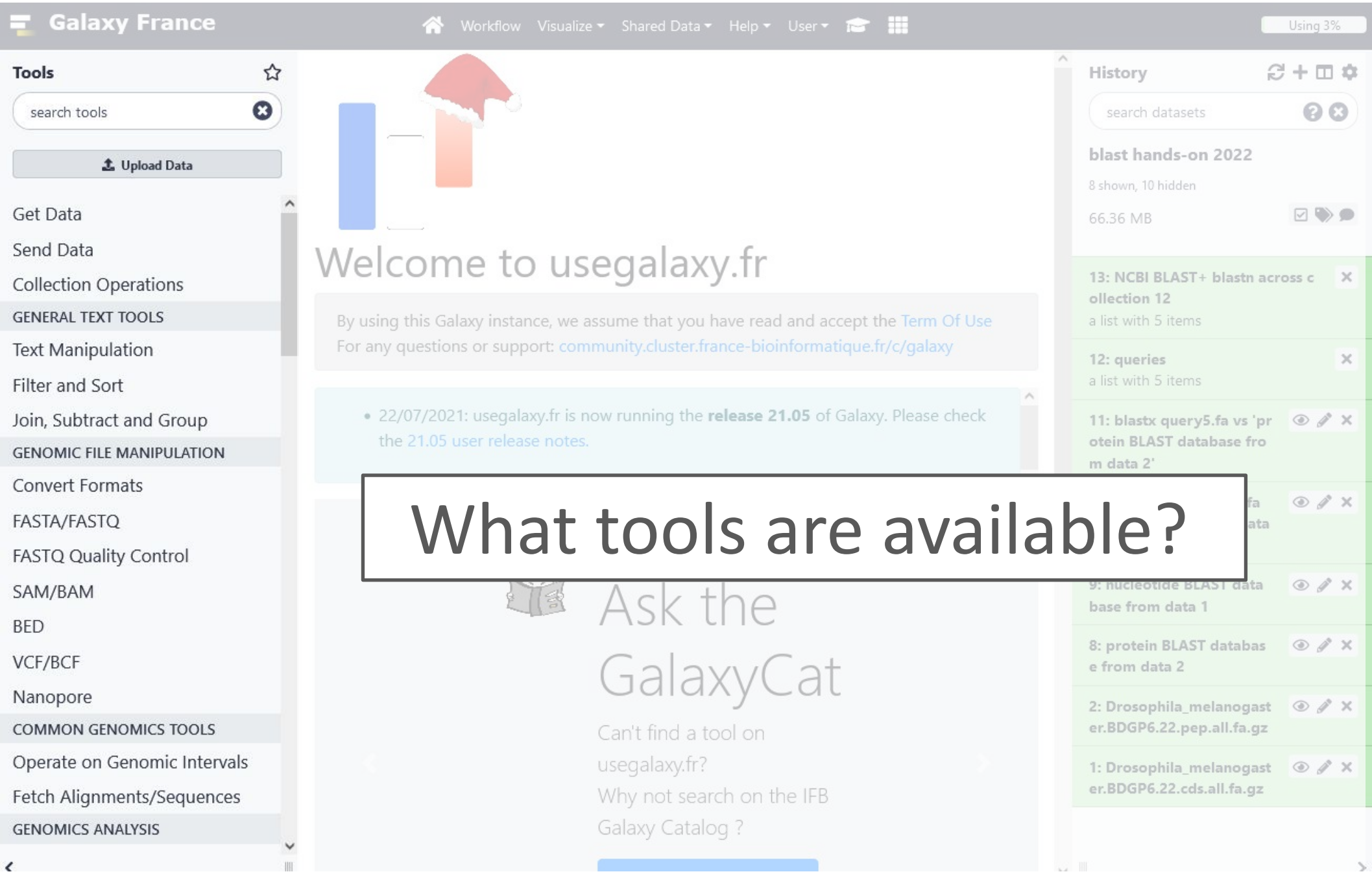
2. Upload files into Galaxy (exons.bed, snps.bed, repeats.bed)
   a. From disk
   b. Using FTP

TOOLS

Tools - panel


Galaxy France
Workflow
Visualize
Shared Data
Help
User
Using 3%
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
Welcome to usegalaxy.fr
By using this Galaxy instance, we assume that you have read and accept the Term Of Use
For any questions or support: community.cluster.france-bioinformatique.fr/c/galaxy
22/07/2021: usegalaxy.fr is now running the release 21.05 of Galaxy. Please check the 21.05 user release notes.
Ask the GalaxyCat
Can't find a tool on usegalaxy.fr?
Why not search on the IFB Galaxy Catalog ?
History
search datasets
blast hands-on 2022
8 shown, 10 hidden
66.36 MB
13: NCBI BLAST+ blastn across collection 12
a list with 5 items
12: queries
a list with 5 items
11: blastx query5.fa vs 'protein BLAST database from data 2'
10: megablast query5.fa vs 'nucleotide BLAST database from data 1'
9: nucleotide BLAST database from data 1
8: protein BLAST database from data 2
2: Drosophila_melanogaster.BDGP6.22.pep.all.fa.gz
1: Drosophila_melanogaster.BDGP6.22.cds.all.fa.gz

Tools - panel



What tools are available?

Tools - panel

130+ platforms for using Galaxy: https://galaxyproject.org/use

[Servers, clouds, deployable resources]

Tools - panel

130+ platforms for using Galaxy: https://galaxyproject.org/use
[Servers, clouds, deployable resources]

GENERAL PURPOSE:

IFB
INSTITUT FRANÇAIS DE BIOINFORMATIQUE

usegalaxy.fr, usegalaxy.org, usegalaxy.eu, usegalaxy.org.au

Tools - panel

130+ platforms for using Galaxy: https://galaxyproject.org/use
[Servers, clouds, deployable resources]

GENERAL PURPOSE:

usegalaxy.fr, usegalaxy.org, usegalaxy.eu, usegalaxy.org.au

DOMAIN SPECIFIC:

Metabarcoding:

Metabolomics:

Proteomics:

SBR-tools:

ChIP-seq:

Tools - panel

130+ platforms for using Galaxy: https://galaxyproject.org/use
[Servers, clouds, deployable resources]

GENERAL PURPOSE:

usegalaxy.fr, usegalaxy.org, usegalaxy.eu, usegalaxy.org.au

DOMAIN SPECIFIC:

Metabarcoding: FROGS

Metabolomics: Workflow4metabolomics

Proteomics: ProteoRE

SBR-tools: ABiMS

ChIP-seq: GalaxEast, nebula

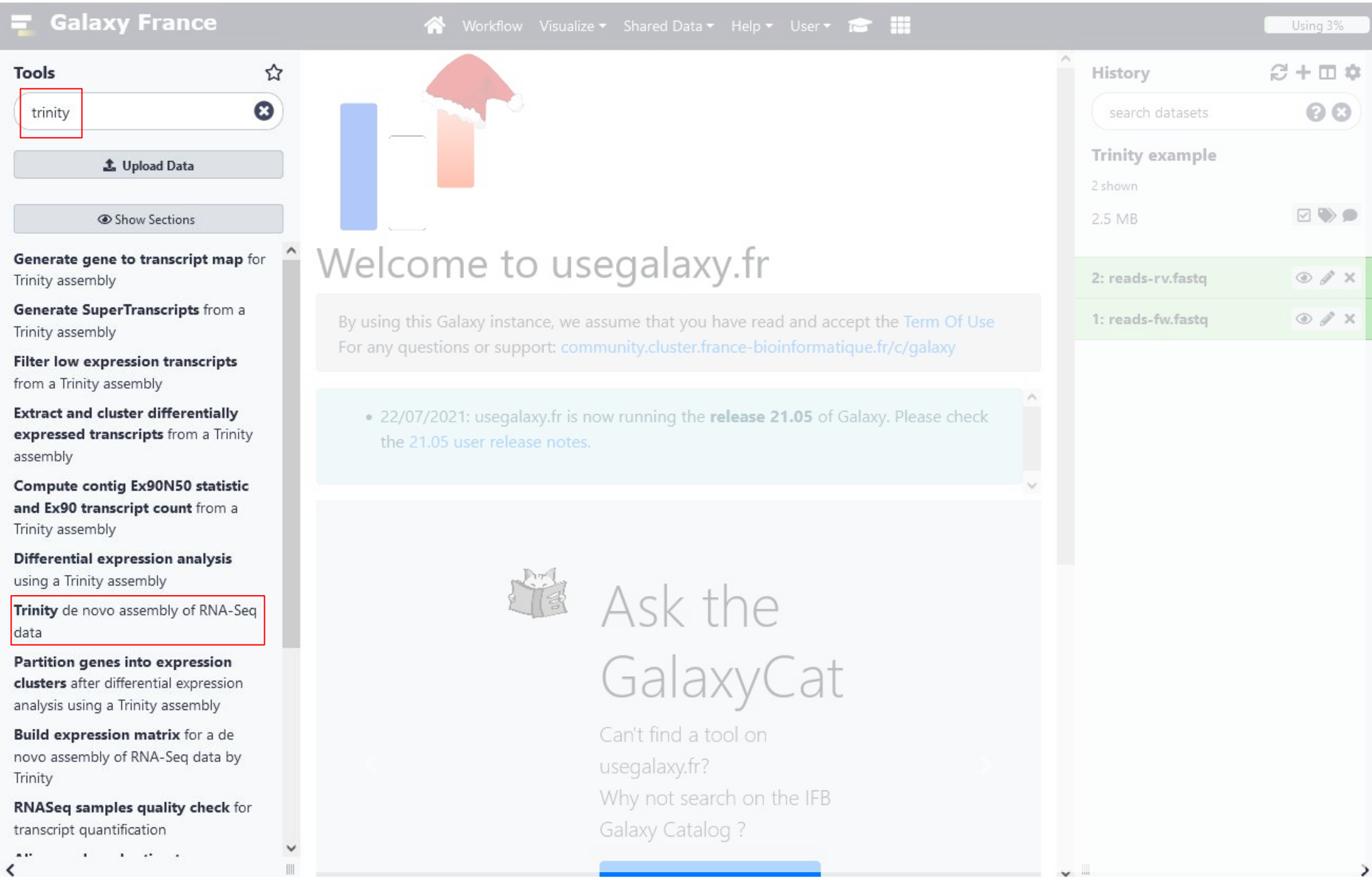
Catalog of French Galaxy tools:
http://galaxycat.france-bioinformatique.fr

Tools - panel

Free "app" store: Galaxy Tool Shed

- Thousands of tools already available
- Most software can be integrated
  - If a tool is not available, ask the Galaxy community for help!
- Only a Galaxy admin can install tools

Tools - panel


Galaxy France
Workflow
Visualize
Shared Data
Help
User
Using 3%
Tools
trinity
Upload Data
Show Sections
Generate gene to transcript map for Trinity assembly
Generate SuperTranscripts from a Trinity assembly
Filter low expression transcripts from a Trinity assembly
Extract and cluster differentially expressed transcripts from a Trinity assembly
Compute contig Ex90N50 statistic and Ex90 transcript count from a Trinity assembly
Differential expression analysis using a Trinity assembly
Trinity de novo assembly of RNA-Seq data
Partition genes into expression clusters after differential expression analysis using a Trinity assembly
Build expression matrix for a de novo assembly of RNA-Seq data by Trinity
RNASeq samples quality check for transcript quantification
Welcome to usegalaxy.fr
By using this Galaxy instance, we assume that you have read and accept the Term Of Use
For any questions or support: community.cluster.france-bioinformatique.fr/c/galaxy
22/07/2021: usegalaxy.fr is now running the release 21.05 of Galaxy. Please check the 21.05 user release notes.
Ask the GalaxyCat
Can't find a tool on usegalaxy.fr?
Why not search on the IFB Galaxy Catalog ?
History
search datasets
Trinity example
2 shown
2.5 MB
2: reads-rv.fastq
1: reads-fw.fastq

Tools - form

Galaxy France | Workflow | Visualize | Shared Data | Help | User | Using 3%

Tools

trinity

Upload Data

Show Sections

Generate gene to transcript map for Trinity assembly

Generate SuperTranscripts from a Trinity assembly

Filter low expression transcripts from a Trinity assembly

Extract and cluster differentially expressed transcripts from a Trinity assembly

Compute contig Ex90N50 statistic and Ex90 transcript count from a Trinity assembly

Differential expression analysis using a Trinity assembly

Trinity de novo assembly of RNA-Seq data

Partition genes into expression clusters after differential expression analysis using a Trinity assembly

Build expression matrix for a de novo assembly of RNA-Seq data by Trinity

RNASeq samples quality check for transcript quantification

Trinity de novo assembly of RNA-Seq data (Galaxy Version 2.9.1+galaxy2) | Favorite | Versions | Options

Are you pooling sequence datasets?

Yes

Paired or Single-end data?

Single-end

Single-end reads

2: reads-rv.fastq
1: reads-fw.fastq

(--single)

Strand specific data

No

Run in silico normalization of reads

Yes

Defaults to max. read coverage of 50. (--no_normalize_reads)

Additional Options

Email notification

No

Send an email notification when the job completes.

Execute

Trinity assembles transcript sequences from Illumina RNA-Seq data.

History

search datasets

Trinity example

2 shown

2.5 MB

2: reads-rv.fastq

1: reads-fw.fastq

Tools - form

Galaxy France | Workflow | Visualize | Shared Data | Help | User | Using 3%

Tools

trinity

Upload Data

Show Sections

Generate gene to transcript map for Trinity assembly

Generate SuperTranscripts from a Trinity assembly

Filter low expression transcripts from a Trinity assembly

Extract and cluster differentially expressed transcripts from a Trinity assembly

Compute contig Ex90N50 statistic and Ex90 transcript count from a Trinity assembly

Differential expression analysis using a Trinity assembly

Trinity de novo assembly of RNA-Seq data

Partition genes into expression clusters after differential expression analysis using a Trinity assembly

Build expression matrix for a de novo assembly of RNA-Seq data by Trinity

RNASeq samples quality check for transcript quantification

Trinity de novo assembly of RNA-Seq data (Galaxy Version 2.9.1+galaxy2) | Favorite | Versions | Options

Are you pooling sequence datasets?

Yes

Paired or Single-end data?

Single-end

Single-end reads

2: reads-rv.fastq
1: reads-fw.fastq

(--single)

Strand specific data

No

Run in silico normalization of reads

Yes

Defaults to max. read coverage of 50. (--no_normalize_reads)

Additional Options

Email notification

No

Send an email notification when the job completes.

Execute

Trinity assembles transcript sequences from Illumina RNA-Seq data.

History

search datasets

Trinity example

2 shown

2.5 MB

2: reads-rv.fastq

1: reads-fw.fastq

Tools

trinity

Upload Data

Show Sections

Generate gene to transcript map for Trinity assembly

Generate SuperTranscripts from a Trinity assembly

Filter low expression transcripts from a Trinity assembly

Extract and cluster differentially expressed transcripts from a Trinity assembly

Compute contig Ex90N50 statistic and Ex90 transcript count from a Trinity assembly

Differential expression analysis using a Trinity assembly

Trinity de novo assembly of RNA-Seq data

Partition genes into expression clusters after differential expression analysis using a Trinity assembly

Build expression matrix for a de novo assembly of RNA-Seq data by Trinity

RNASeq samples quality check for transcript quantification

Run in silico normalization of reads

Yes

Defaults to max. read coverage of 50. (--no_normalize_reads)

Additional Options

Minimum Contig Length

200

All contigs shorter than this will be discarded (--min_contig_length)

Use the genome guided mode?

No

If you already mapped the reads to the genome, Trinity can use this information

Error-corrected or circular consensus (CCS) pac bio reads

No fasta dataset available.

Experimental feature! Long reads must be in the same orientation as short reads if they are strand specific (--long_reads)

Minimum count for K-mers to be assembled

1

(--min_kmer_cov)

Email notification

No

Send an email notification when the job completes.

Execute

Trinity assembles transcript sequences from Illumina RNA-Seq data.

Citations:

History

search datasets

Trinity example

2 shown

2.5 MB

2: reads-rv.fastq

1: reads-fw.fastq

Tools - form

Galaxy France | Workflow | Visualize | Shared Data | Help | User | Using 3%

Tools

trinity

Upload Data

Show Sections

Generate gene to transcript map for Trinity assembly

Generate SuperTranscripts from a Trinity assembly

Filter low expression transcripts from a Trinity assembly

Extract and cluster differentially expressed transcripts from a Trinity assembly

Compute contig Ex90N50 statistic and Ex90 transcript count from a Trinity assembly

Differential expression analysis using a Trinity assembly

Trinity de novo assembly of RNA-Seq data

Partition genes into expression clusters after differential expression analysis using a Trinity assembly

Build expression matrix for a de novo assembly of RNA-Seq data by Trinity

RNASeq samples quality check for transcript quantification

Use the genome guided mode?

No

If you already mapped the reads to the genome, Trinity can use this information

Error-corrected or circular consensus (CCS) pac bio reads

No fasta dataset available.

Experimental feature! Long reads must be in the same orientation as short reads if they are strand specific (--long_reads)

Minimum count for K-mers to be assembled

1

(--min_kmer_cov)

Email notification

No

Send an email notification when the job completes.

Execute

Trinity assembles transcript sequences from Illumina RNA-Seq data.

Citations:

- Grabherr, M. G., Haas, B. J., Yassour, M., Levin, J. Z., Thompson, D. A., Amit, I., ... Regev, A. (2011). Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nature Biotechnology*, *29*(7), 644–652. https://doi.org/10.1038/nbt.1883

Requirements:

- trinity (Version 2.9.1)
- coreutils (Version 8.32)
- rsync (Version 3.2.3)

References:

- bio.tools: trinity (bio.tools) (OpenEBench)

History

search datasets

Trinity example

2 shown

2.5 MB

2: reads-rv.fastq

1: reads-fw.fastq

Tools - form

Galaxy France | Workflow | Visualize | Shared Data | Help | User | Using 3%

Tools

trinity

Upload Data

Show Sections

Generate gene to transcript map for Trinity assembly

Generate SuperTranscripts from a Trinity assembly

Filter low expression transcripts from a Trinity assembly

Extract and cluster differentially expressed transcripts from a Trinity assembly

Compute contig Ex90N50 statistic and Ex90 transcript count from a Trinity assembly

Differential expression analysis using a Trinity assembly

Trinity de novo assembly of RNA-Seq data

Partition genes into expression clusters after differential expression analysis using a Trinity assembly

Build expression matrix for a de novo assembly of RNA-Seq data by Trinity

RNASeq samples quality check for transcript quantification

Trinity de novo assembly of RNA-Seq data (Galaxy Version 2.9.1+galaxy2)

Favorite | Versions | Options

- Switch to 2.9.1+galaxy1
- Switch to 2.9.1
- Switch to 2.8.5

Are you pooling sequence datasets?

Yes

Paired or Single-end data?

Single-end

Single-end reads

2: reads-rv.fastq
1: reads-fw.fastq

(--single)

Strand specific data

No

Run in silico normalization of reads

Yes

Defaults to max. read coverage of 50. (--no_normalize_reads)

Additional Options

Minimum Contig Length

200

All contigs shorter than this will be discarded (--min_contig_length)

Use the genome guided mode?

No

If you already mapped the reads to the genome, Trinity can use this information

History

search datasets

Trinity example

2 shown

2.5 MB

2: reads-rv.fastq

1: reads-fw.fastq

Tools - form

Galaxy France | Workflow | Visualize | Shared Data | Help | User | Using 3%

Tools

trinity

Upload Data

Show Sections

Generate gene to transcript map for Trinity assembly

Generate SuperTranscripts from a Trinity assembly

Filter low expression transcripts from a Trinity assembly

Extract and cluster differentially expressed transcripts from a Trinity assembly

Compute contig Ex90N50 statistic and Ex90 transcript count from a Trinity assembly

Differential expression analysis using a Trinity assembly

Trinity de novo assembly of RNA-Seq data

Partition genes into expression clusters after differential expression analysis using a Trinity assembly

Build expression matrix for a de novo assembly of RNA-Seq data by Trinity

RNASeq samples quality check for transcript quantification

Run in silico normalization of reads

Yes

Defaults to max. read coverage of 50. (--no_normalize_reads)

Additional Options

Minimum Contig Length

200

All contigs shorter than this will be discarded (--min_contig_length)

Use the genome guided mode?

No

If you already mapped the reads to the genome, Trinity can use this information

Error-corrected or circular consensus (CCS) pac bio reads

No fasta dataset available.

Experimental feature! Long reads must be in the same orientation as short reads if they are strand specific (--long_reads)

Minimum count for K-mers to be assembled

1

(--min_kmer_cov)

Email notification

No

Send an email notification when the job completes.

Execute

Trinity assembles transcript sequences from Illumina RNA-Seq data.

Citations:

History

search datasets

Trinity example

2 shown

2.5 MB

2: reads-rv.fastq

1: reads-fw.fastq

Tools - form

Galaxy France | Workflow | Visualize | Shared Data | Help | User | Using 3%

Tools

trinity

Upload Data

Show Sections

Generate gene to transcript map for Trinity assembly

Generate SuperTranscripts from a Trinity assembly

Filter low expression transcripts from a Trinity assembly

Extract and cluster differentially expressed transcripts from a Trinity assembly

Compute contig Ex90N50 statistic and Ex90 transcript count from a Trinity assembly

Differential expression analysis using a Trinity assembly

Trinity de novo assembly of RNA-Seq data

Partition genes into expression clusters after differential expression analysis using a Trinity assembly

Build expression matrix for a de novo assembly of RNA-Seq data by Trinity

RNASeq samples quality check for transcript quantification

Executed **Trinity** and successfully added 1 job to the queue.

The tool uses 2 inputs:

- **1: reads-fw.fastq**
- **2: reads-rv.fastq**

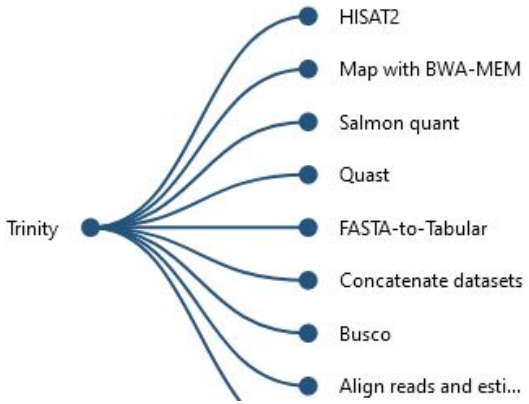
It produces 2 outputs:

- **3: Trinity on data 2 and data 1: Assembled Transcripts**
- **4: Trinity on data 2 and data 1: Gene to transcripts map**

You can check the status of queued jobs and view the resulting data by refreshing the History panel. When the job has been run the status will change from 'running' to 'finished' if completed successfully or 'error' if problems were encountered.

Tool recommendation

You have used trinity tool. For further analysis, you could try using the following/recommended tools. The recommended tools are shown in the decreasing order of their scores predicted using machine learning analysis on workflows. Therefore, tools at the top may be more useful than the ones at the bottom. Please click on one of the following/recommended tools to open its definition.



History

search datasets

Trinity example

4 shown

2.5 MB

4: Trinity on data 2 and data 1: Gene to transcripts map

3: Trinity on data 2 and data 1: Assembled Transcripts

2: reads-rv.fastq

1: reads-fw.fastq

Tools - Job status

Job is waiting to run

= the job is in the scheduler « queue »

Duration time of this status depends on the amount of actual queued jobs and on the requested number of processors

Tools - Job status

Job is currently running

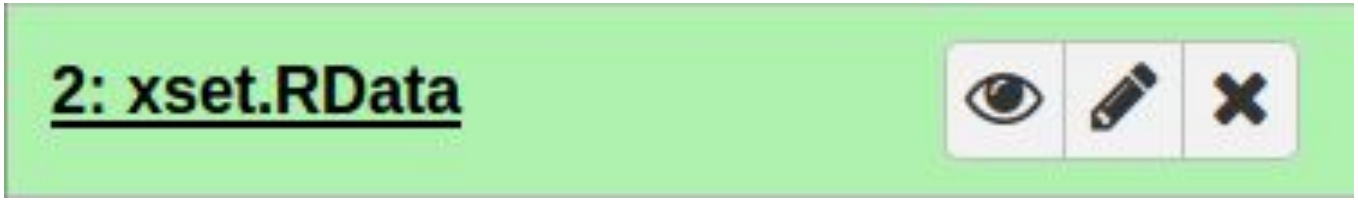
= the job is being executed on the computing cluster

Duration time of this status depends on the job's attributes and the computing resources allocated.

Some programs are executed with several processors (using 4, 8 or 16 Gb of RAM).

And others are mono-threaded ☹

Tools - Job status



Job has finished successfully

But warnings or errors can be hidden behind!

Tools - Job status

Job has finished but with an error status

= the program sends an error

The error is often explained by the program but sometimes … not.

Tools - Job status

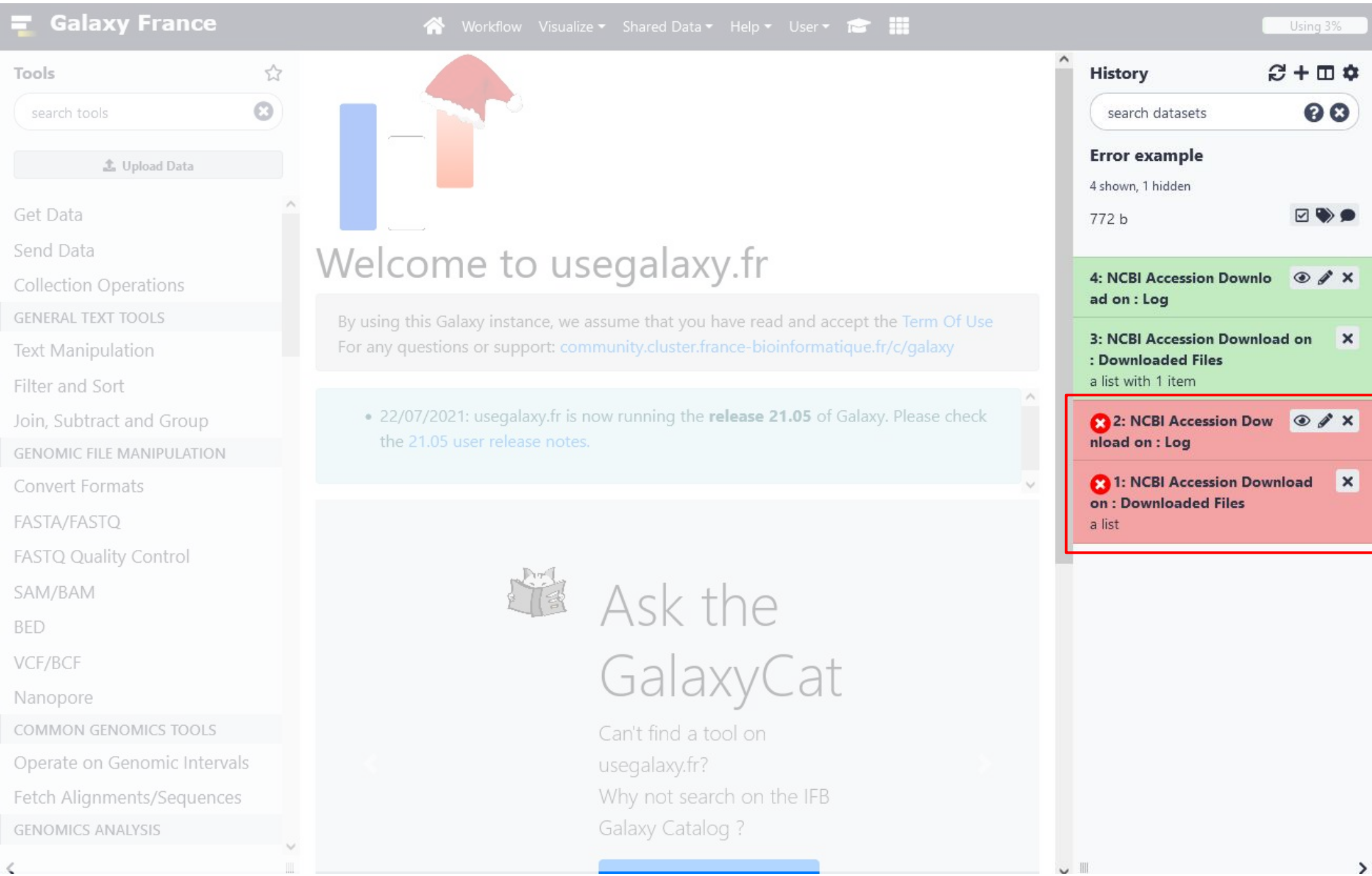
Job has finished but with an error status

= the program sends an error

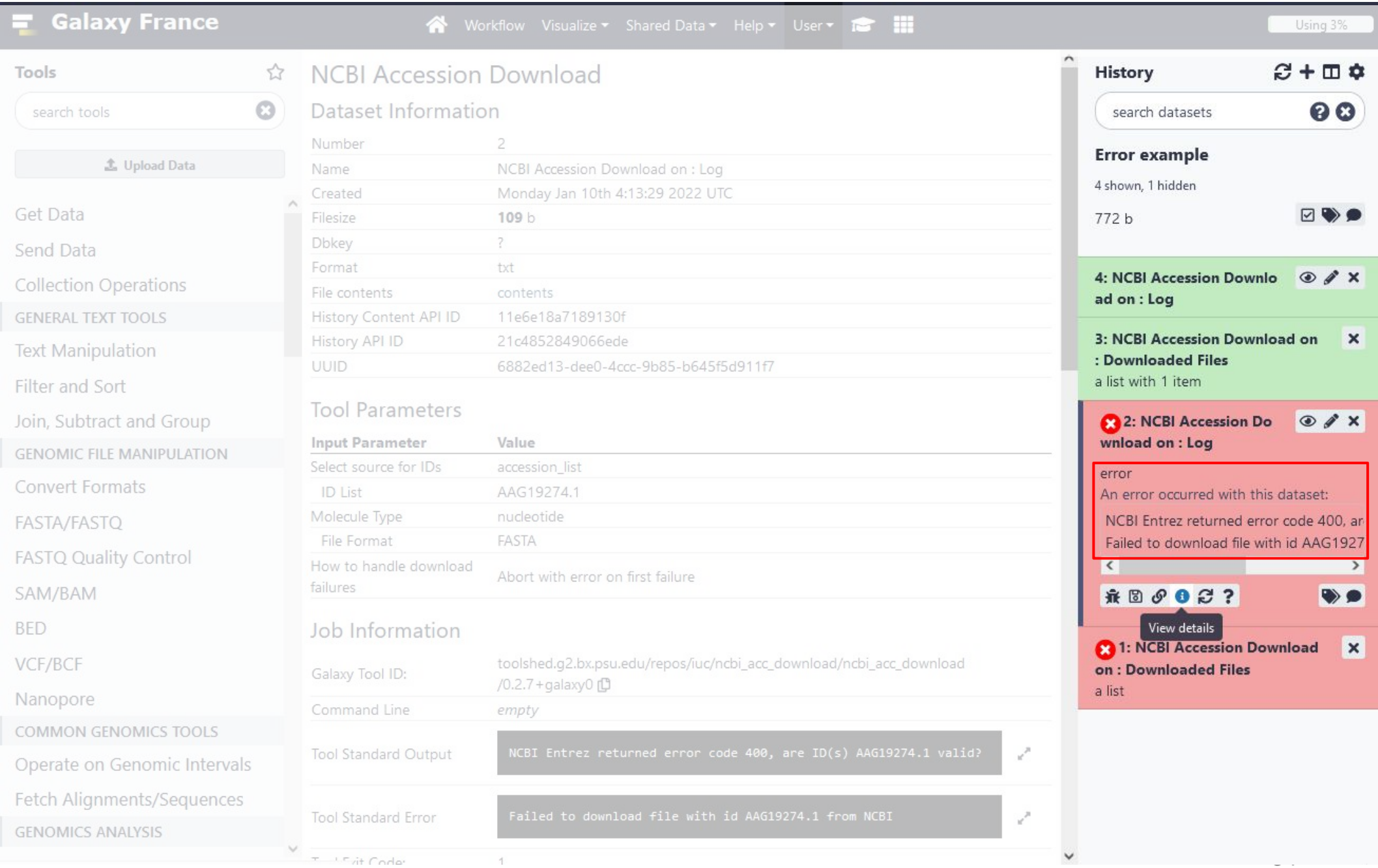
Possible causes of error :

- Bad usage : input file, format or option
- Bad integration of the program into Galaxy (exceeded memory allocation…)
- Non anticipated crash of the program

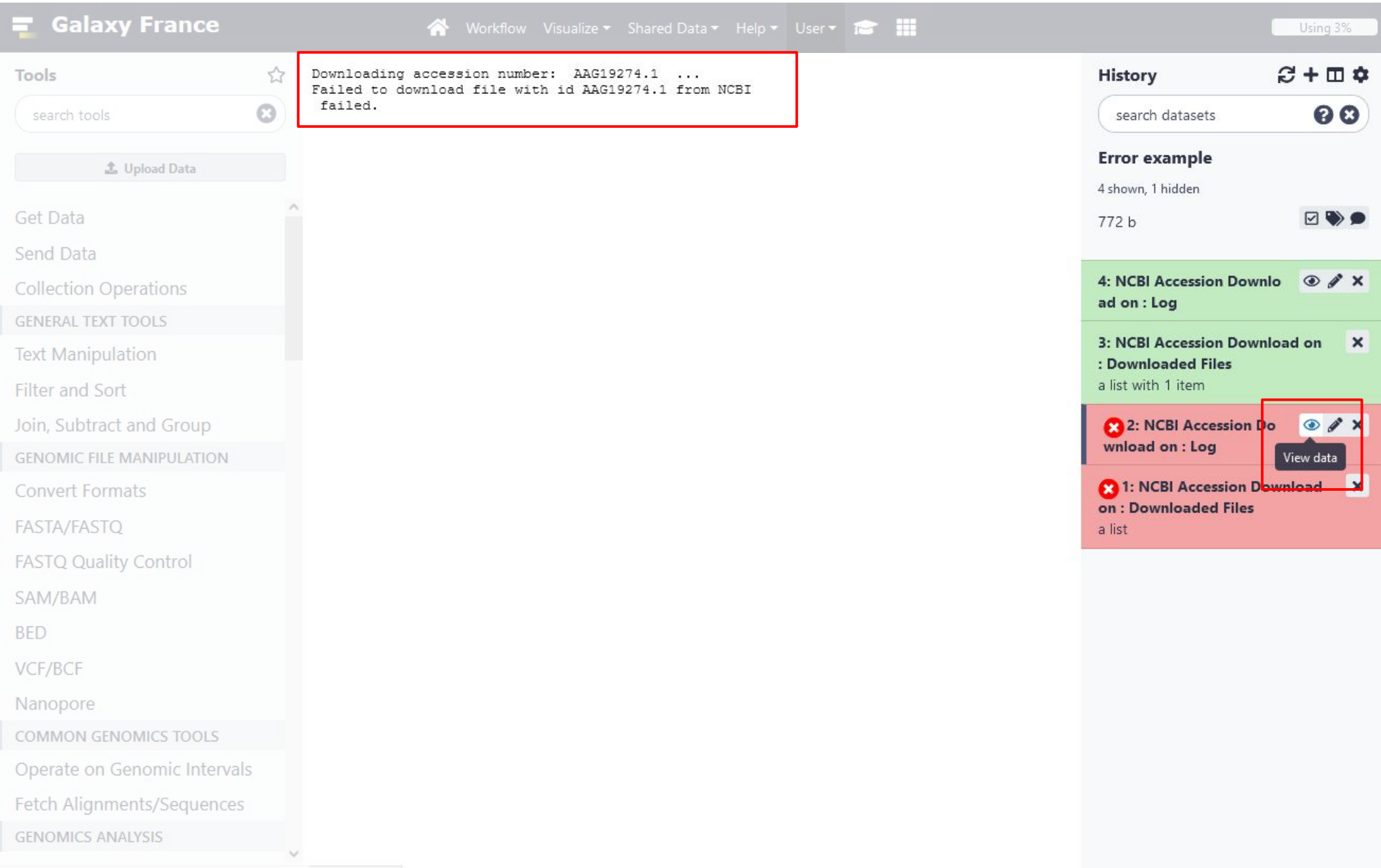
Tools - Handle errors



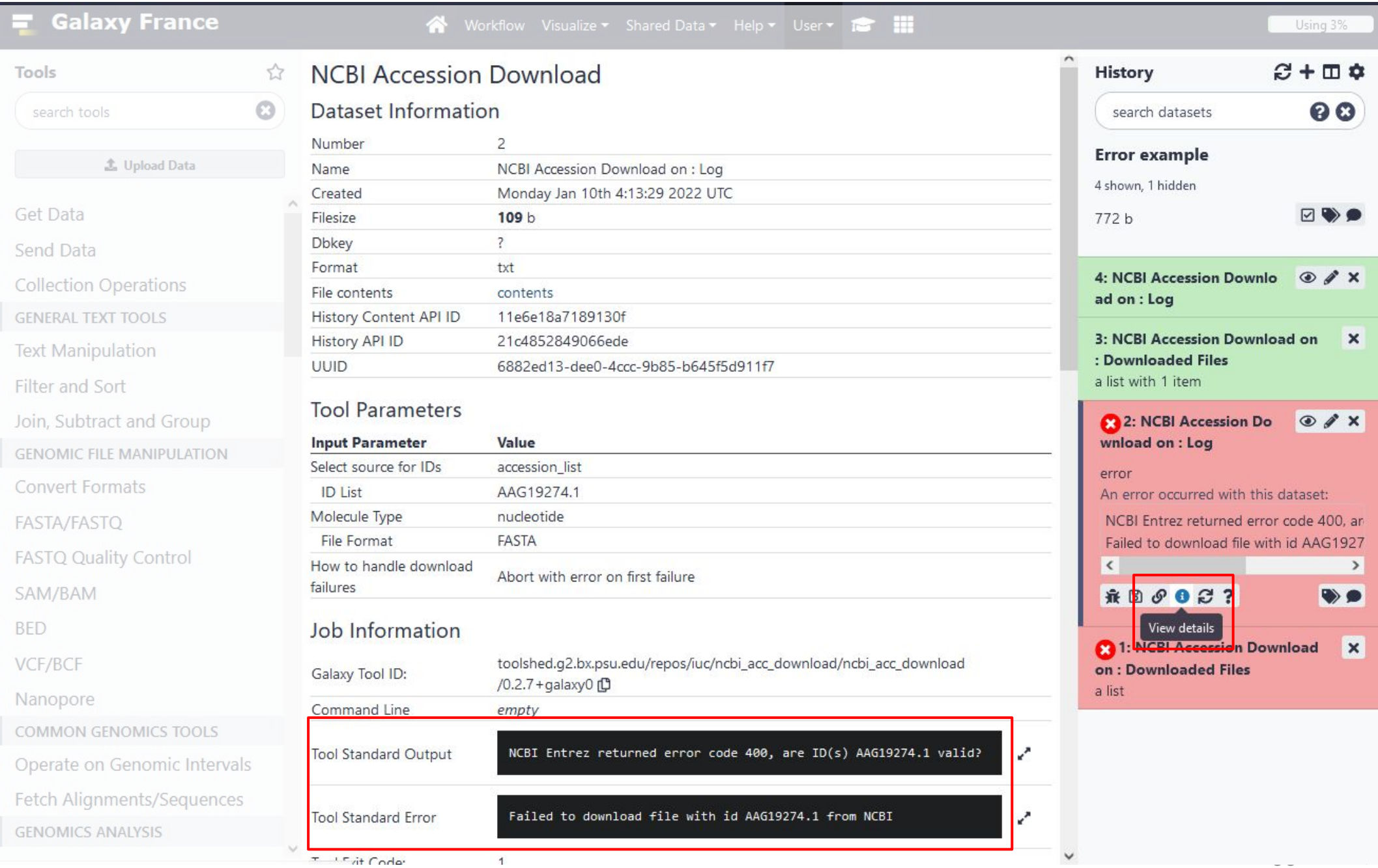
Tools - Handle errors



Tools - Handle errors



Tools - Handle errors



Tools - Handle errors

Galaxy France | Workflow | Visualize | Shared Data | Help | User | Using 3%

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

Dataset Error Report

An error occurred while running the tool **toolshed.g2.bx.psu.edu/repos/iuc/ncbi_acc_download/ncbi_acc_download/0.2.7+galaxy0**.

Details

Execution resulted in the following messages:

```
Fatal error: Exit code 1 ()
```

Tool generated the following standard error:

```
Failed to download file with id AAG19274.1 from NCBI
```

Troubleshooting

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

Issue Report

Please provide detailed information on the activities leading to this issue:

Report

History

search datasets

Error example

4 shown, 1 hidden

772 b

4: NCBI Accession Download on : Log

3: NCBI Accession Download on : Downloaded Files
a list with 1 item

2: NCBI Accession Download on : Log
error
An error occurred with this dataset:
NCBI Entrez returned error code 400, ar
Failed to download file with id AAG1927

View or report this error

1: NCBI Accession Download on : Downloaded Files
a list

Tools - Handle errors

Galaxy France | Workflow | Visualize | Shared Data | Help | User | Using 3%

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

Dataset Error Report

An error occurred while running the tool **toolshed.g2.bx.psu.edu/repos/iuc/ncbi_acc_download/ncbi_acc_download/0.2.7+galaxy0**.

Details

Execution resulted in the following messages:

```
Fatal error: Exit code 1 ()
```

Tool generated the following standard error:

```
Failed to download file with id AAG19274.1 from NCBI
```

Troubleshooting

https://galaxyproject.org/support/tool-error/

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

Issue Report

Please provide detailed information on the activities leading to this issue:

Report

History

search datasets

Error example

4 shown, 1 hidden

772 b

4: NCBI Accession Download on : Log

3: NCBI Accession Download on : Downloaded Files
a list with 1 item

2: NCBI Accession Download on : Log
error
An error occurred with this dataset:
NCBI Entrez returned error code 400, ar
Failed to download file with id AAG1927

View or report this error

1: NCBI Accession Download on : Downloaded Files
a list

Tools - Handle errors

Galaxy France | Workflow | Visualize | Shared Data | Help | User | Using 3%

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

Dataset Error Report

An error occurred while running the tool **toolshed.g2.bx.psu.edu/repos/iuc/ncbi_acc_download/ncbi_acc_download/0.2.7+galaxy0**.

Details

Execution resulted in the following messages:

```
Fatal error: Exit code 1 ()
```

Tool generated the following standard error:

```
Failed to download file with id AAG19274.1 from NCBI
```

Troubleshooting

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

Issue Report

Please provide detailed information on the activities leading to this issue:

Report

History

search datasets

Error example

4 shown, 1 hidden

772 b

4: NCBI Accession Download on : Log

3: NCBI Accession Download on : Downloaded Files
a list with 1 item

2: NCBI Accession Download on : Log
error
An error occurred with this dataset:
NCBI Entrez returned error code 400, ar
Failed to download file with id AAG1927

View or report this error

1: NCBI Accession Download on : Downloaded Files
a list

Bug report sent to the local galaxy administrators

Tools - Handle errors

Galaxy France | Workflow | Visualize | Shared Data | Help | User | Using 3%

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

Dataset Error Report

An error occurred while running the tool **toolshed.g2.bx.psu.edu/repos/iuc/ncbi_acc_download/ncbi_acc_download/0.2.7+galaxy0**.

Details

Execution resulted in the following messages:

```
Fatal error: Exit code 1 ()
```

Tool generated the following standard error:

```
Failed to download file with id AAG19274.1 from NCBI
```

Troubleshooting

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

Issue Report

Please provide detailed information on the activities leading to this issue:

Report

History

search datasets

Error example

4 shown, 1 hidden

772 b

4: NCBI Accession Download on : Log

3: NCBI Accession Download on : Downloaded Files
a list with 1 item

2: NCBI Accession Download on : Log
error
An error occurred with this dataset:
NCBI Entrez returned error code 400, ar
Failed to download file with id AAG1927

View or report this error

1: NCBI Accession Download on : Downloaded Files
a list

For usegalaxy.fr, create a new request on
https://community.france-bioinformatique.fr/c/galaxy

HISTORY AND DATASETS

History

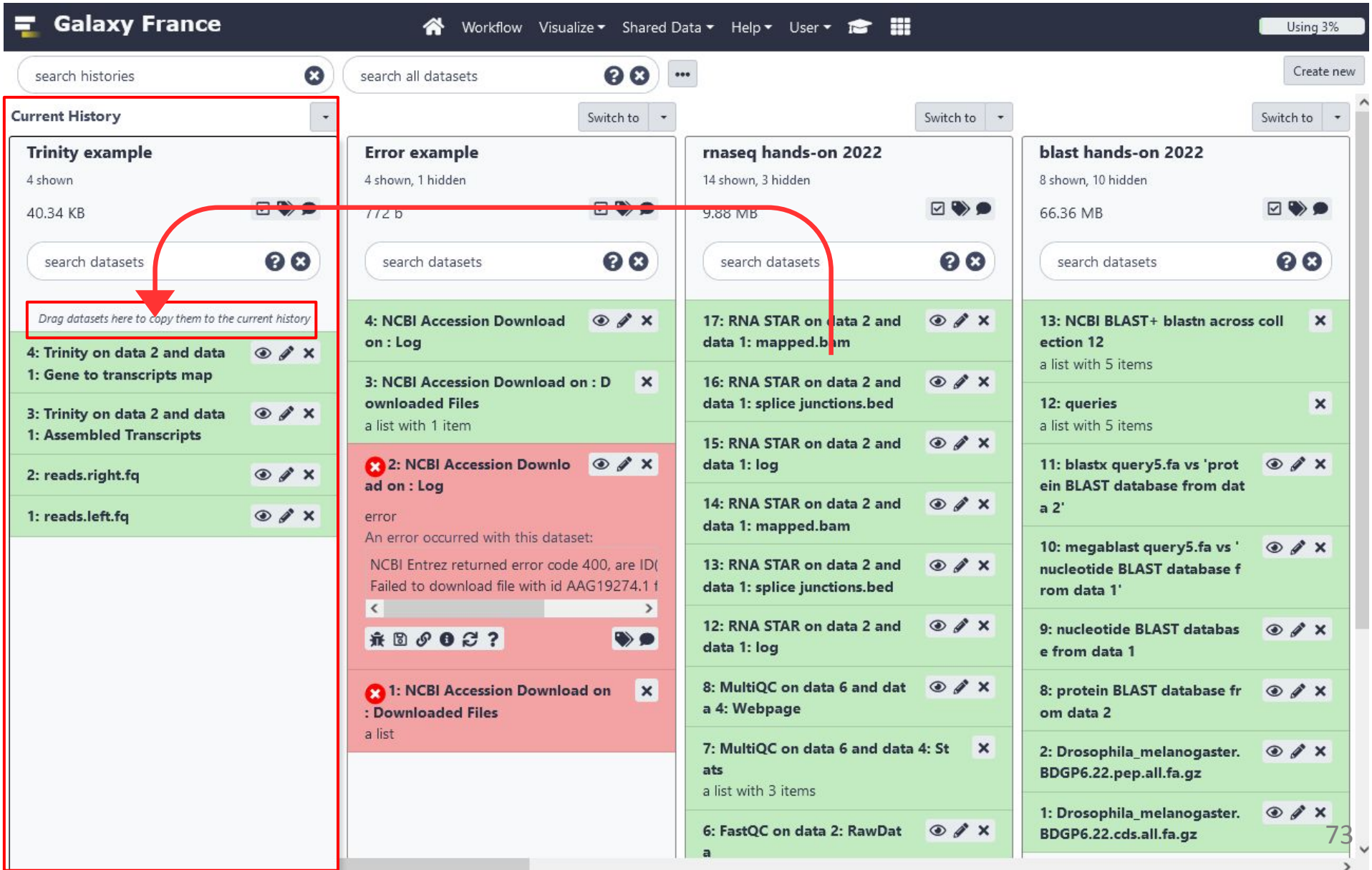
History

- Location of all analyses
  - collects all datasets produced by tools
  - collects all operations performed on the data
- For each dataset (the heart of Galaxy's reproducibility), the history tracks
  - name, format, size, creation time, datatype-specific metadata
  - tool id, version, inputs, parameters
  - standard output (`stdout`) and error (`stderr`)
  - state (waiting, running, success, failed)
  - hidden, deleted, purged

https://training.galaxyproject.org/training-material/topics/introduction/slides/introduction.html#12

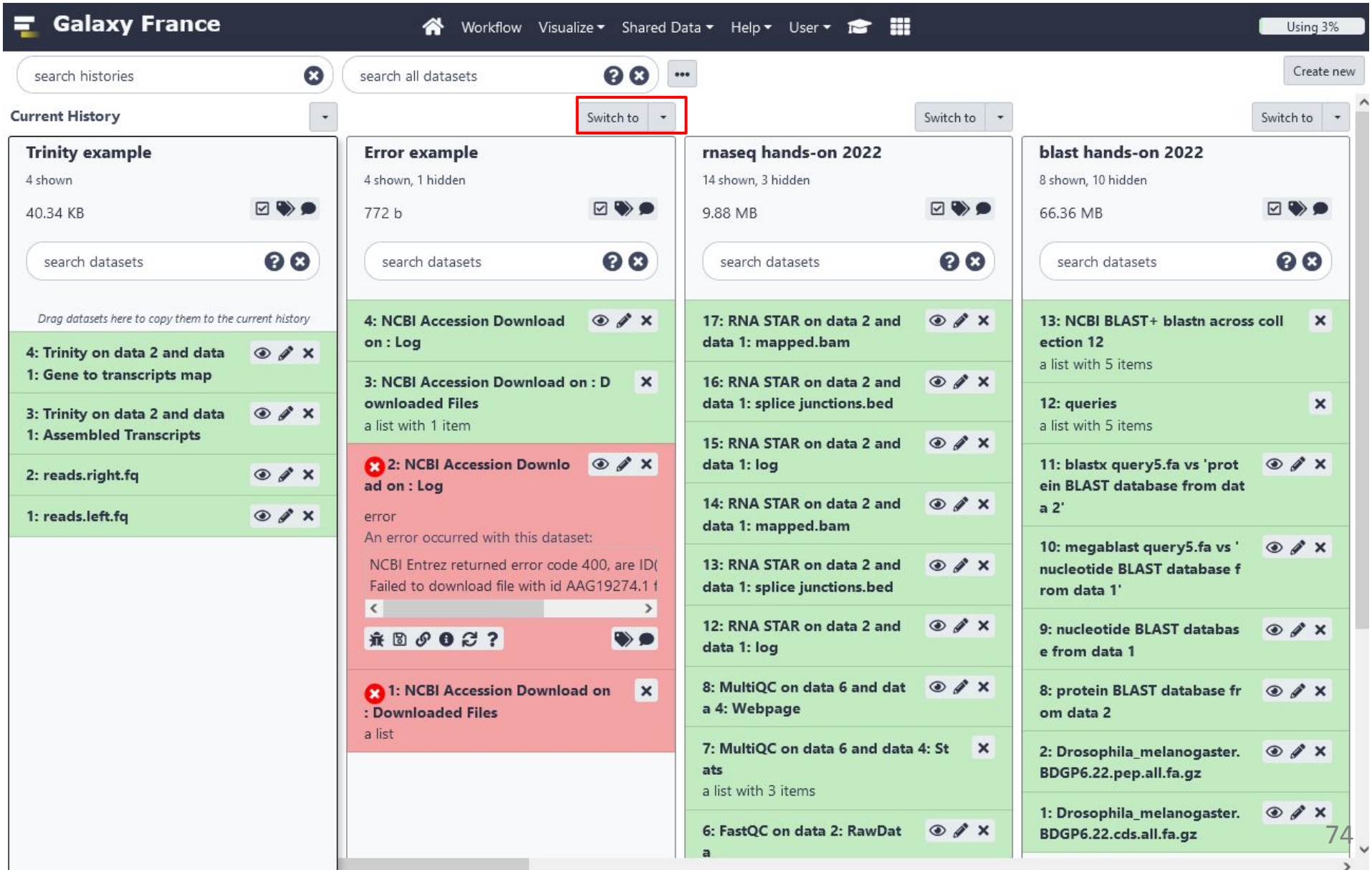
History

Current history, drag and drop datasets

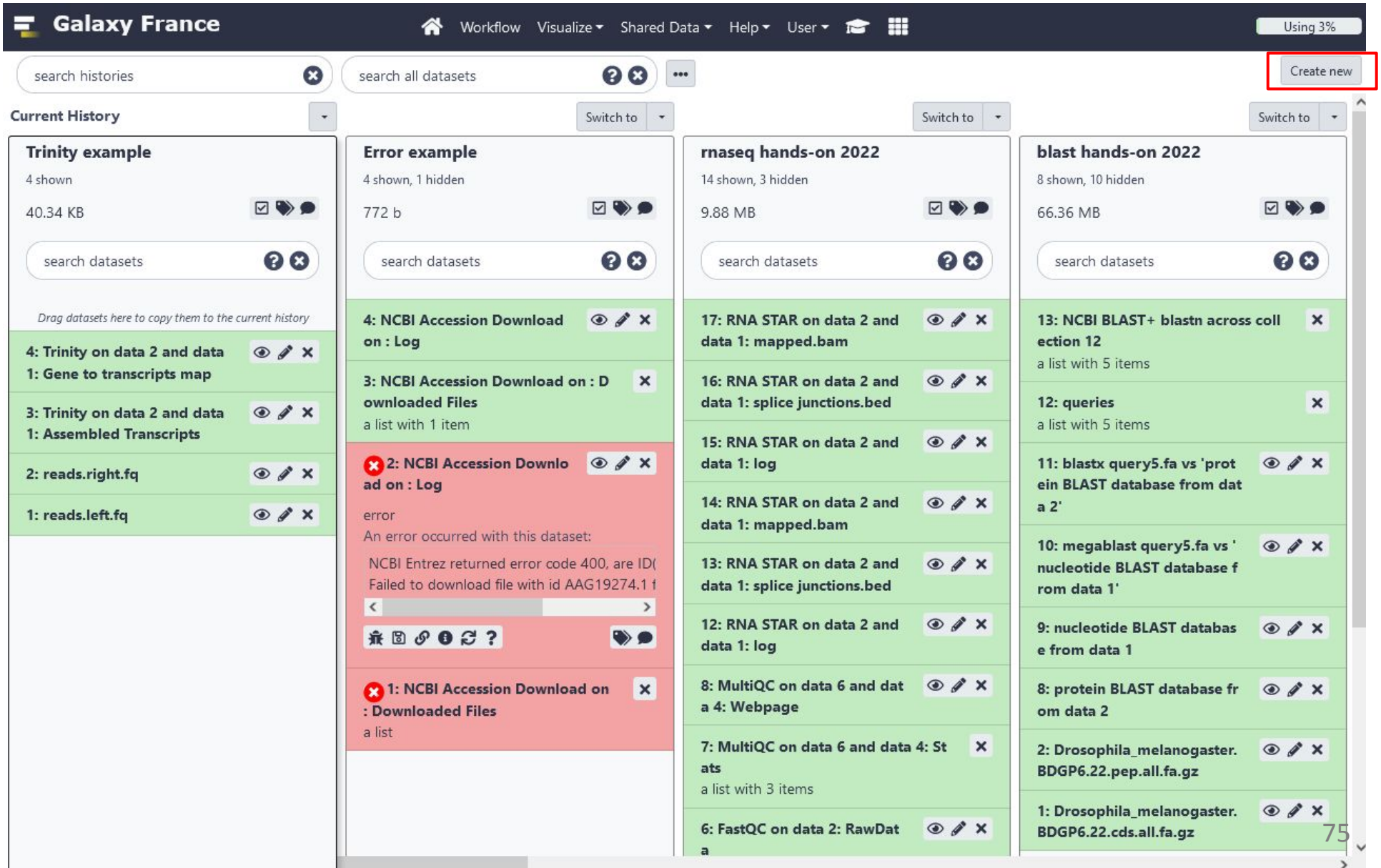


History

Switch

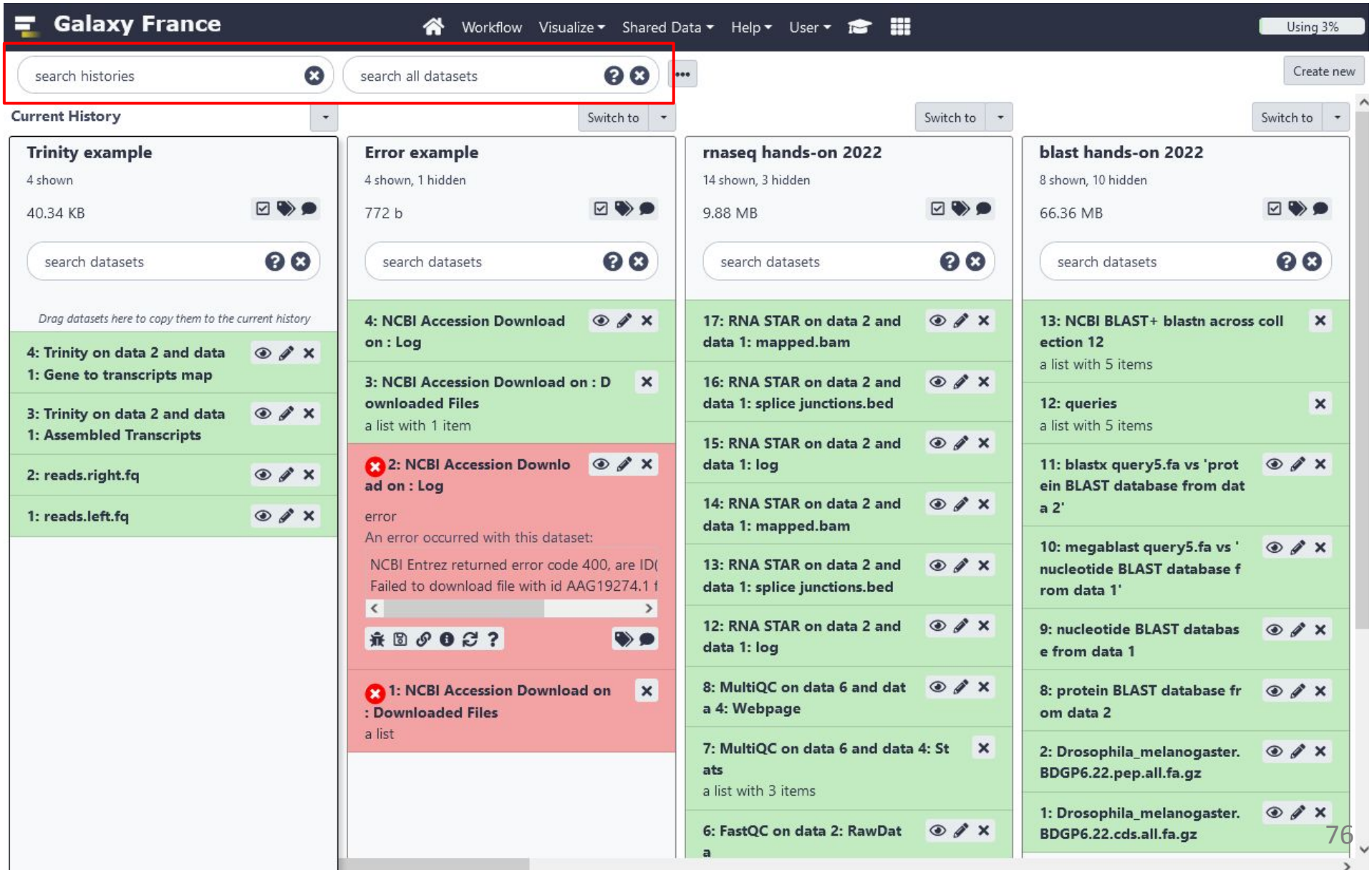


Create new

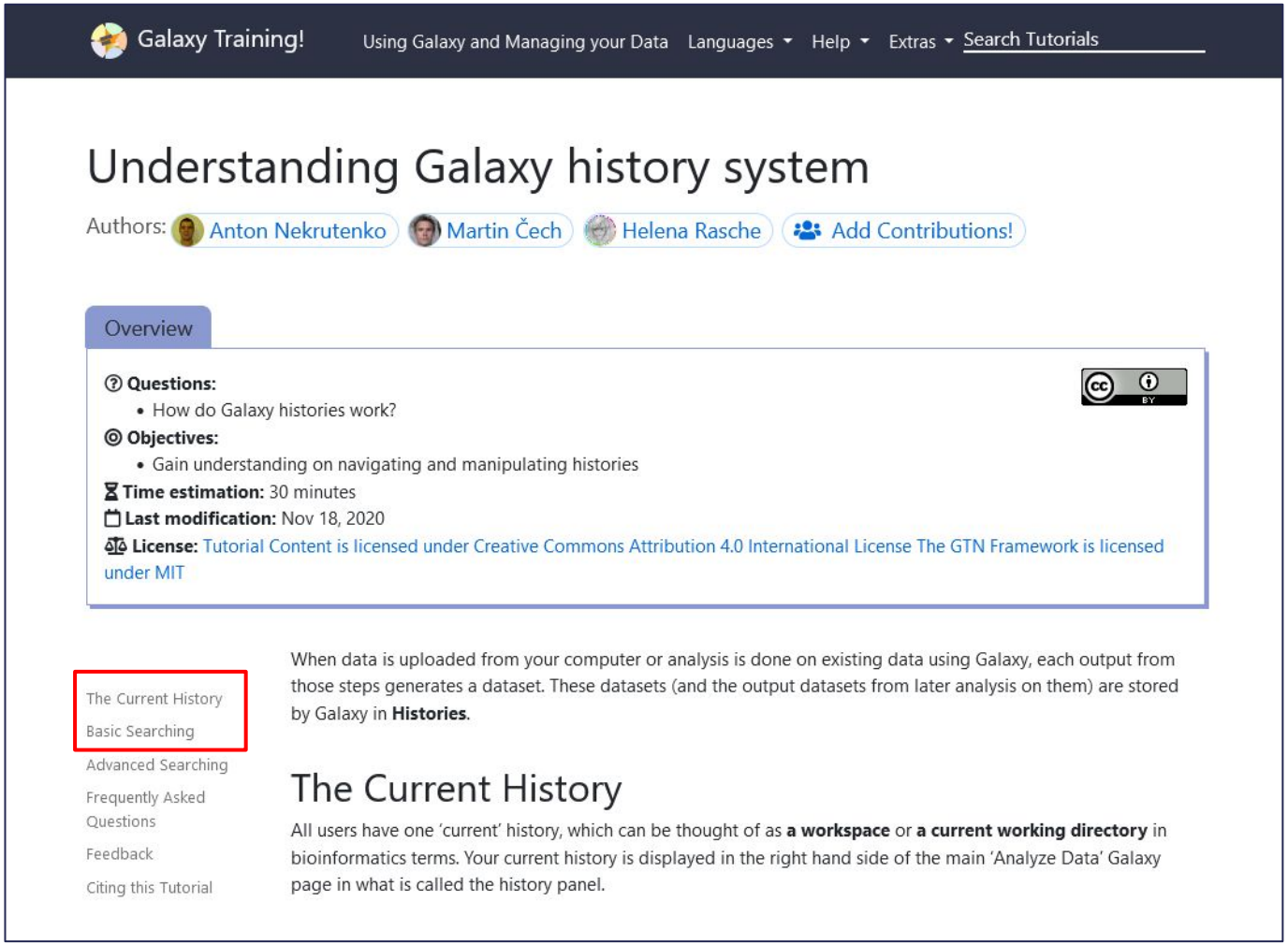

Galaxy France
Workflow
Visualize
Shared Data
Help
User
Using 3%
search histories
search all datasets
Create new
Current History
Switch to
Switch to
Switch to
Trinity example
4 shown
40.34 KB
search datasets
Drag datasets here to copy them to the current history
4: Trinity on data 2 and data 1: Gene to transcripts map
3: Trinity on data 2 and data 1: Assembled Transcripts
2: reads.right.fq
1: reads.left.fq
Error example
4 shown, 1 hidden
772 b
search datasets
4: NCBI Accession Download on : Log
3: NCBI Accession Download on : Downloaded Files
a list with 1 item
2: NCBI Accession Download on : Log
error
An error occurred with this dataset:
NCBI Entrez returned error code 400, are ID(
Failed to download file with id AAG19274.1 f
1: NCBI Accession Download on : Downloaded Files
a list
rnaseq hands-on 2022
14 shown, 3 hidden
9.88 MB
search datasets
17: RNA STAR on data 2 and data 1: mapped.bam
16: RNA STAR on data 2 and data 1: splice junctions.bed
15: RNA STAR on data 2 and data 1: log
14: RNA STAR on data 2 and data 1: mapped.bam
13: RNA STAR on data 2 and data 1: splice junctions.bed
12: RNA STAR on data 2 and data 1: log
8: MultiQC on data 6 and data 4: Webpage
7: MultiQC on data 6 and data 4: Stats
a list with 3 items
6: FastQC on data 2: RawData
blast hands-on 2022
8 shown, 10 hidden
66.36 MB
search datasets
13: NCBI BLAST+ blastn across collection 12
a list with 5 items
12: queries
a list with 5 items
11: blastx query5.fa vs 'protein BLAST database from data 2'
10: megablast query5.fa vs 'nucleotide BLAST database from data 1'
9: nucleotide BLAST database from data 1
8: protein BLAST database from data 2
2: Drosophila_melanogaster.BDGP6.22.pep.all.fa.gz
1: Drosophila_melanogaster.BDGP6.22.cds.all.fa.gz

History

Search histories and datasets



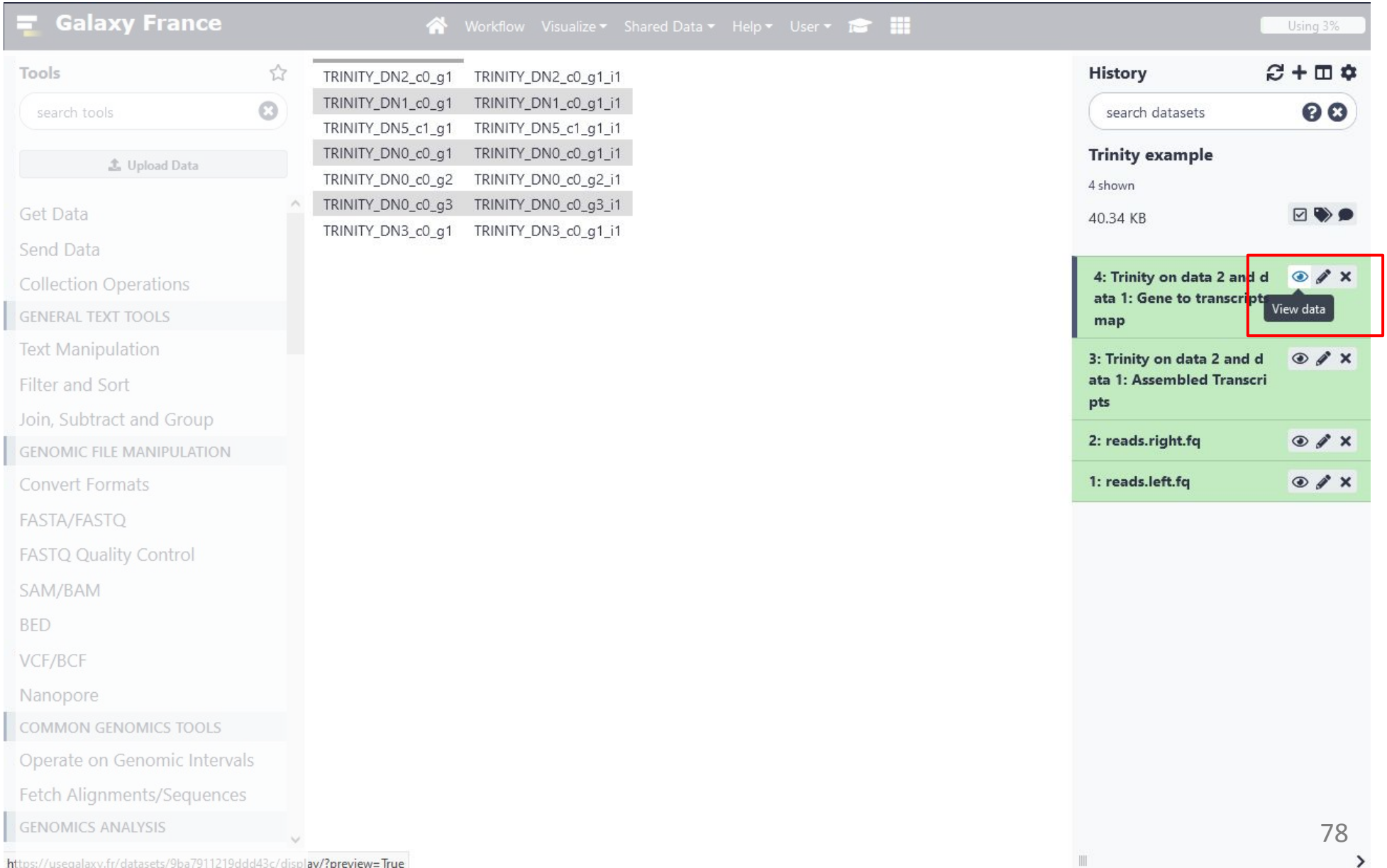
History



https://training.galaxyproject.org/training-material/topics/galaxy-interface/tutorials/history/tutorial.html

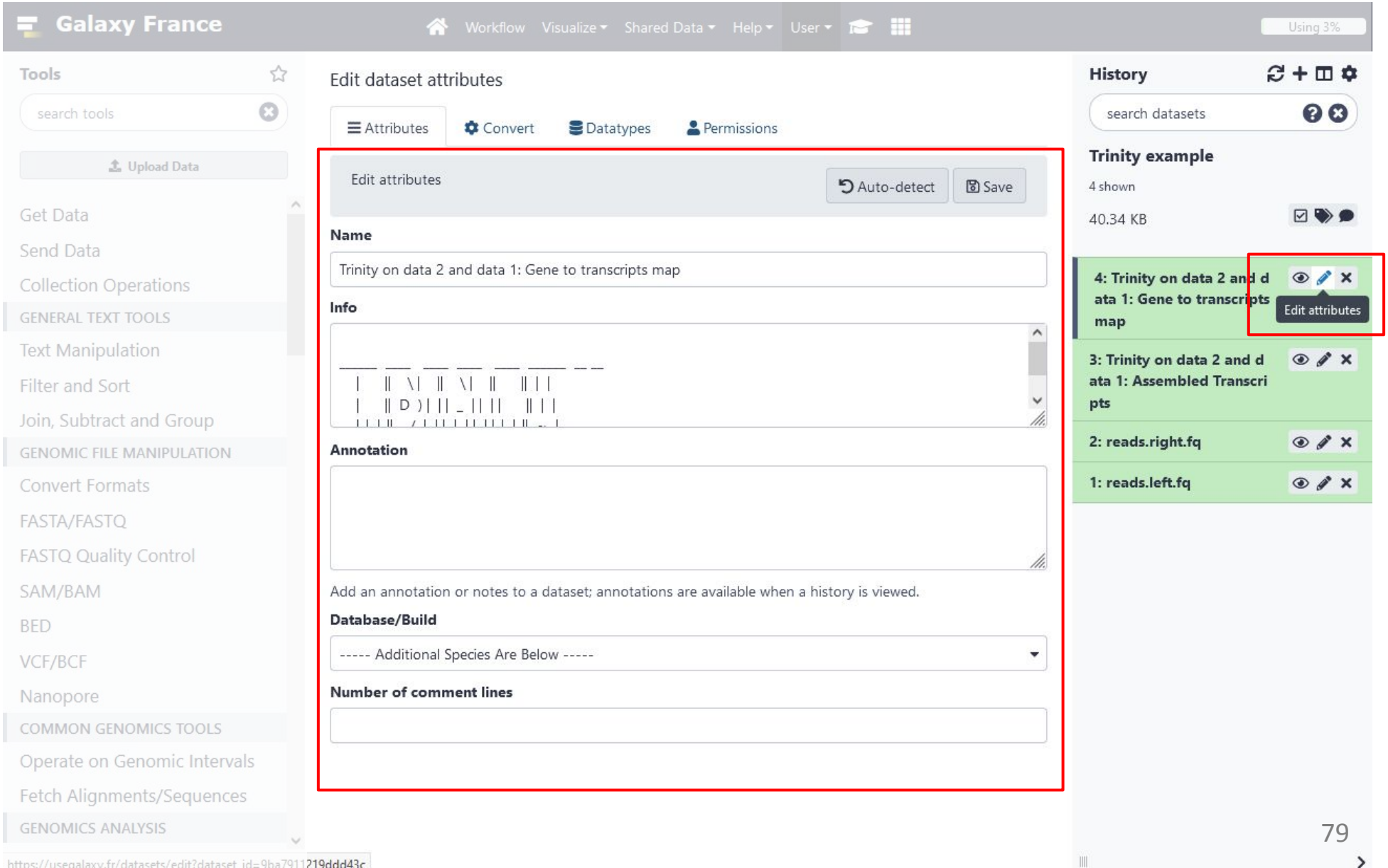
Dataset

Dataset display : text, tabular, pdf, picture, html …



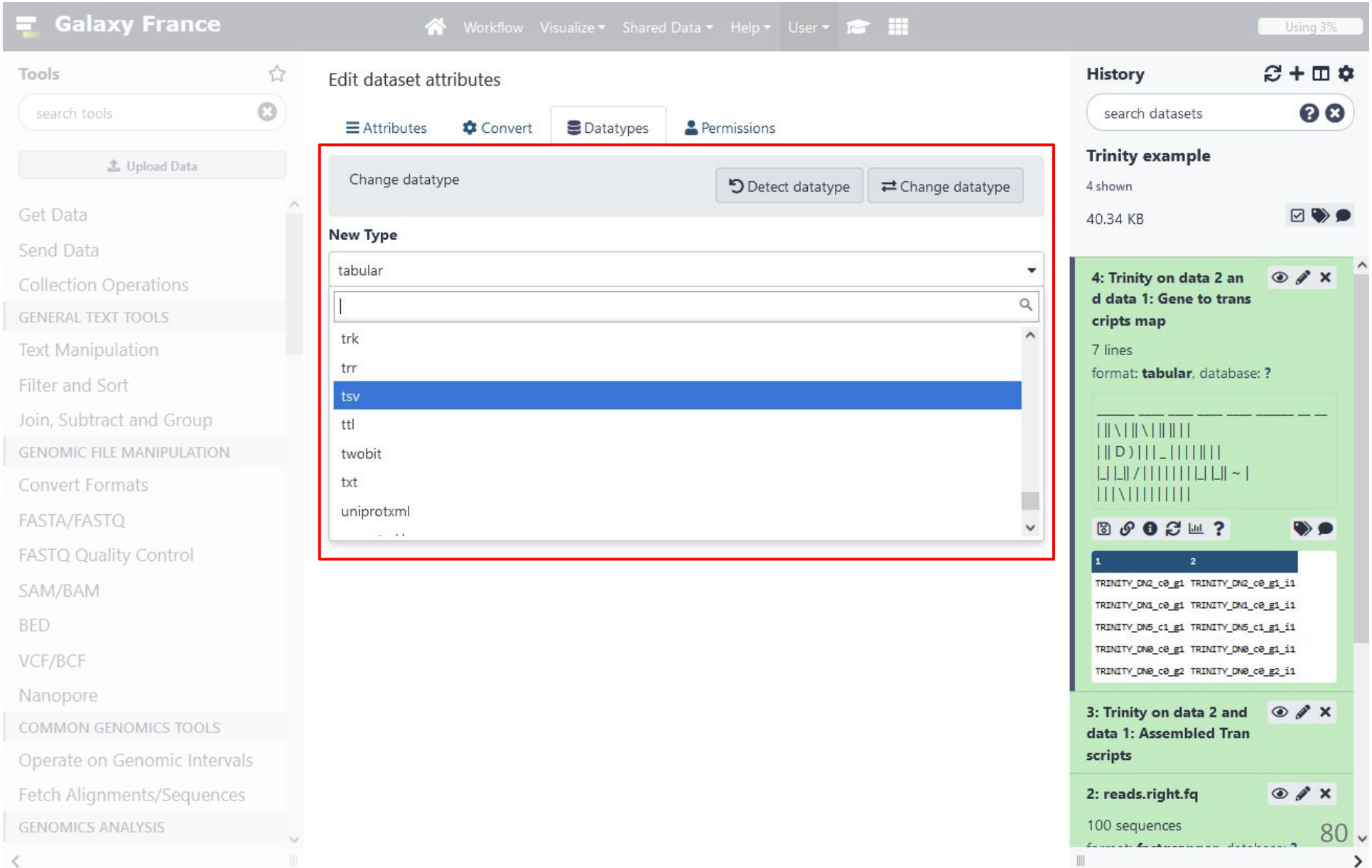
Dataset

Rename, annotate, change datatype…

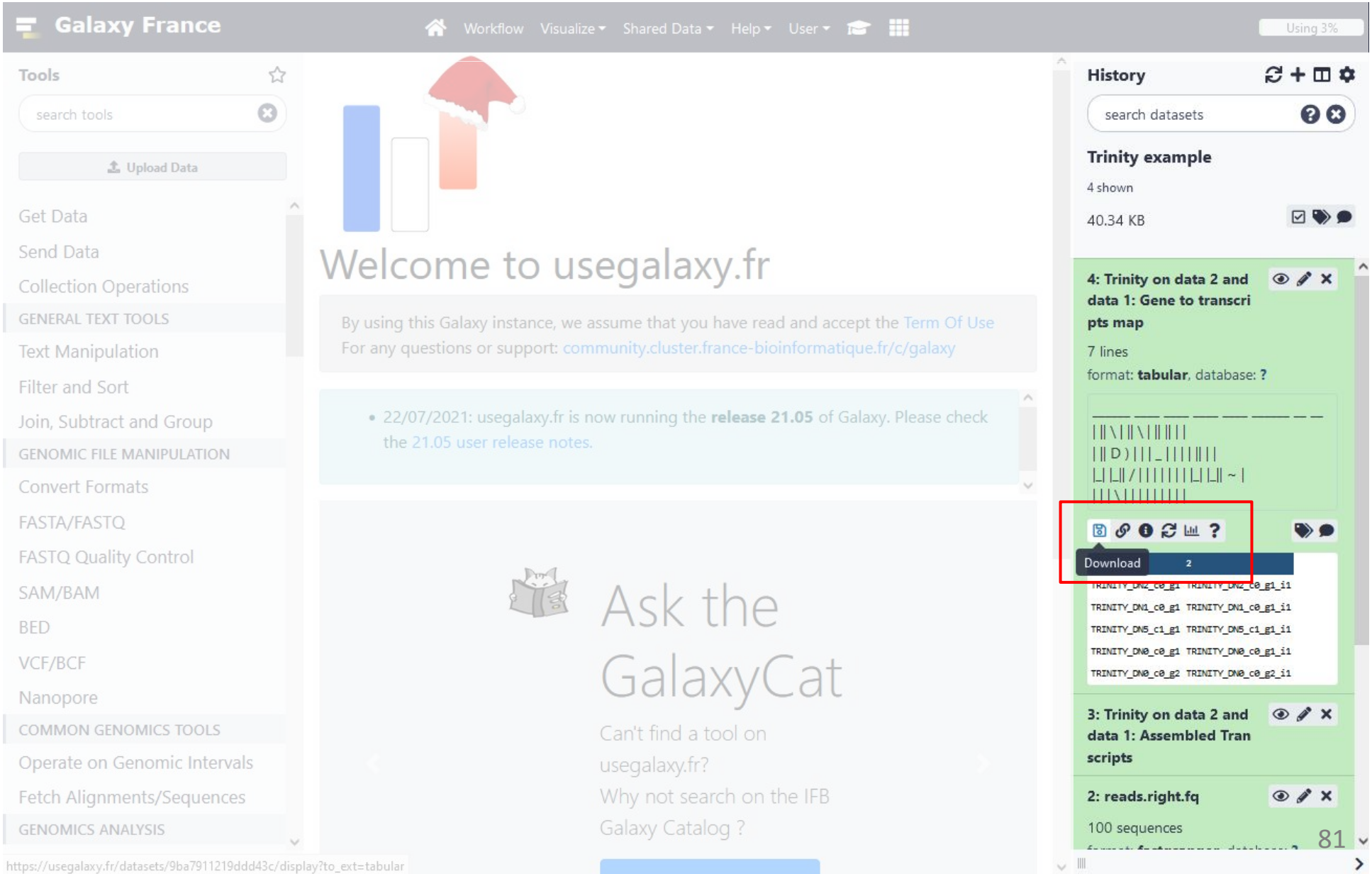


Dataset

Rename, annotate, change datatype…

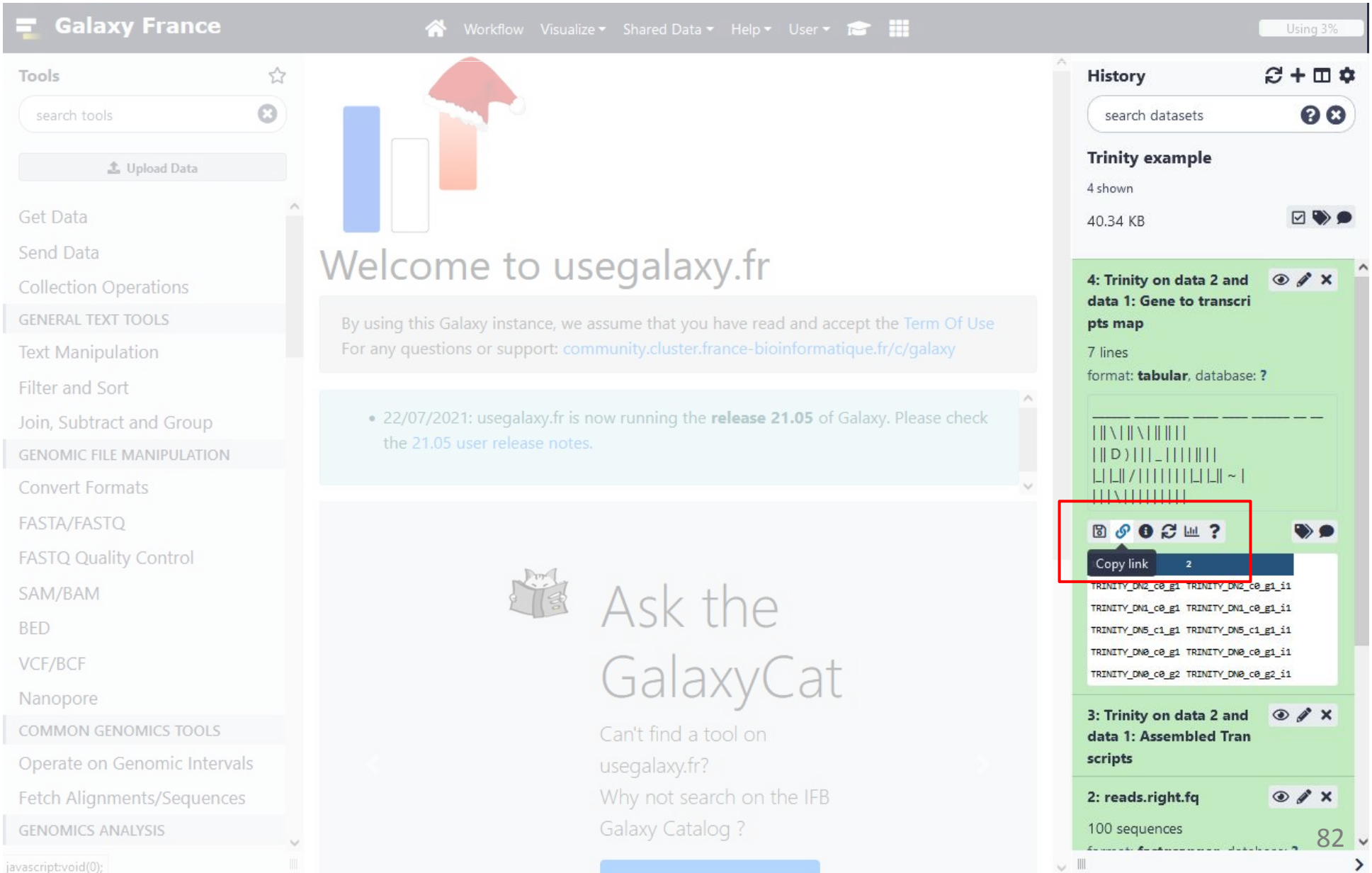


Download


Galaxy France
Workflow
Visualize
Shared Data
Help
User
Using 3%
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
Welcome to usegalaxy.fr
By using this Galaxy instance, we assume that you have read and accept the Term Of Use
For any questions or support: community.cluster.france-bioinformatique.fr/c/galaxy
22/07/2021: usegalaxy.fr is now running the release 21.05 of Galaxy. Please check the 21.05 user release notes.
Ask the GalaxyCat
Can't find a tool on usegalaxy.fr?
Why not search on the IFB Galaxy Catalog ?
History
search datasets
Trinity example
4 shown
40.34 KB
4: Trinity on data 2 and data 1: Gene to transcripts map
7 lines
format: tabular, database: ?
Download
TRINITY_DN1_c0_g1 TRINITY_DN1_c0_g1_i1
TRINITY_DN5_c1_g1 TRINITY_DN5_c1_g1_i1
TRINITY_DN0_c0_g1 TRINITY_DN0_c0_g1_i1
TRINITY_DN0_c0_g2 TRINITY_DN0_c0_g2_i1
3: Trinity on data 2 and data 1: Assembled Transcripts
2: reads.right.fq
100 sequences
https://usegalaxy.fr/datasets/9ba7911219ddd43c/display?to_ext=tabular

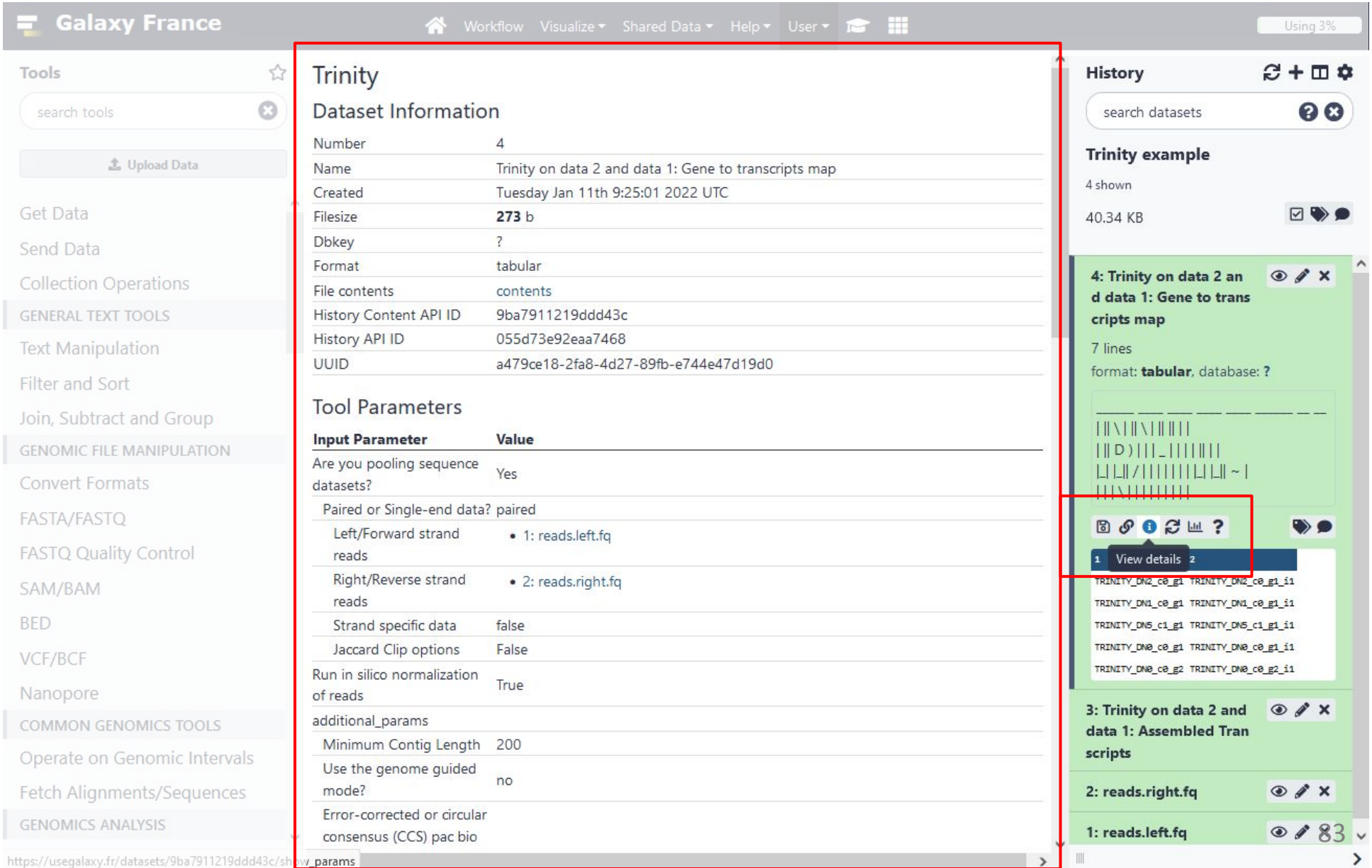
Dataset

Link to data

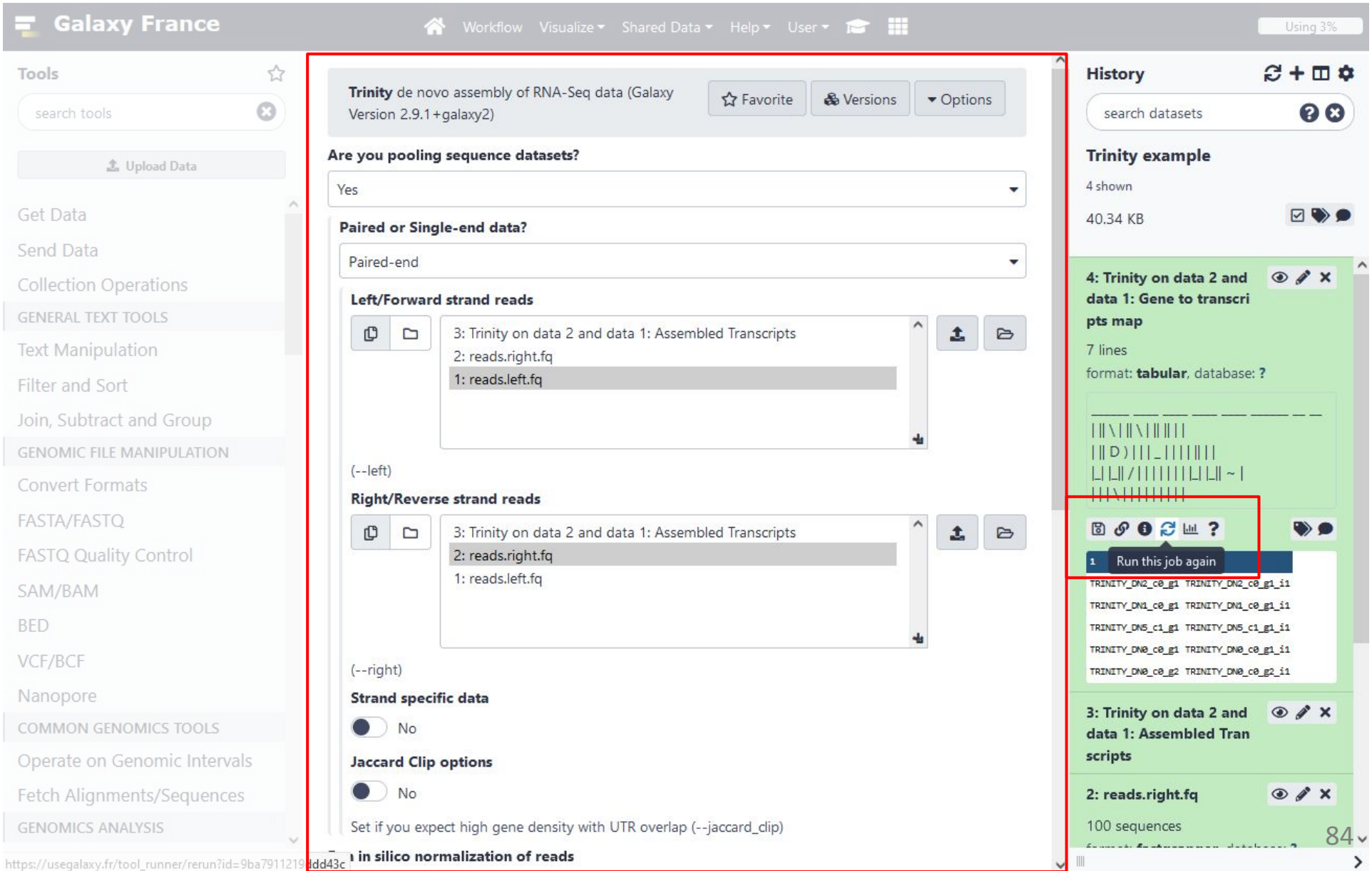


Dataset

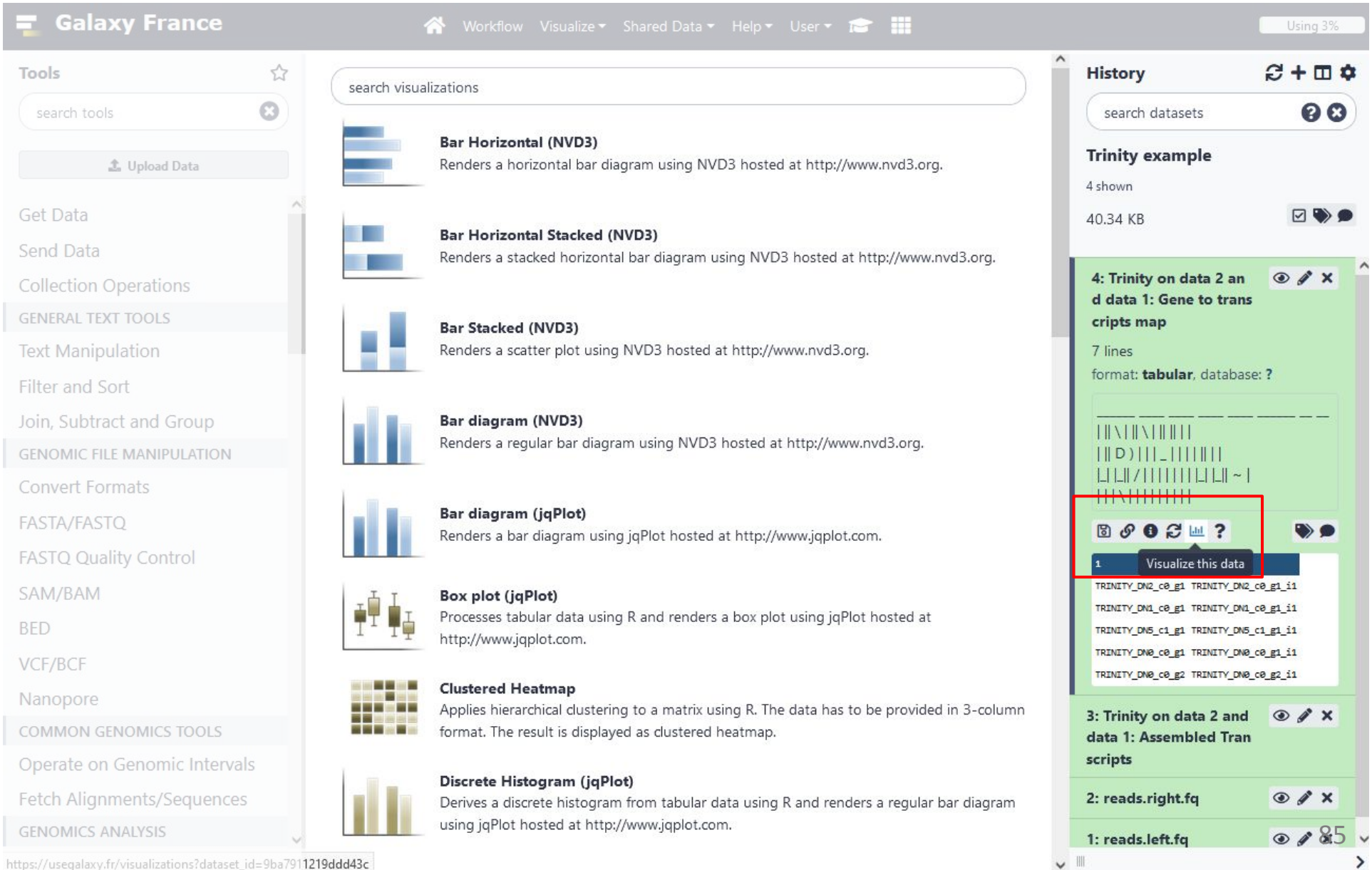
Informations



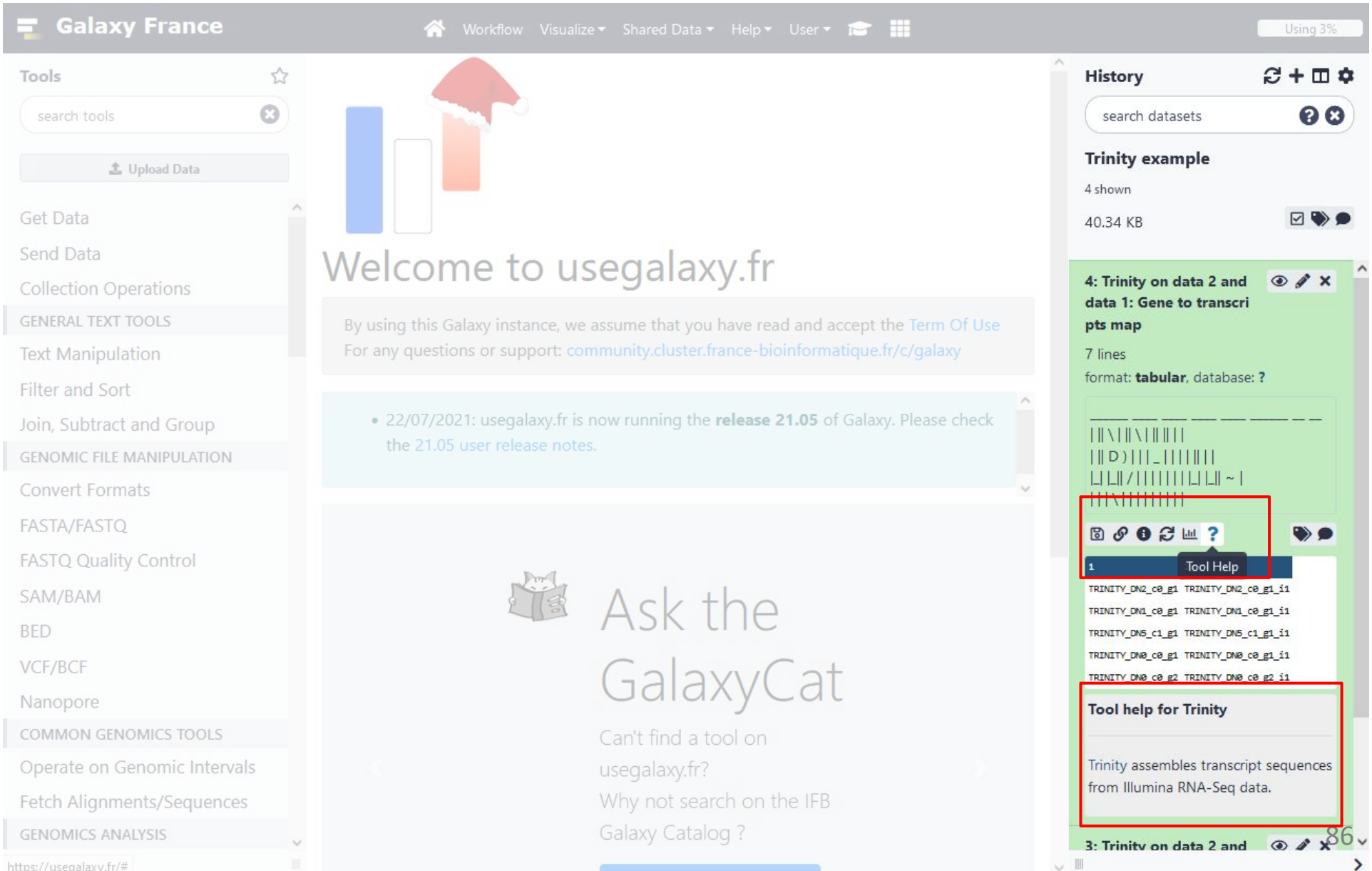
Re-run a job


Galaxy France
Workflow
Visualize
Shared Data
Help
User
Using 3%
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
Trinity de novo assembly of RNA-Seq data (Galaxy Version 2.9.1+galaxy2)
Favorite
Versions
Options
Are you pooling sequence datasets?
Yes
Paired or Single-end data?
Paired-end
Left/Forward strand reads
3: Trinity on data 2 and data 1: Assembled Transcripts
2: reads.right.fq
1: reads.left.fq
(--left)
Right/Reverse strand reads
3: Trinity on data 2 and data 1: Assembled Transcripts
2: reads.right.fq
1: reads.left.fq
(--right)
Strand specific data
No
Jaccard Clip options
No
Set if you expect high gene density with UTR overlap (--jaccard_clip)
History
search datasets
Trinity example
4 shown
40.34 KB
4: Trinity on data 2 and data 1: Gene to transcripts map
7 lines
format: tabular, database: ?
Run this job again
3: Trinity on data 2 and data 1: Assembled Transcripts
2: reads.right.fq
100 sequences

Visualisation tools



Help


Galaxy France
Workflow
Visualize
Shared Data
Help
User
Using 3%
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
Welcome to usegalaxy.fr
By using this Galaxy instance, we assume that you have read and accept the Term Of Use
For any questions or support: community.cluster.france-bioinformatique.fr/c/galaxy
22/07/2021: usegalaxy.fr is now running the release 21.05 of Galaxy. Please check the 21.05 user release notes.
Ask the GalaxyCat
Can't find a tool on usegalaxy.fr?
Why not search on the IFB Galaxy Catalog ?
History
search datasets
Trinity example
4 shown
40.34 KB
4: Trinity on data 2 and data 1: Gene to transcripts map
7 lines
format: tabular, database: ?
Tool Help
TRINITY_DN2_c0_g1 TRINITY_DN2_c0_g1_i1
TRINITY_DN1_c0_g1 TRINITY_DN1_c0_g1_i1
TRINITY_DN5_c1_g1 TRINITY_DN5_c1_g1_i1
TRINITY_DN0_c0_g1 TRINITY_DN0_c0_g1_i1
TRINITY_DN0_c0_g2 TRINITY_DN0_c0_g2_i1
Tool help for Trinity
Trinity assembles transcript sequences from Illumina RNA-Seq data.
3: Trinity on data 2 and

Dataset - Visualisations

Visualisations in Galaxy

Helena Rasche Saskia Hiltemann Add Contributions!

Updated: Jul 9, 2021
Or view the plain-text slides without JS
Tip: press P to view the presenter notes

https://training.galaxyproject.org/training-material/topics/visualisation/slides/introduction.html

Hands-on

TOOLS

Galaxy 101

Authors: Saskia Hiltemann Nicola Soranzo Clemens Blank Anton Nekrutenko Björn Grüning Anne Pajon Helena Rasche Add Contributions!

Overview

Questions:
- Which coding exon has the highest number of single nucleotide polymorphisms (SNPs) on human chromosome 22?

Objectives:
- Familiarize yourself with the basics of Galaxy
- Learn how to obtain data from external sources
- Learn how to run tools
- Learn how histories work
- Learn how to create a workflow
- Learn how to share your work

Time estimation: 1 hour

Level: Introductory

Supporting Materials:

Topic Overview slides Datasets Workflows GTN Video Library Available on these Galaxies

Last modification: Nov 17, 2021

Introduction

This tutorial aims to familiarize you with the Galaxy user interface. It will teach you how to perform basic tasks such as importing data, running tools, working with histories, creating workflows, and sharing your work.

Results may vary

Your results may be slightly different from the ones presented in this tutorial due to differing versions of tools, reference data, external databases, or because of stochastic processes in the algorithms.

Agenda

https://training.galaxyproject.org/training-material/topics/introduction/tutorials/galaxy-intro-101/tutorial.html

Deleting

DATASET

https://training.galaxyproject.org/training-material/topics/galaxy-interface/tutorials/download-delete-data/tutorial.html#deleting-data-from-galaxy

Downloading and Deleting Data in Galaxy

Authors: Simon Gladman Helena Rasche Add Contributions!

Overview

Questions:
- How can I efficiently download my data from Galaxy once I've completed my analyses?
- How do I delete old data to make more room in my quota?

Objectives:
- Have a greater knowledge of how Galaxy handles data downloads and deletions
- Be able to successfully download and delete their own data

Time estimation: 20 minutes
Level: Introductory
Last modification: Oct 5, 2020

Introduction

This tutorial outlines the process to get your data out of Galaxy and to delete it from Galaxy afterwards.

After you have completed your analysis on a Galaxy server, you may need to download the results for use in reports, papers, or other requirements. Galaxy has a myriad of methods for downloading data from individual datasets to a collection of data to an entire history along with all of its metadata. The first section of this tutorial will walk you through all of these methods.

On some Galaxy servers, user disk quotas have been enabled meaning that users are only able to store a certain amount of data on that particular Galaxy server. Quotas can differ on different servers. User disk quotas for the `usegalaxy.*` servers are shown in the table below:

Galaxy Server	Unregistered User	Registered User
usegalaxy.org	5 GB	250 GB
usegalaxy.eu	5 GB	250 GB or 500 GB for Elixir Members
usegalaxy.org.au	5 GB	100 GB or 600GB for Australian Users

Other Galaxy servers may have different quota systems.

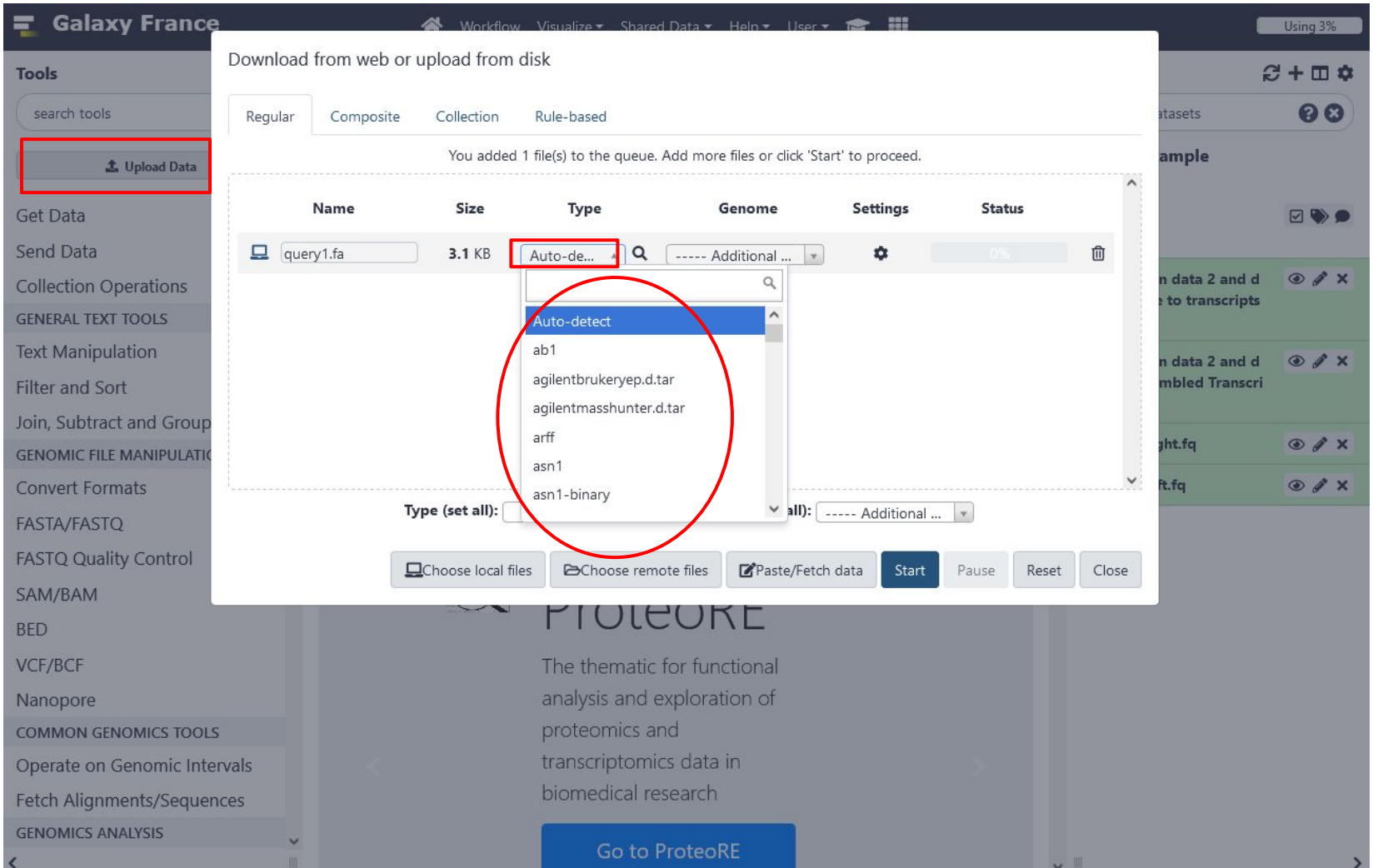
As you use a Galaxy server and upload data, perform analyses etc. you will use up your quota. If you have filled your quota and you want to do some new work, you can download your old work and delete it off the Galaxy server to free up quota space. Galaxy uses a two step process to delete user data from disk. The first step is to delete the data - this marks the data as deleted but it can be undeleted for a certain time[1]. The second step is to

Datatypes

DATASET

Dataset - Datatypes

- Every Galaxy dataset is associated with a datatype.
- Datasets produced by a tool have their datatype assigned by the tool



Dataset - Datatypes

Common text formats:

- ***txt***: plain text ('.txt')
- ***tabular***: tab delimited ('.tab', '.txt', etc.)

```
          wt_37_2  wt_37_3  wt_37_1
TR24|c0_g1_i1     90.00    67.00    85.00
TR2779|c0_g1_i1   186.00   137.00   217.00
TR127|c1_g1_i1    9.00 23.00    16.00
```

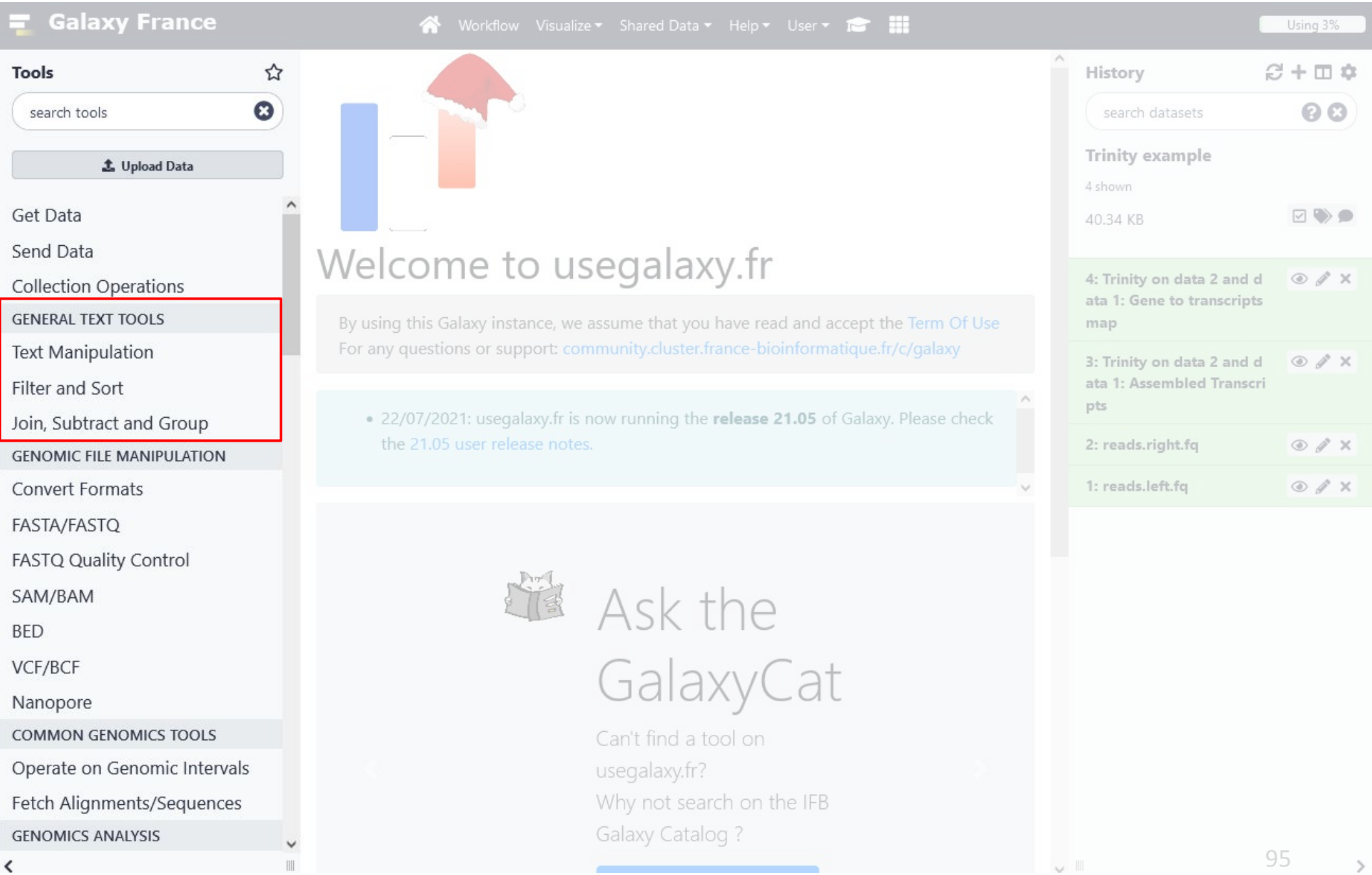
- ***csv***: comma-separated values ('.csv')

```
Year,Make,Model
1997,Ford,E350
2000,Mercury,Couga
r
```

- ***html***: standard language for web pages

```
<!DOCTYPE html>
<html>
  <head>
    <title>This is a
title</title>
  </head>
  <body>
    <p>Hello world!</p>
  </body>
</html>
```

Dataset - Datatypes



Dataset - Datatypes

Common binary formats:

- ***data***: generic binary format
- ***zip, tar***: archives
- ***pdf, png, jpg, bmp, tiff, gif***: images
- ***rdata***: statistical computing program R
- ***bam*, wig, bigwig**: sequence alignment

Dataset - Datatypes

Sequence file formats:

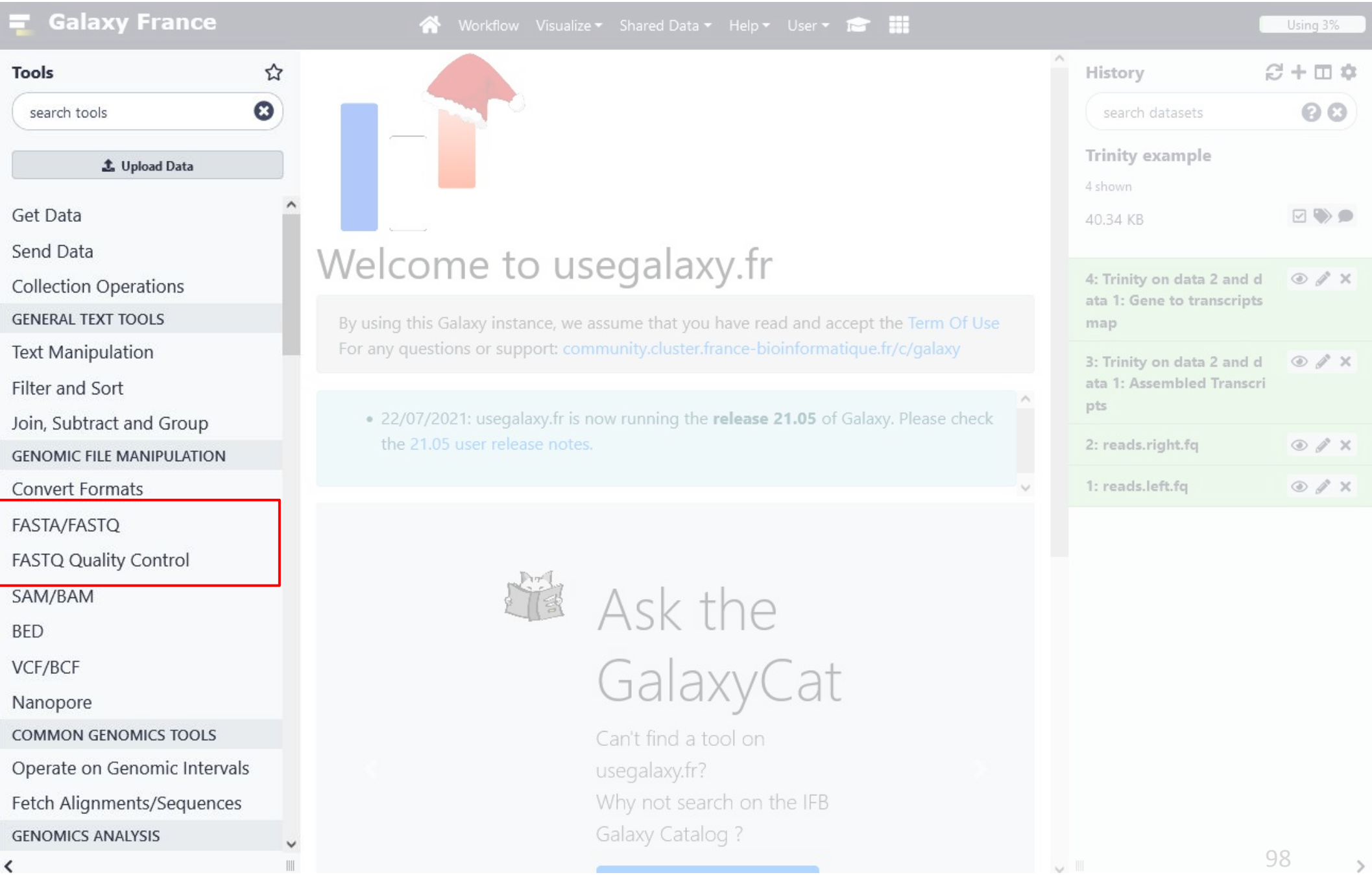
- ***fasta***: a single-line description with '>', followed by lines of sequence data ('.fasta', '.fas')

```
>sequence1
atgcgtttgcgtgcatgcgtttgcgtgcatgcgtttgcgtgcatgcgtttgcgtg
c
atgcgtttgcgtgc
>sequence2
tttcgtgcgtatagtttcgtgcgtatagtttcgtgcgtatagtttcgtgcgtata
g
tggcgcggt
```

- ***fastq***: sequence + quality score ('.fastq', '.fq')

```
@SEQ_ID
GATTTGGGGTTCAAAGCAGTATCGATCAAATAGTAAATCCATTTGTTCAACTCACAGTT
T
+
!''*((((***+))%%%++)(%%%%).1***-+*''))**55CCF>>>>>>CCCCCCC6
5
@SEQ_ID2
GATTTGGGGTTCAAAGCAGTATCGATCAAATAGTAAATCCATTTGTTCAACTCACAGTT
T
+
!''*((((***+))%%%++)(%%%%).1***-+*''))**55CCF>>>>>>CCCCCCC6
```

Dataset - Datatypes



Dataset - Datatypes

Sequence file formats:

- ***gff3, bed, genbank***: sequence + annotations

bed

```
track name=pairedReads description="Clone Paired Reads"
useScore=1
chr22 1000 5000 cloneA 960 + 1000 5000 0 2 567,488, 0,3512
chr22 2000 6000 cloneB 900 - 2000 6000 0 2 433,399, 0,3601
```

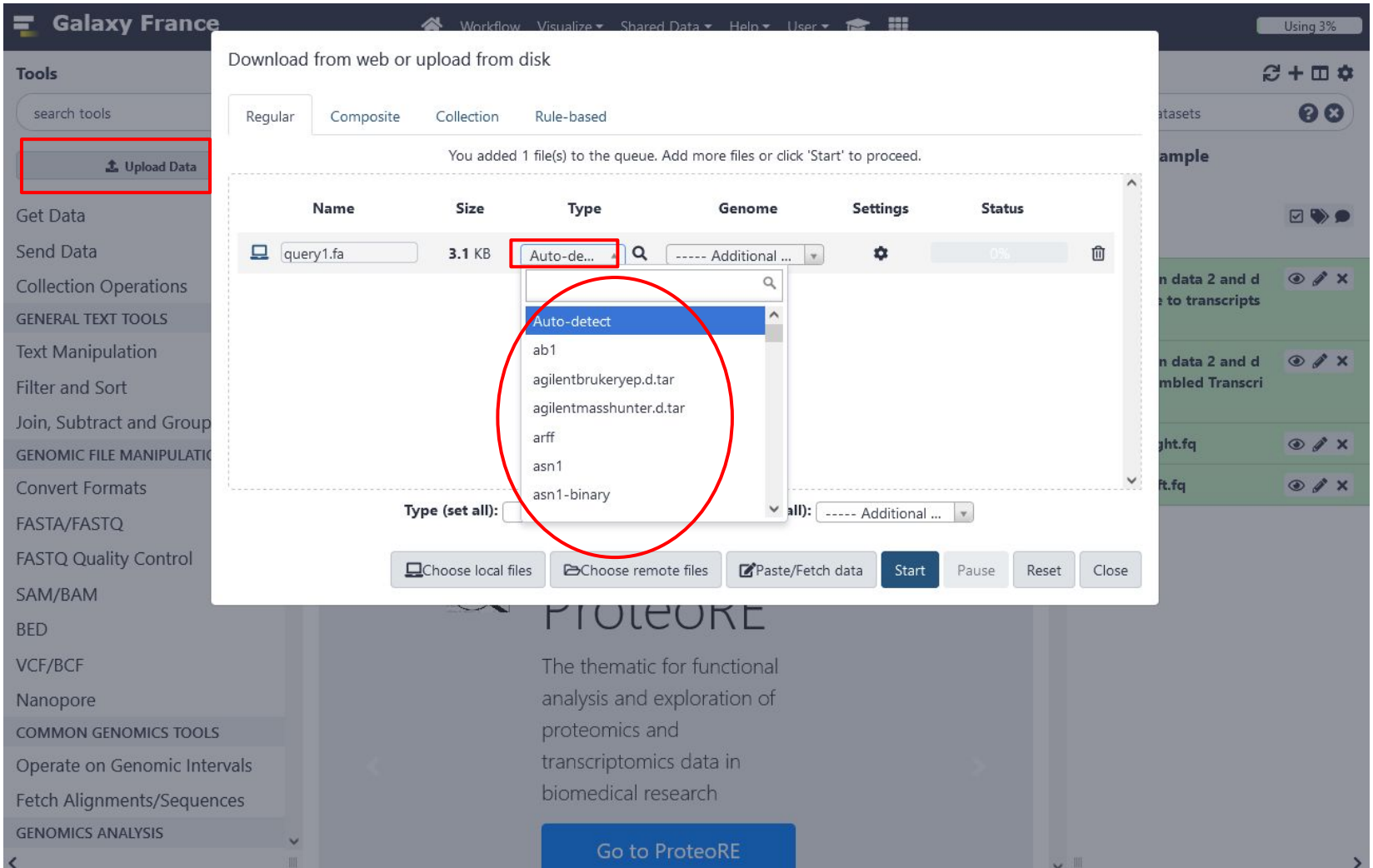
gff3

```
##gff-version 3
ctg123  .  exon  1300  1500  .  +  .
ID=exon00001
ctg123  .  exon  1050  1500  .  +  .
ID=exon00002
ctg123  .  exon  3000  3902  .  +  .
ID=exon00003
##FASTA
>ctg123
cttctgggcgtacccgattctcggagaacttgccgcaccattccgcctt
g
tgttcattgctgcctgcatgttcattgtctacctcggctacgtgtggct
g
…
```

https://genome.ucsc.edu/FAQ/FAQformat.html#format3

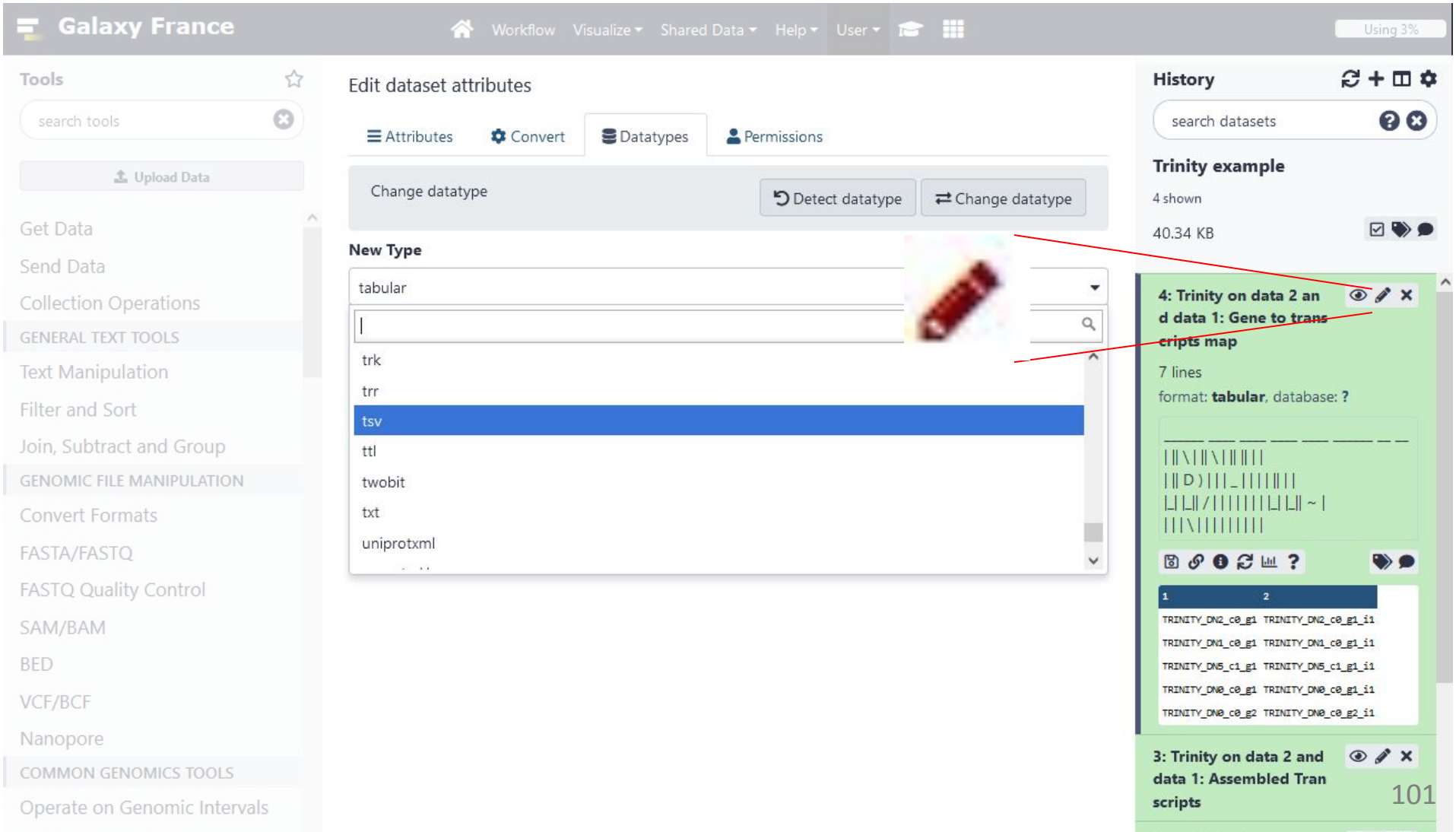
Dataset - Datatypes

- On upload, datatype can be detected or user specified.



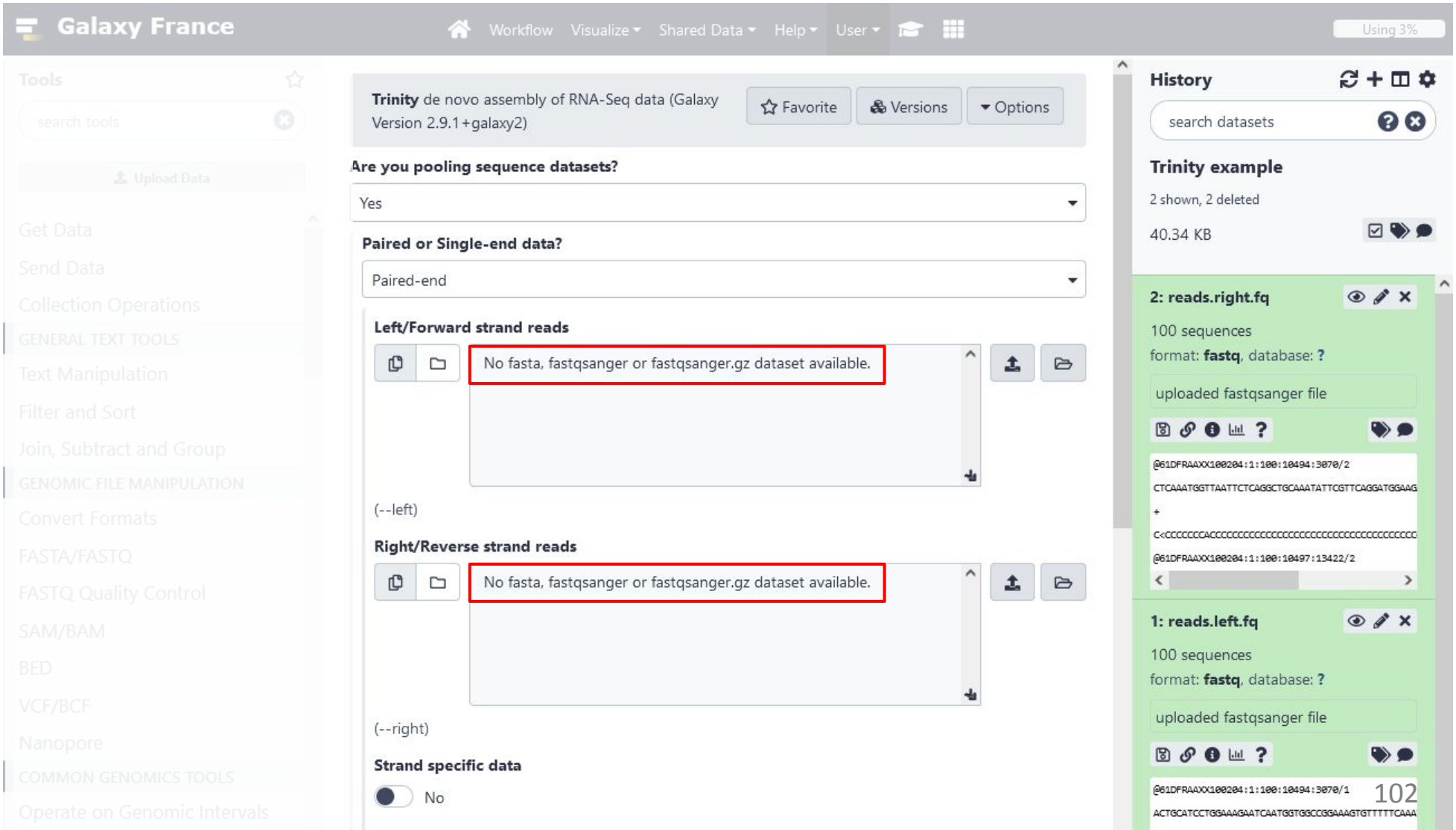
Dataset - Datatypes

- Change the datatype of a dataset



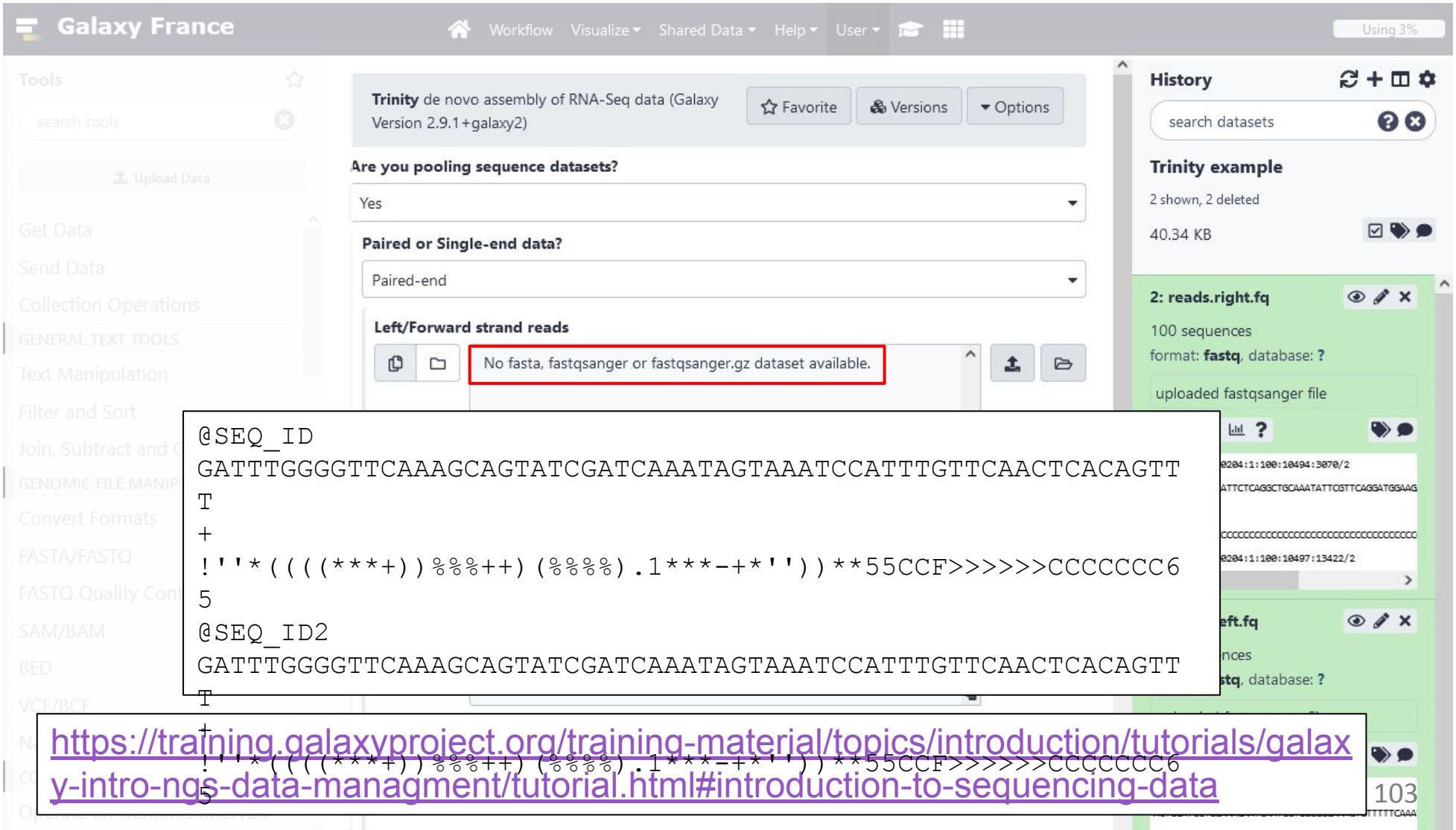
Dataset - Datatypes

- Many tools will only accept input datasets with the appropriate datatype assigned.



Dataset - Datatypes

- Many tools will only accept input datasets with the appropriate datatype assigned.



https://training.galaxyproject.org/training-material/topics/introduction/tutorials/galaxy-intro-ngs-data-managment/tutorial.html#introduction-to-sequencing-data

Hands-on

TOOLS (sequence files manipulation)

Tools (sequence files manipulation) – Hands-on

Part 1:

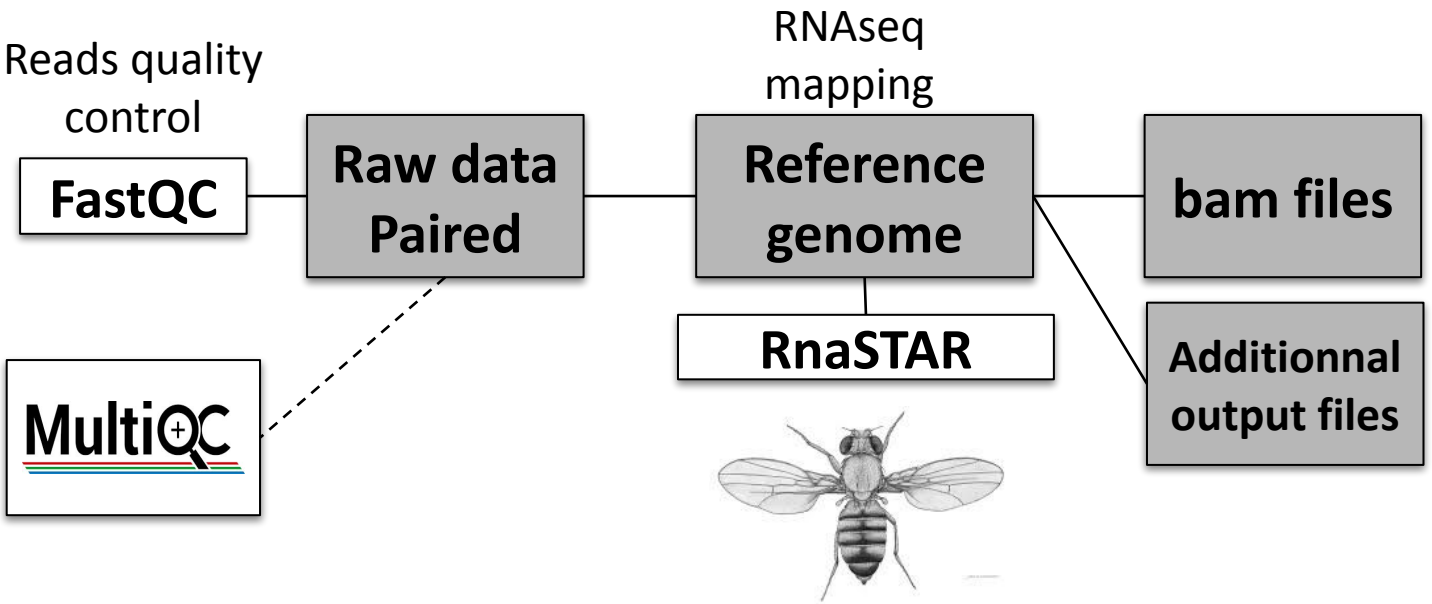
You have sequencing data from your favorite species. You want to check the quality of your sequences and to map on the reference genome.

Case of Drosophila melanogaster (dm)

Part 1:

You have sequencing data from your favorite species. You want to check the quality of your sequences and to map on the reference genome.

Case of Drosophila melanogaster (dm)



Tools (sequence files manipulation) – Hands-on

Part 1:

You have sequencing data from your favorite species. You want to check the quality of your sequences and to map on the reference genome.

Case of Drosophila melanogaster (dm)

1. Look into the description of the tool FastQC, what is its purpose ?
2. Does it takes compressed files ? ".gz"
3. Which encoding of the file (in FastQC results, basic statistics) ? => edit if needed the datatype of your reads files
4. Look at mapping tools, map your reads on genome with RNA STAR.
5. What is the default parameter of the option "Maximum ratio of mismatches to mapped length" in RNA STAR ?
6. How to map on another reference genome or assembly release ?

Tools (sequence files manipulation) – Hands-on

Part 2:

You have new sequences that you want to compare with the gene and protein databases from your favorite species (BLAST).

Case of Drosophila melanogaster (dm)

1. Create new history
2. Import CDS and peptide sequences databases
3. Import query sequences
4. Make BLAST databases
5. Run BLAST against the CDS database
6. Run BLAST against the protein database

Tools (sequence files manipulation) – Hands-on

Part 2:

You have new sequences that you want to compare with the gene and protein databases from your favorite species (BLAST).

Case of Drosophila melanogaster (dm)

1. Create new history
2. Import CDS and peptide sequences databases
3. Import query sequences
4. Make BLAST databases *Blast --> NCBI BLAST+ makeblastdb*
5. Run BLAST against the CDS database
6. Run BLAST against the protein database

Part 2:

You have new sequences that you want to compare with the gene and protein databases from your favorite species (BLAST).

Case of Drosophila melanogaster (dm)

1. Create new history
2. Import CDS and peptide sequences databases
3. Import query sequences
4. Make BLAST databases *Blast --> NCBI BLAST+ makeblastdb*
5. Run BLAST against the CDS database *Blast --> NCBI BLAST+ blastn*
6. Run BLAST against the protein database

Tools (sequence files manipulation) – Hands-on

Part 2:

You have new sequences that you want to compare with the gene and protein databases from your favorite species (BLAST).

Case of Drosophila melanogaster (dm)

1. Create new history
2. Import CDS and peptide sequences databases
3. Import query sequences
4. Make BLAST databases *Blast --> NCBI BLAST+ makeblastdb*
5. Run BLAST against the CDS database *Blast --> NCBI BLAST+ blastn*
6. Run BLAST against the protein database *Blast --> NCBI BLAST+ blastx*

DATASET COLLECTION

Dataset collection

- Problematic: you have a large numbers of datasets to send through the same analysis

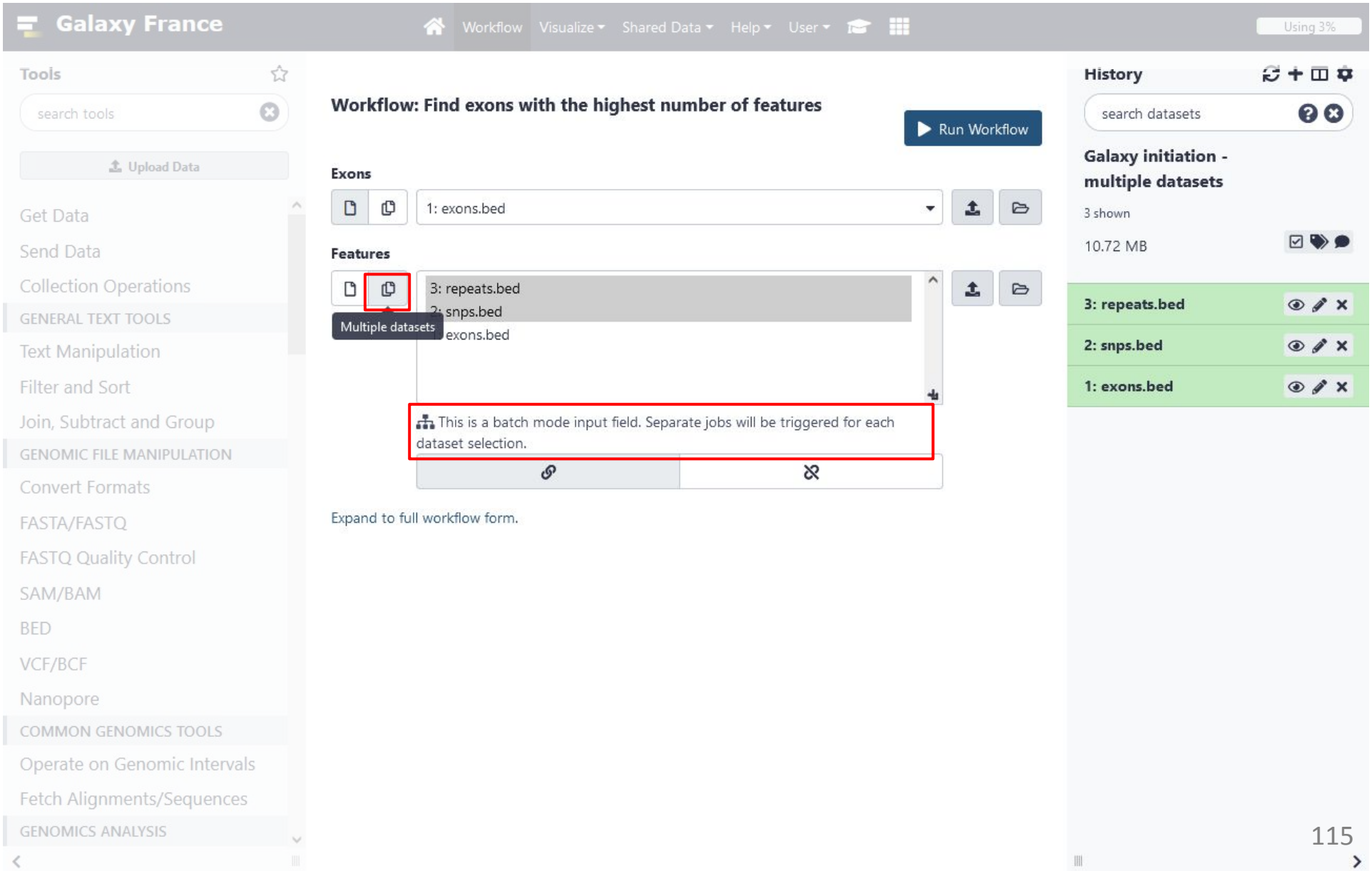
Dataset collection

- Problematic: you have a large numbers of datasets to send through the same analysis

- Solution 1: select multiple datasets as input

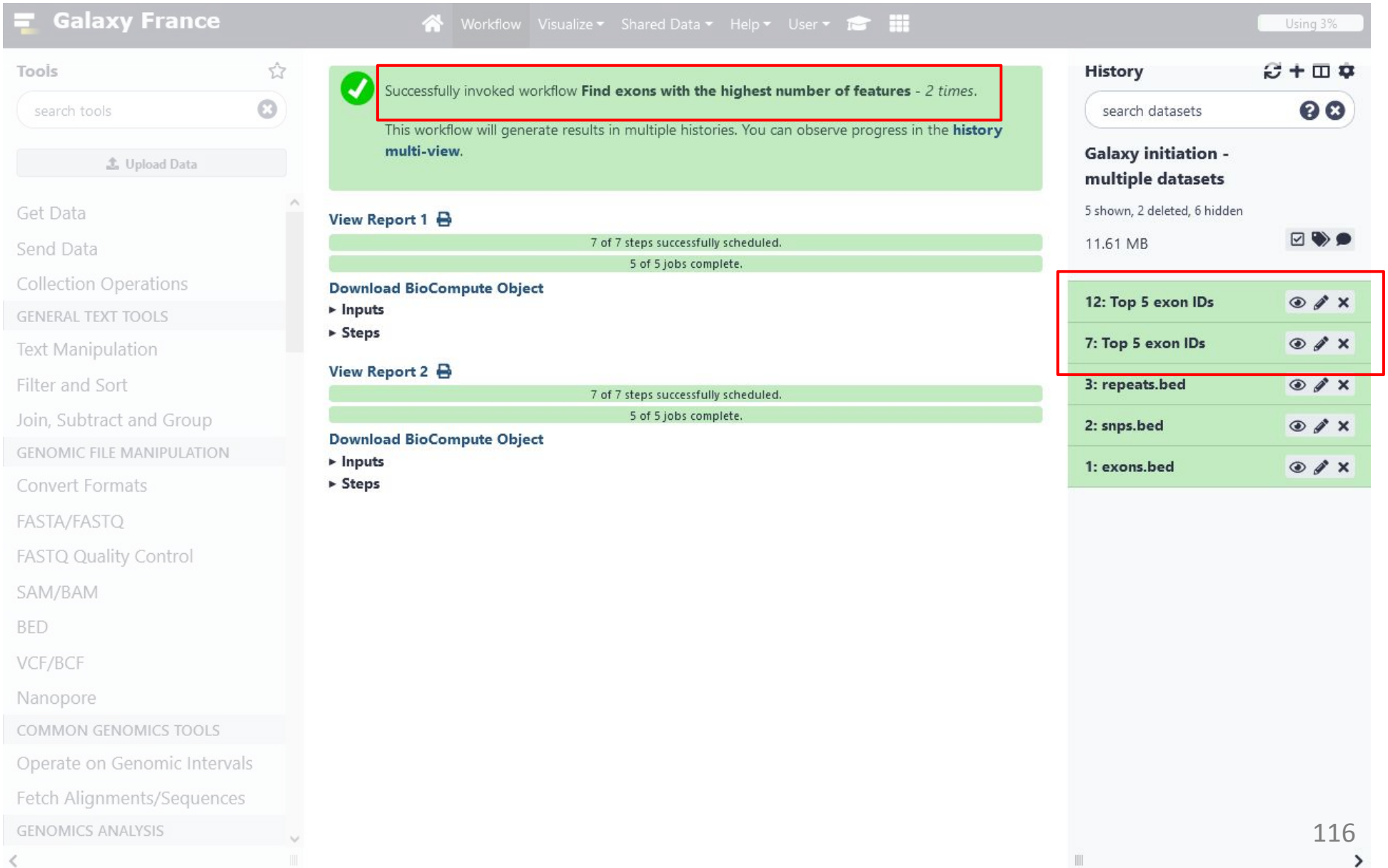
Dataset collection

Select multiple datasets as input



Dataset collection

Select multiple datasets as input



Dataset collection

- Problematic: you have a large numbers of datasets to send through the same analysis

- Solution 1: select multiple datasets as input
- Solution 2: create a dataset collection (any number of datasets bundled as a **single entity**, i.e. minimize clutter)

Dataset collection

- Problematic: you have a large numbers of datasets to send through the same analysis

- Solution 1: select multiple datasets as input
- Solution 2: create a dataset collection (any number of datasets bundled as a **single entity**, i.e. minimize clutter)
  - Dataset list: set of files of the same type
  - Dataset pairs: pairs of read files (forward, reverse)
  - List of dataset pairs

Dataset collection

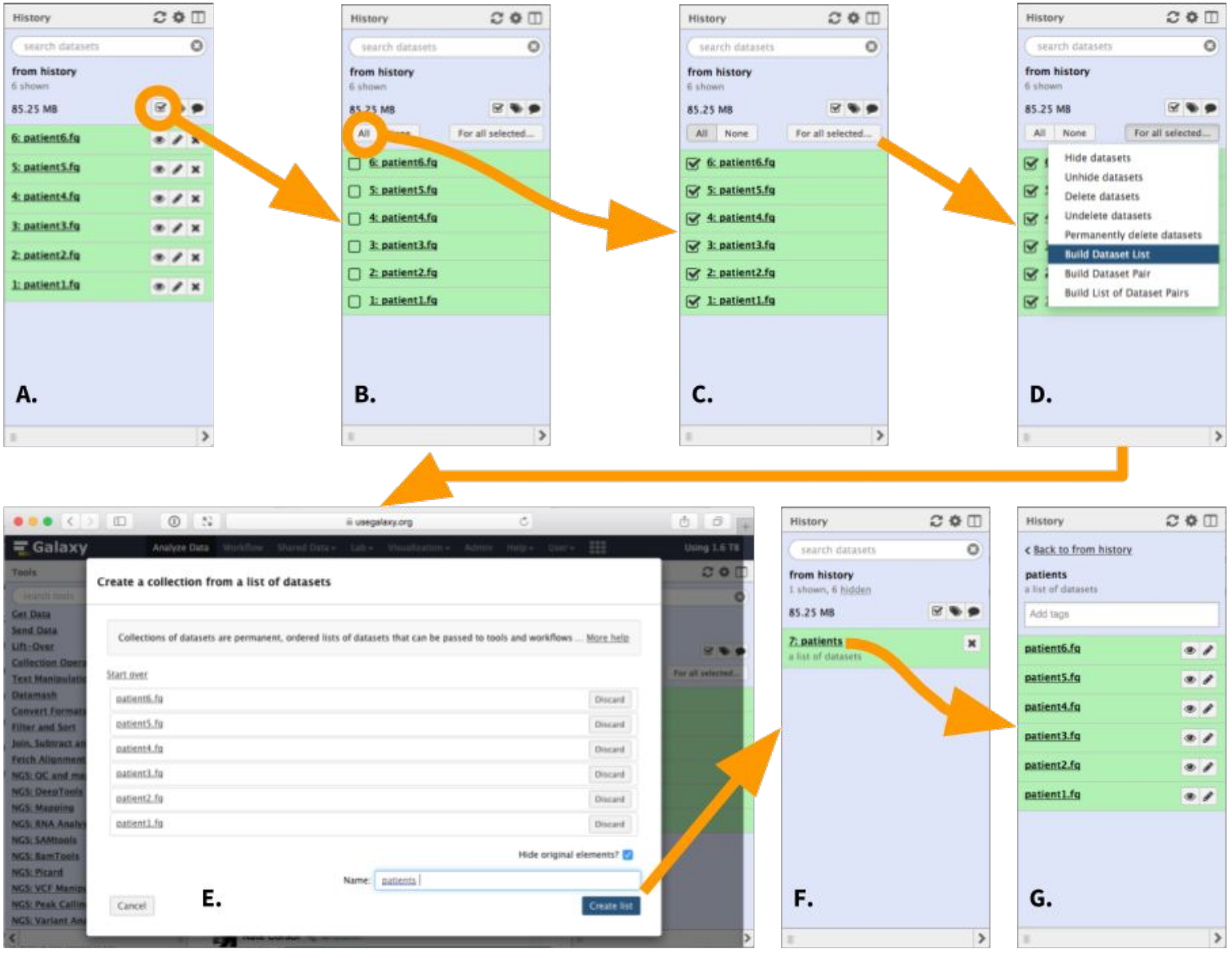
- Problematic: you have a large numbers of datasets to send through the same analysis

- Solution 1: select multiple datasets as input
- Solution 2: create a dataset collection (any number of datasets bundled as a **single entity**, i.e. minimize clutter)
  - Dataset list: set of files of the same type
  - Dataset pairs: pairs of read files (forward, reverse)
  - List of dataset pairs

- Galaxy runs the tool automatically on each dataset using the same settings

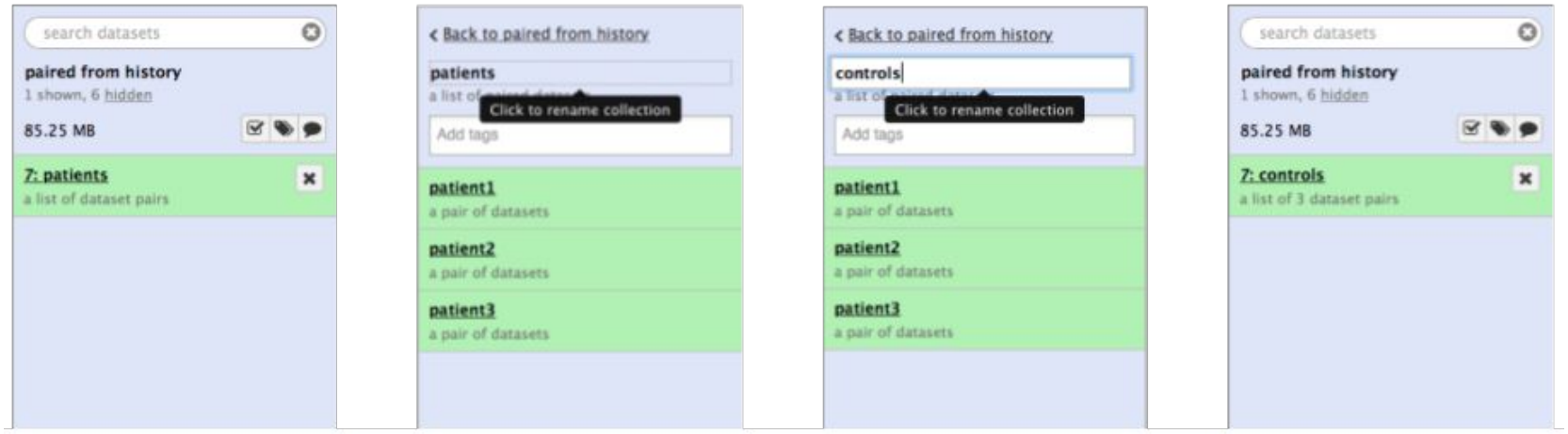
Dataset collection

Create a collection from datasets in your history



Dataset collection

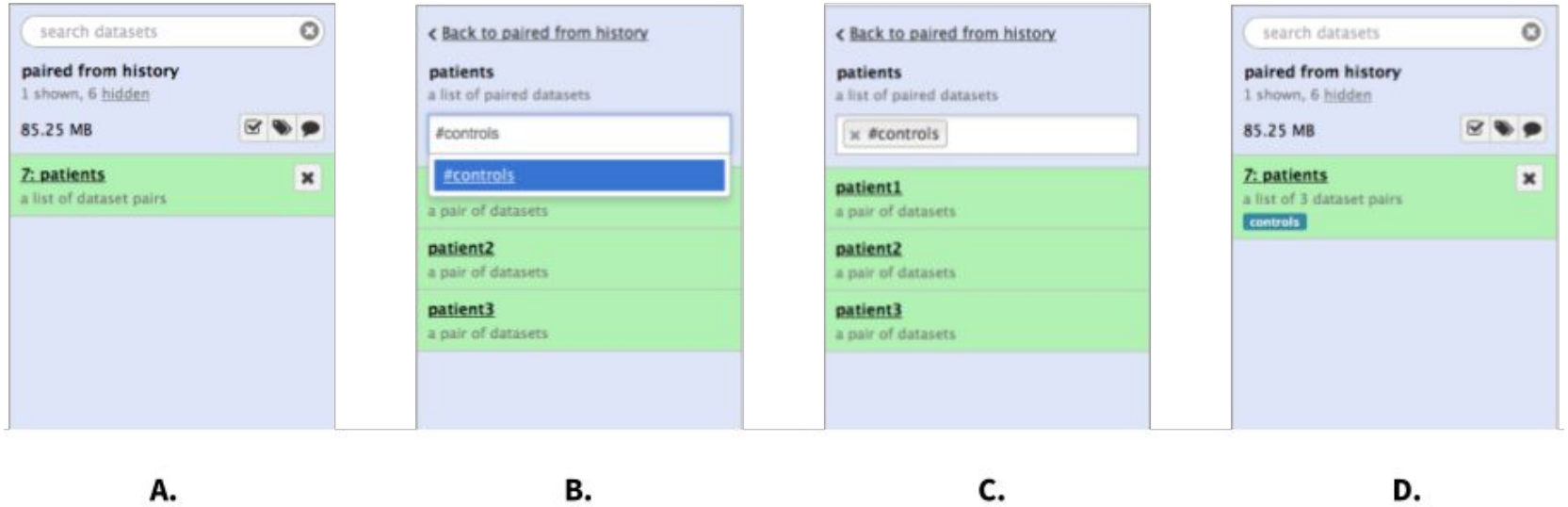
Renaming



A. **B.** **C.** **D.**

Dataset collection

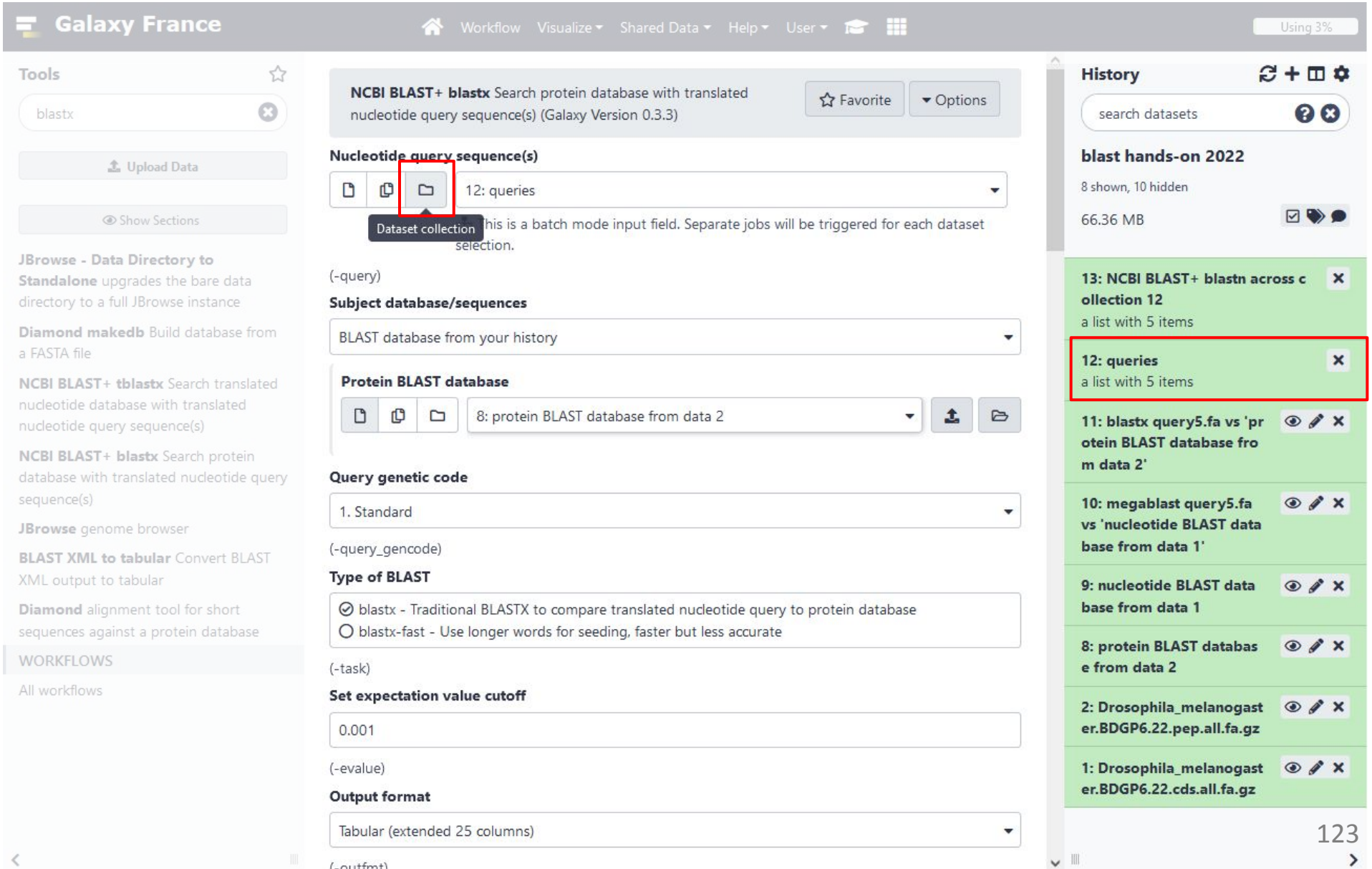
Tagging



A. **B.** **C.** **D.**

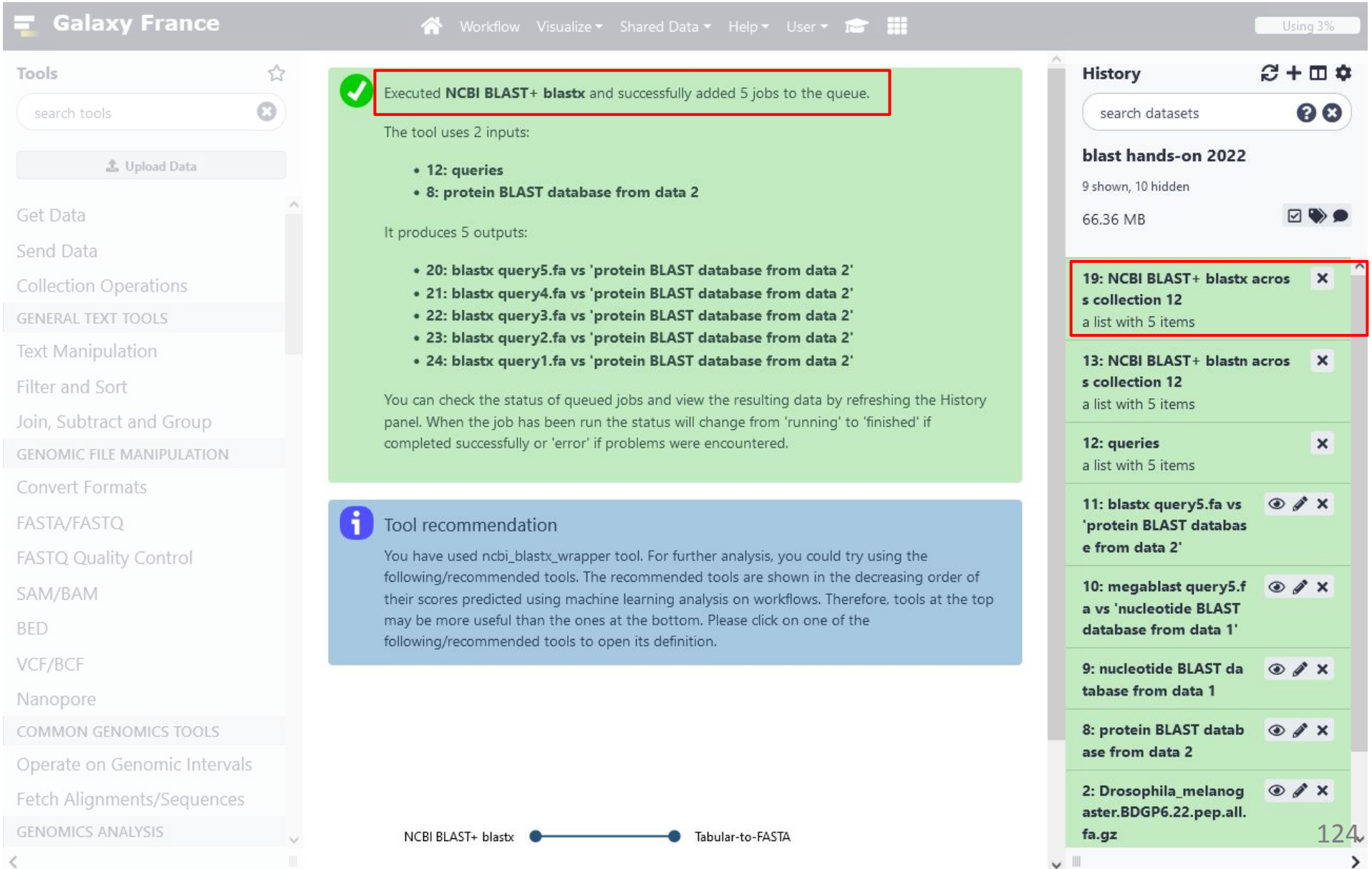
Dataset collection

Use a collection as input



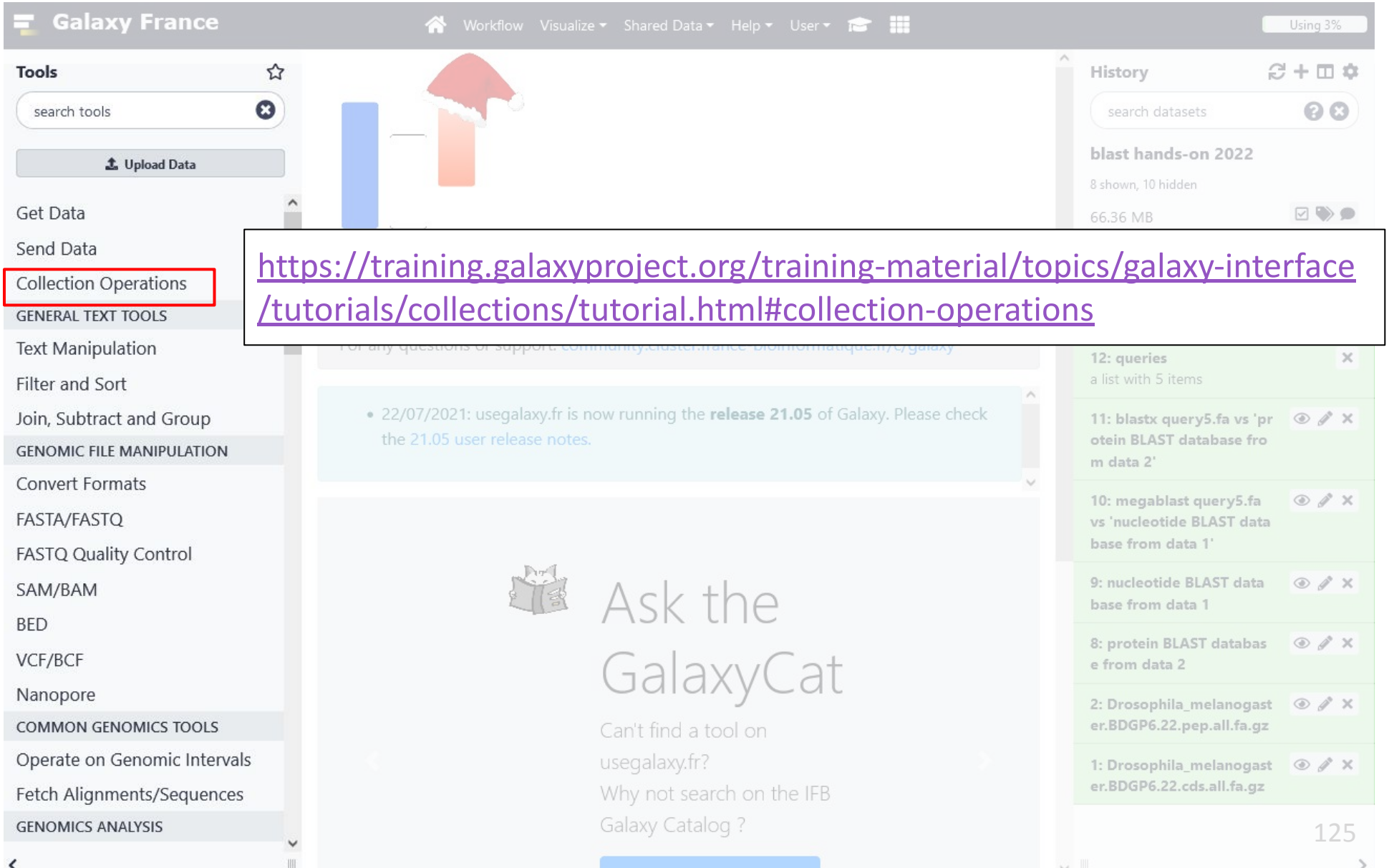
Dataset collection

Use a collection as input



Dataset collection

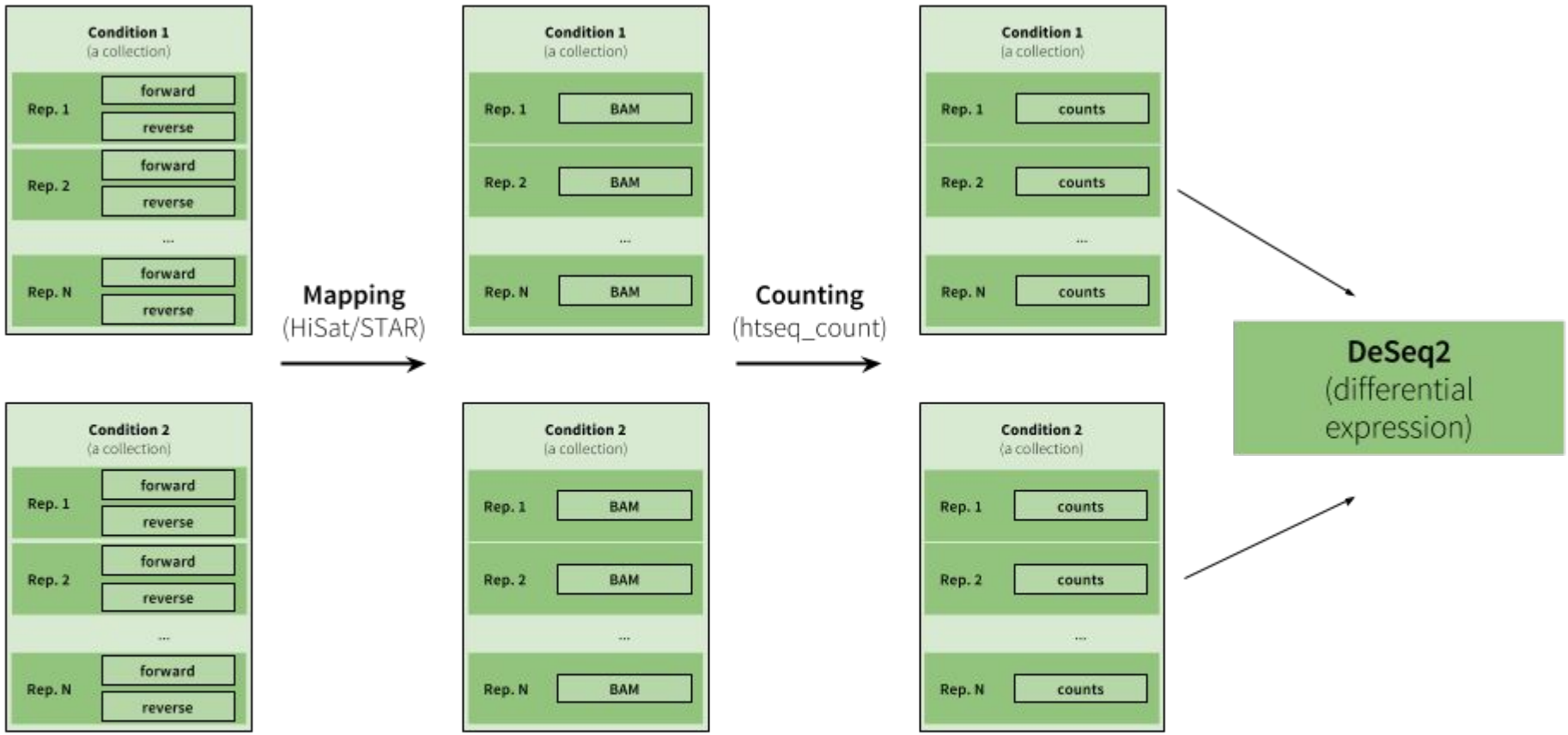
Use dataset collection tools



https://training.galaxyproject.org/training-material/topics/galaxy-interface/tutorials/collections/tutorial.html#collection-operations

Dataset collection

Use-case: RNA-seq analysis



Using dataset collections

Authors: Anton Nekrutenko | Add Contributions!

Overview

Questions:
- How to manipulate large numbers of datasets at once?

Objectives:
- Understand and master dataset collections

Time estimation: 30 minutes

Level: Intermediate

Supporting Materials:

Datasets | GTN Video Library

Last modification: Nov 16, 2021

Here we will show Galaxy features designed to help with the analysis of large numbers of samples. When you have just a few samples - clicking through them is easy. But once you've got hundreds - it becomes very annoying. In Galaxy we have introduced **Dataset collections** that allow you to combine numerous datasets in a single entity that can be easily manipulated.

Getting data

First, we need to upload datasets. Cut and paste the following URLs to Galaxy upload tool (see a **Tip** on how to do this below).

```
https://zenodo.org/record/5119008/files/M117-bl_1.fq.gz
https://zenodo.org/record/5119008/files/M117-bl_2.fq.gz
https://zenodo.org/record/5119008/files/M117-ch_1.fq.gz
https://zenodo.org/record/5119008/files/M117-ch_2.fq.gz
https://zenodo.org/record/5119008/files/M117C1-bl_1.fq.gz
```

https://training.galaxyproject.org/training-material/topics/galaxy-interface/tutorials/collections/tutorial.html

Hands-on

COLLECTION

Collection – Hands-on

Part 2:

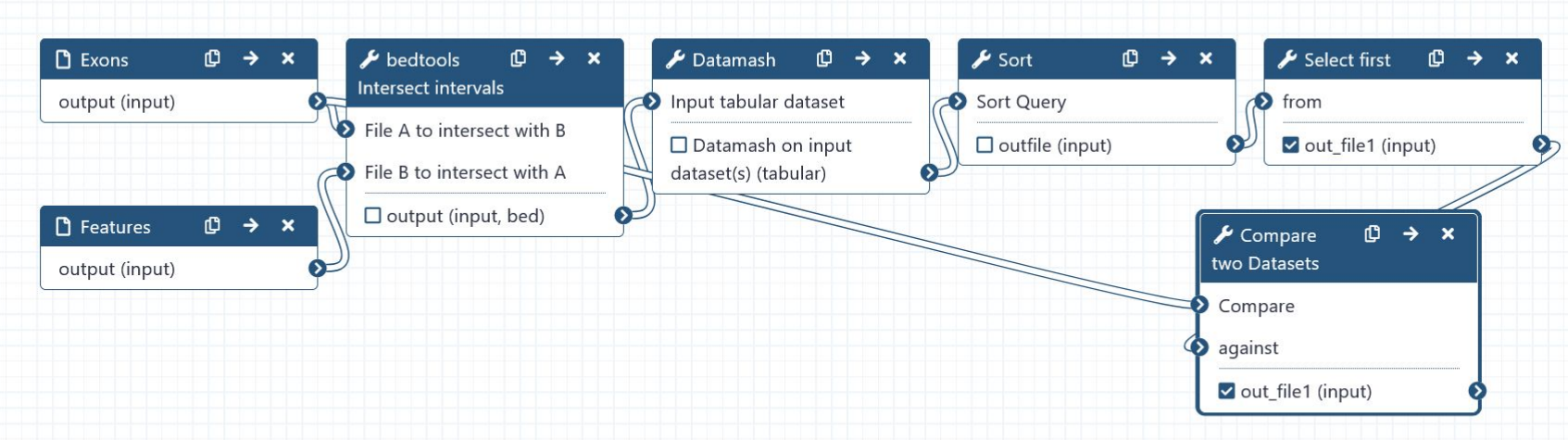
You have new sequences that you want to compare with the gene and protein databases from your favorite species (BLAST).

Case of Drosophila melanogaster (dm)

1. Create new history
2. Import CDS and peptide sequences databases
3. Import query sequences
4. Make BLAST databases
5. Run BLAST against the CDS database
6. Run BLAST against the protein database
7. Create a dataset list with all the query sequences
8. Run BLAST against the CDS database on the dataset list

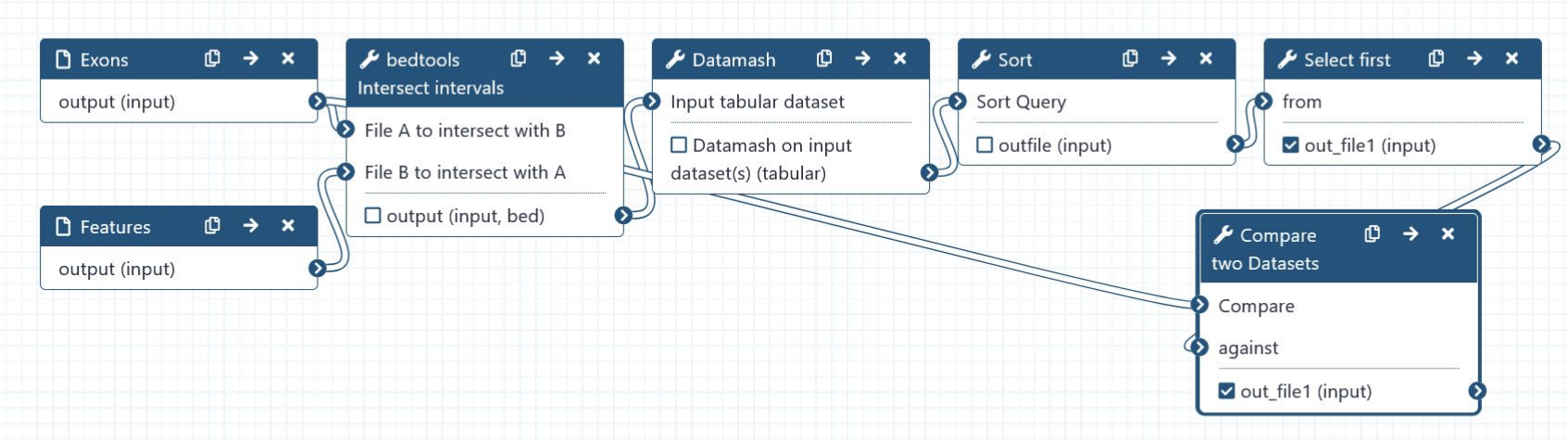
WORKFLOW

Workflow



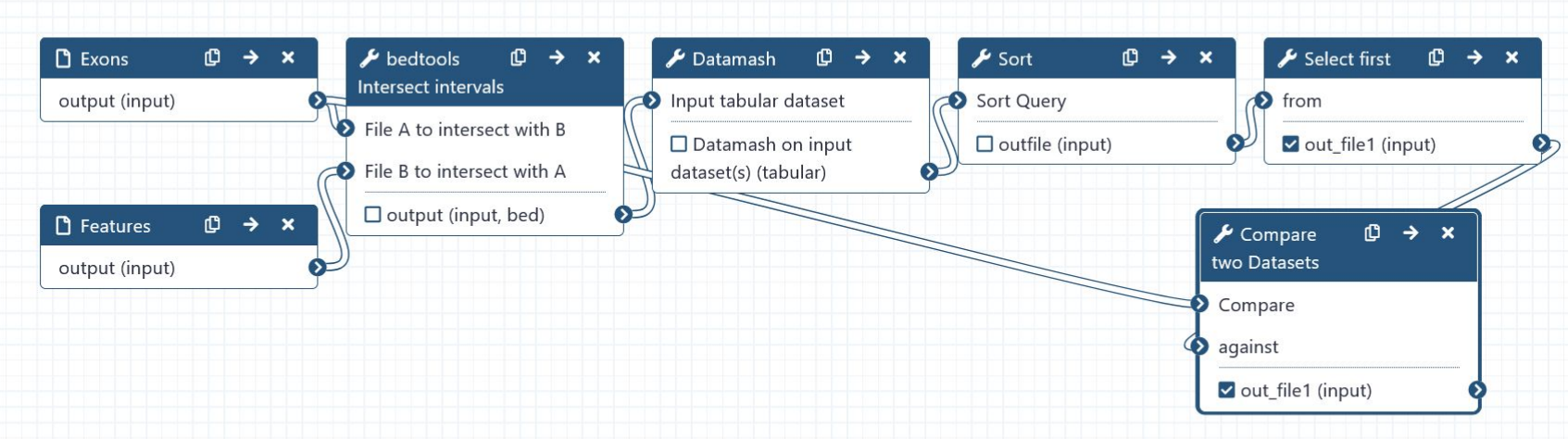
- A workflow is a sequence of tool operations and parameters
- A workflow is built to be replayed (more or less strict)

Workflow



- Extracted from a history
- Built manually by adding and configuring tools using the canvas
- Imported using an existing shared workflow

Workflow

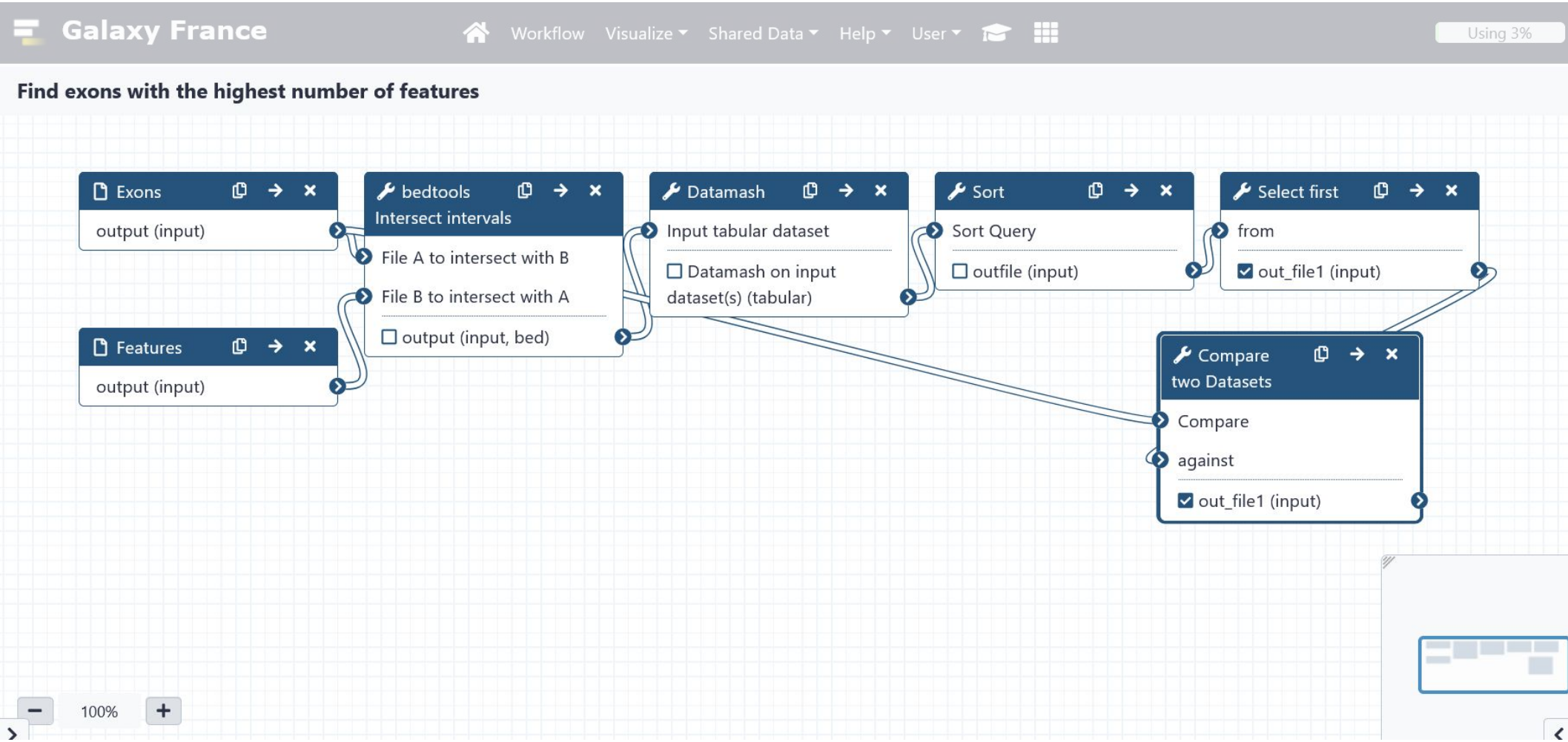


Why would you want to create workflows?

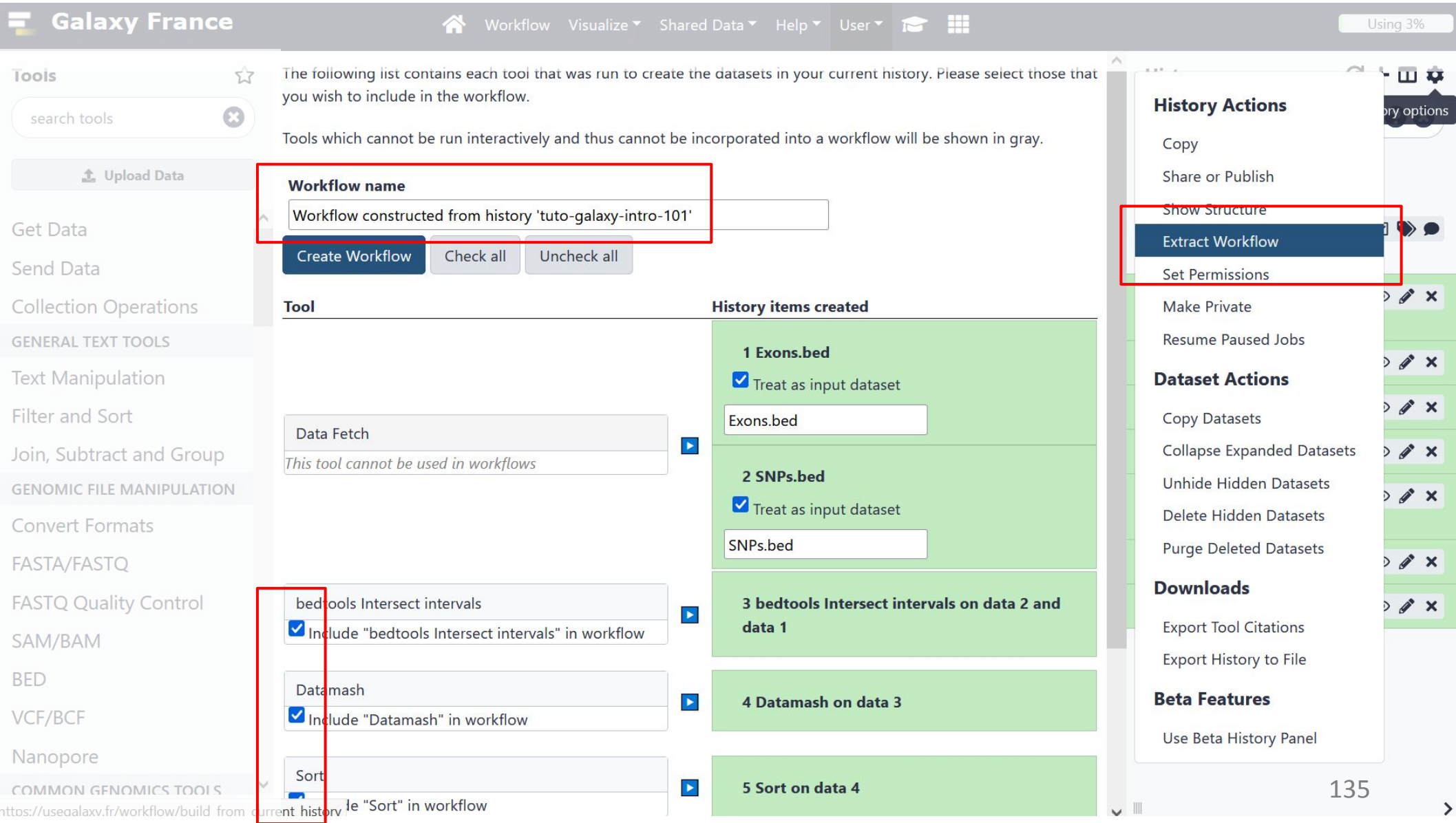
- Re-run the same analysis on different input data sets
- Change parameters before re-running a similar analysis
- Make use of the workflow job scheduling (jobs are submitted as soon as their inputs are ready)
- Share workflows for publication and with the community

Workflow

Our workflow with Galaxy

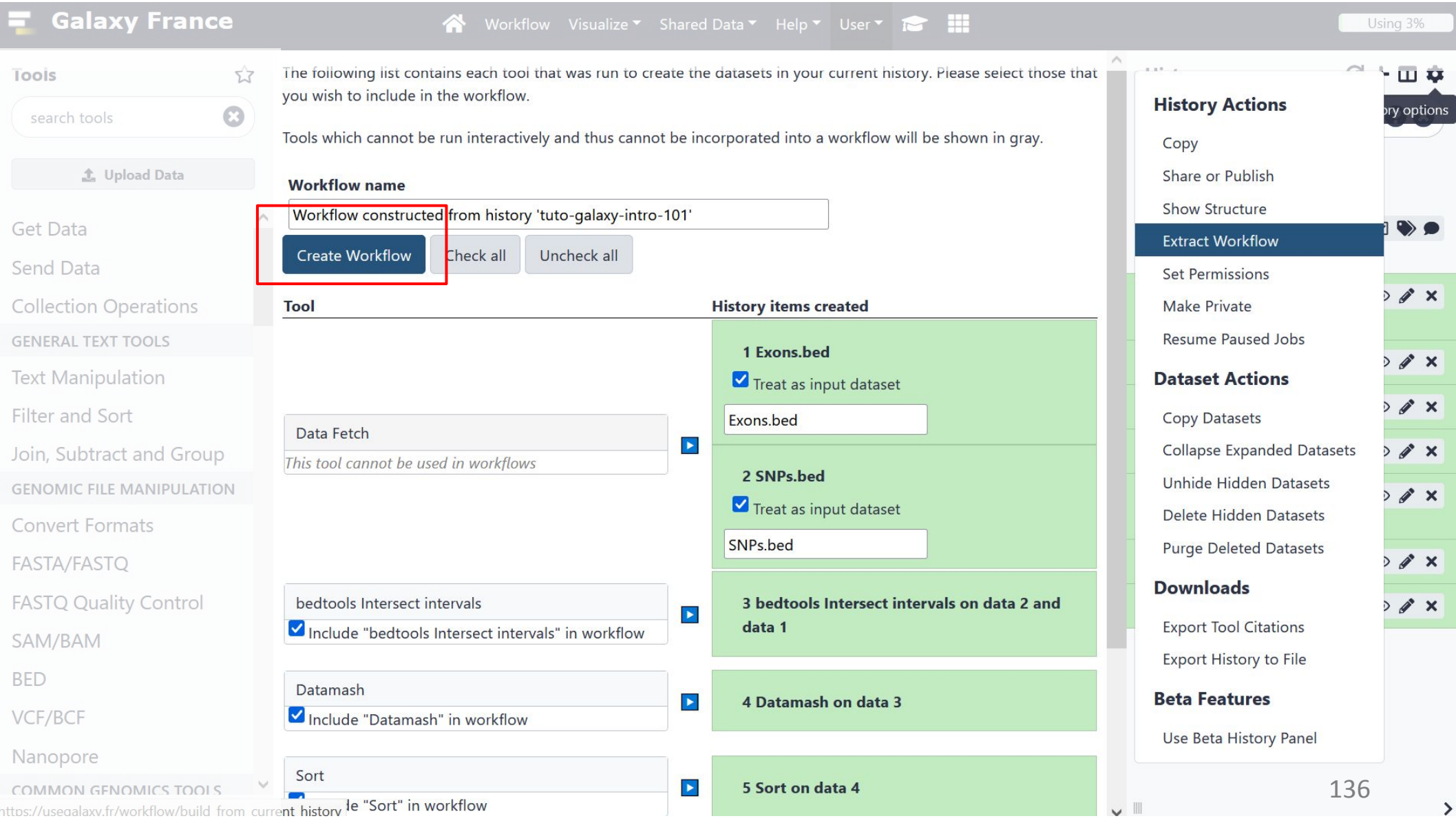


From history



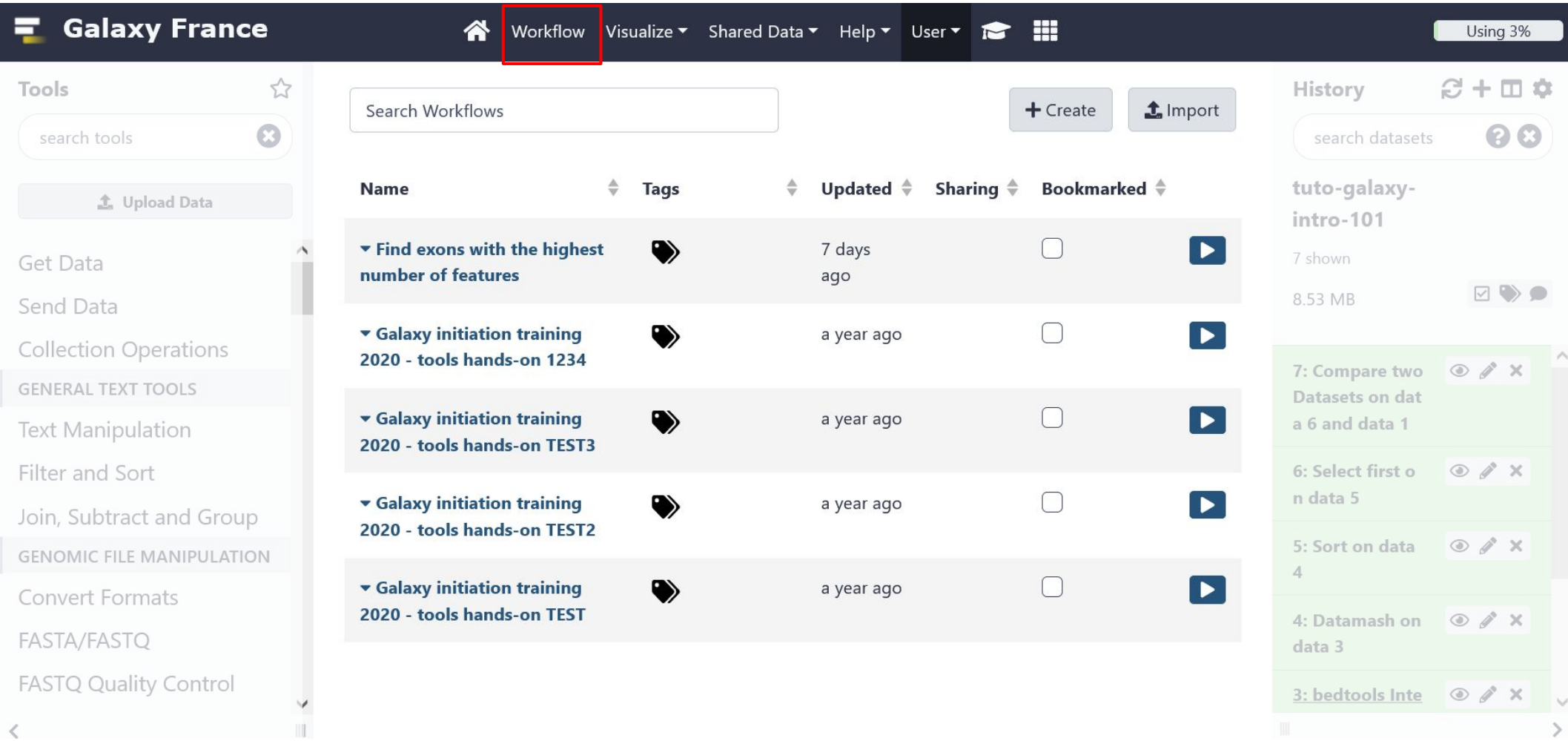
Workflow

From history



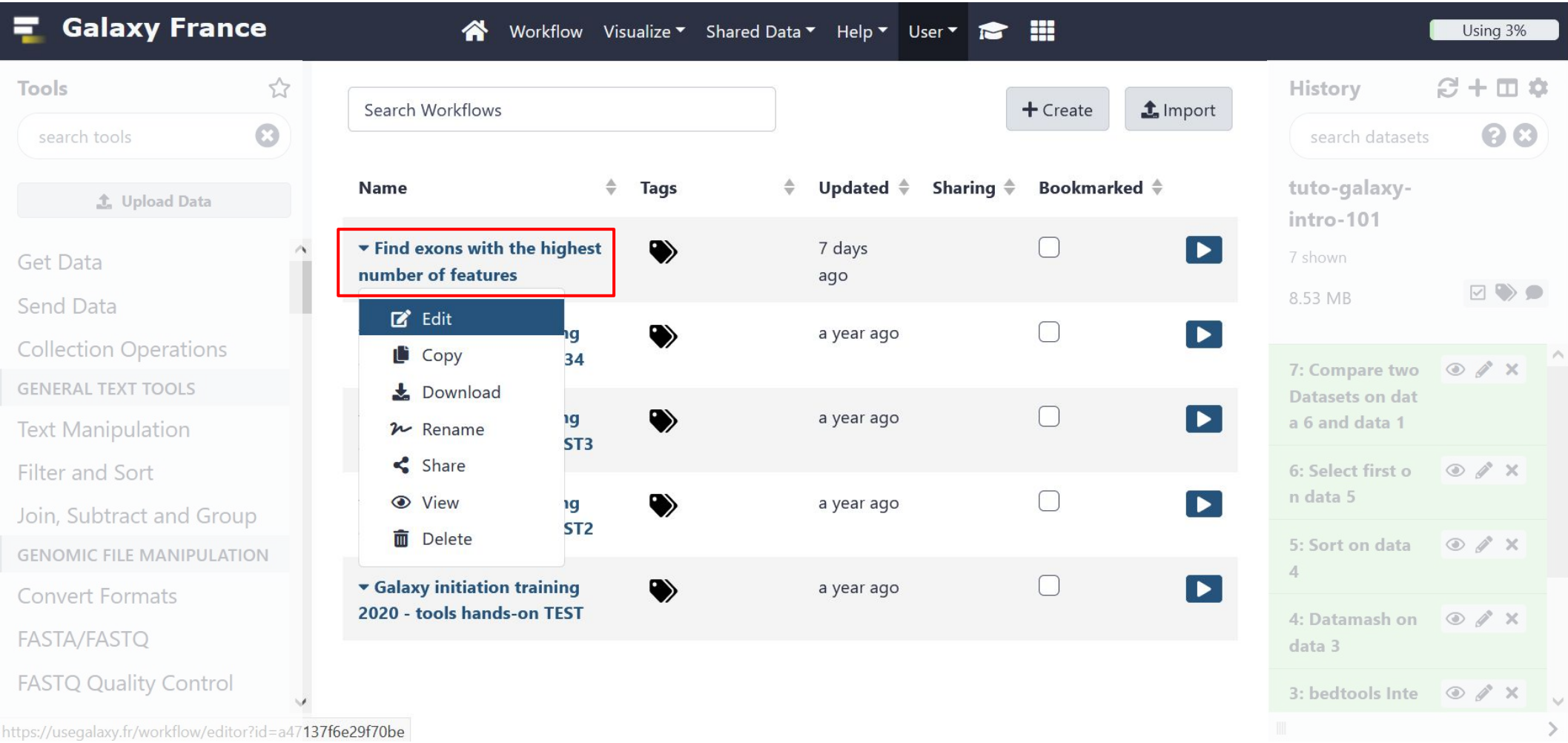
Workflow

Workflow manager

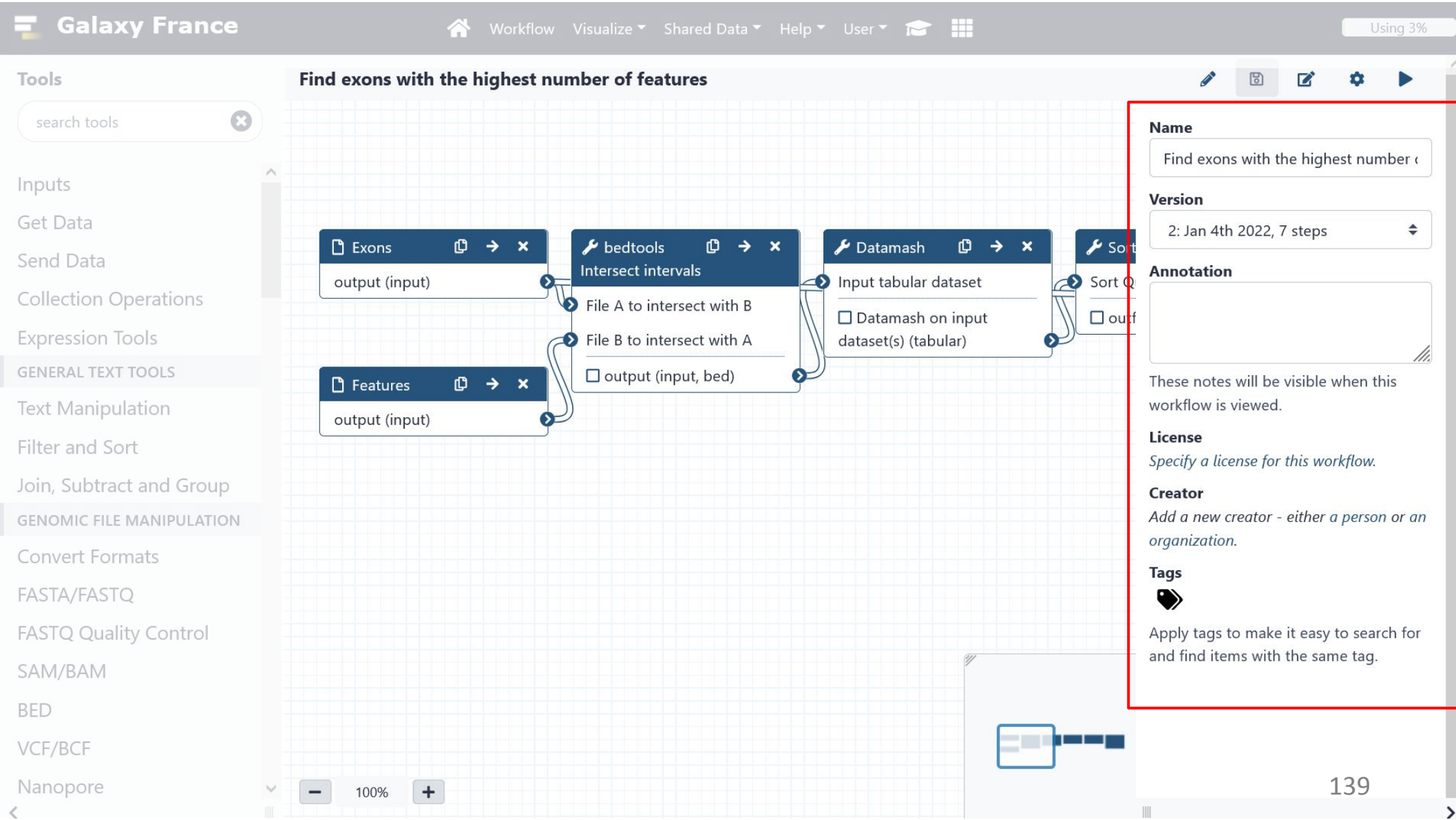


Workflow

Workflow manager

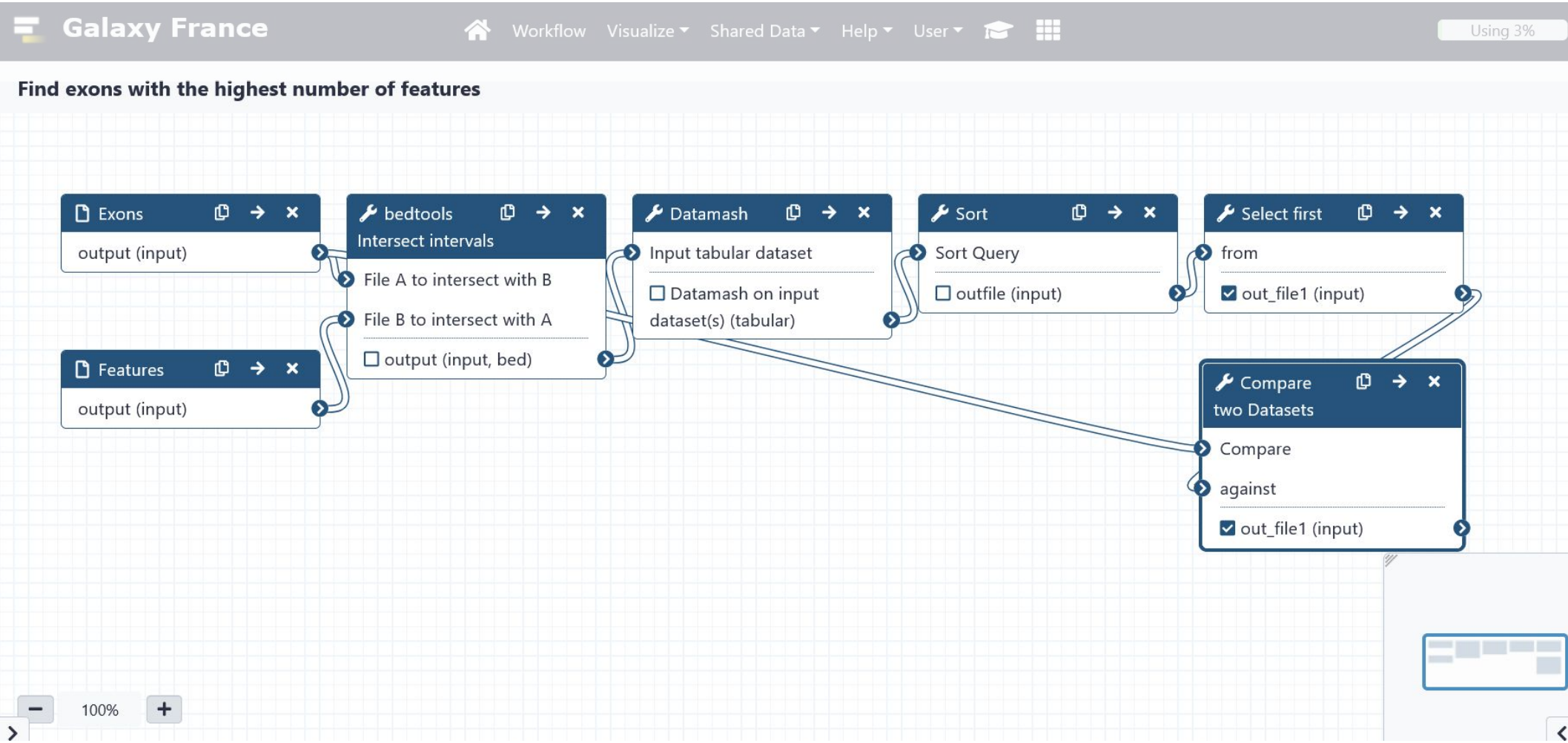


Edit a workflow: attributes



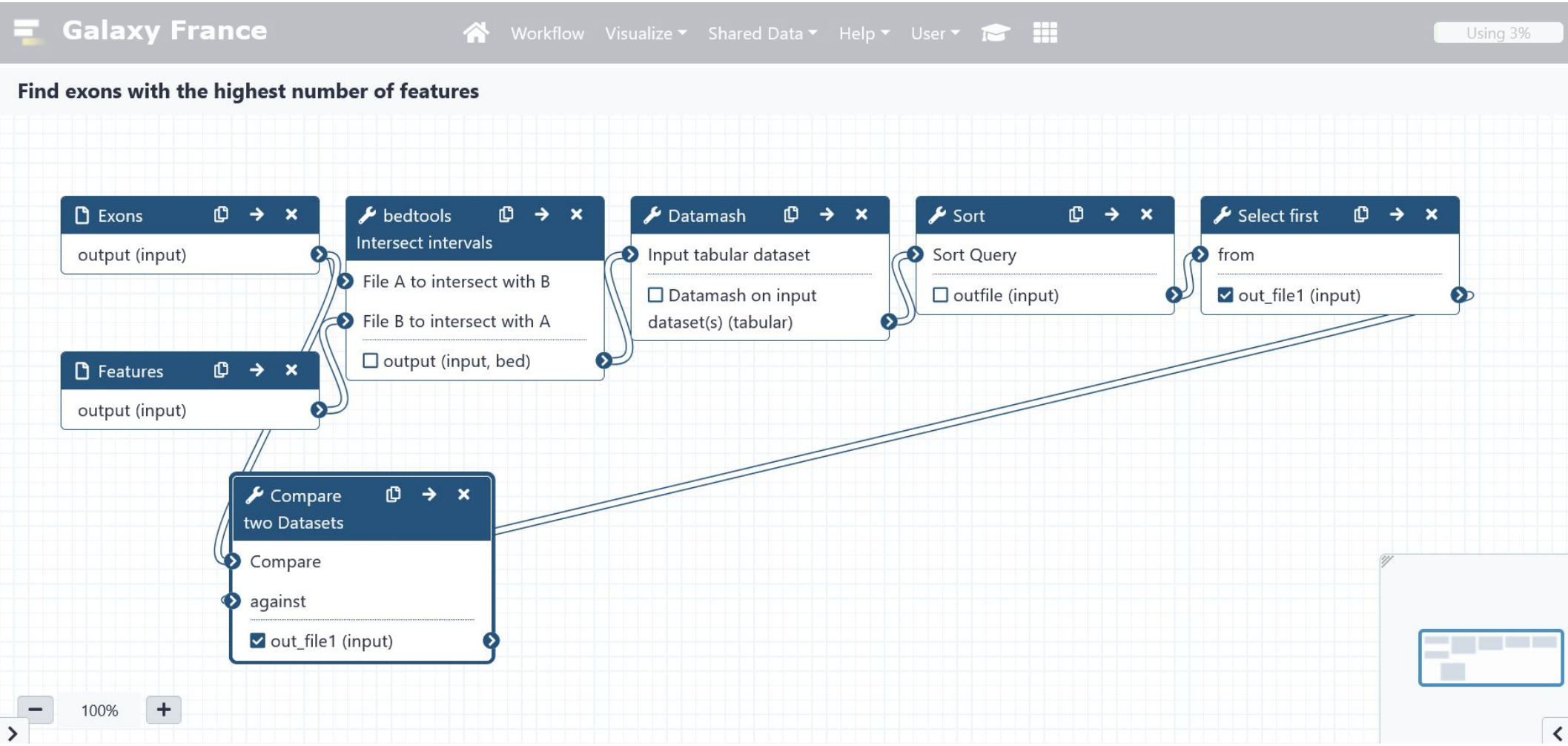
Workflow

Edit a workflow: drag and drop



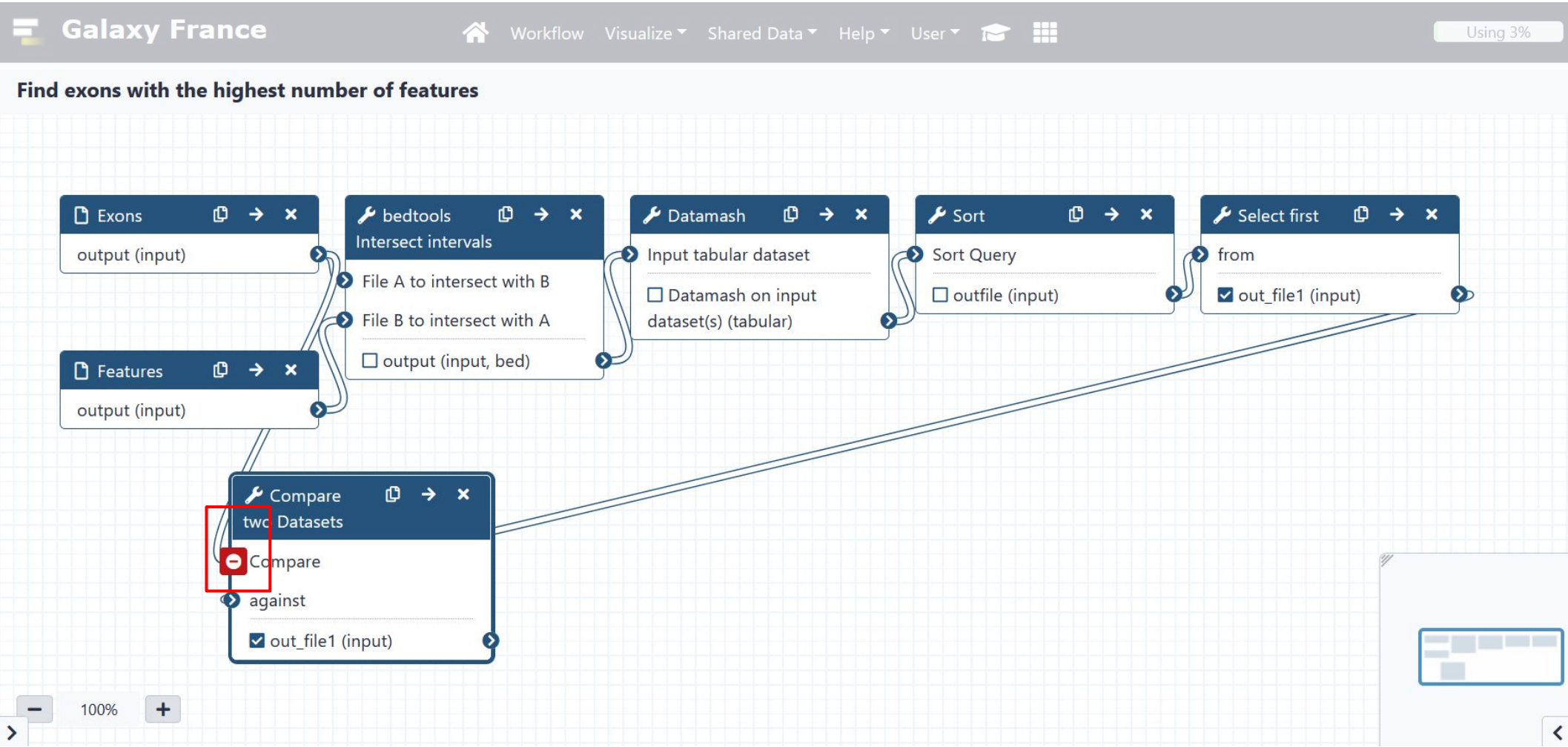
Workflow

Edit a workflow: drag and drop



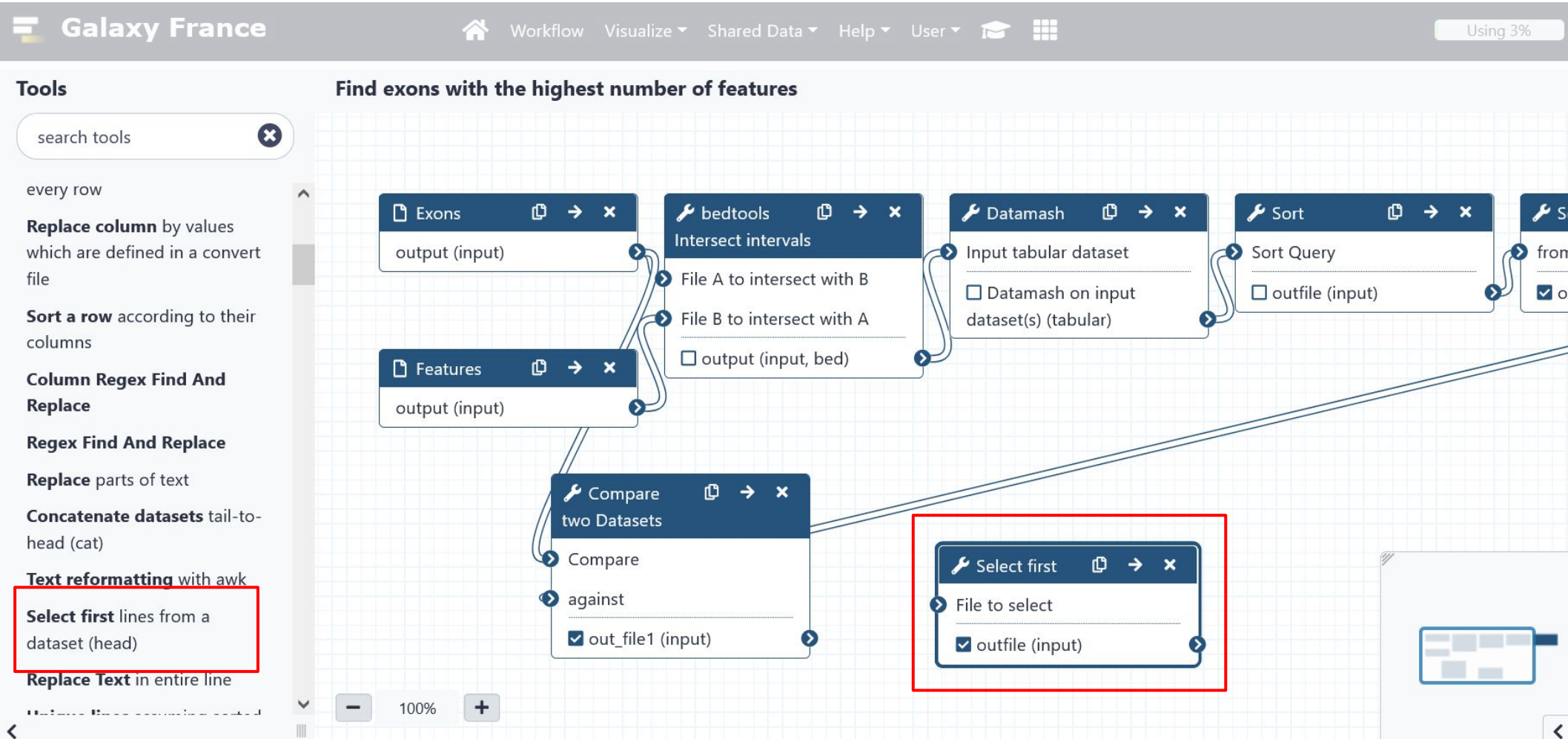
Workflow

Edit a workflow: delete a noodle



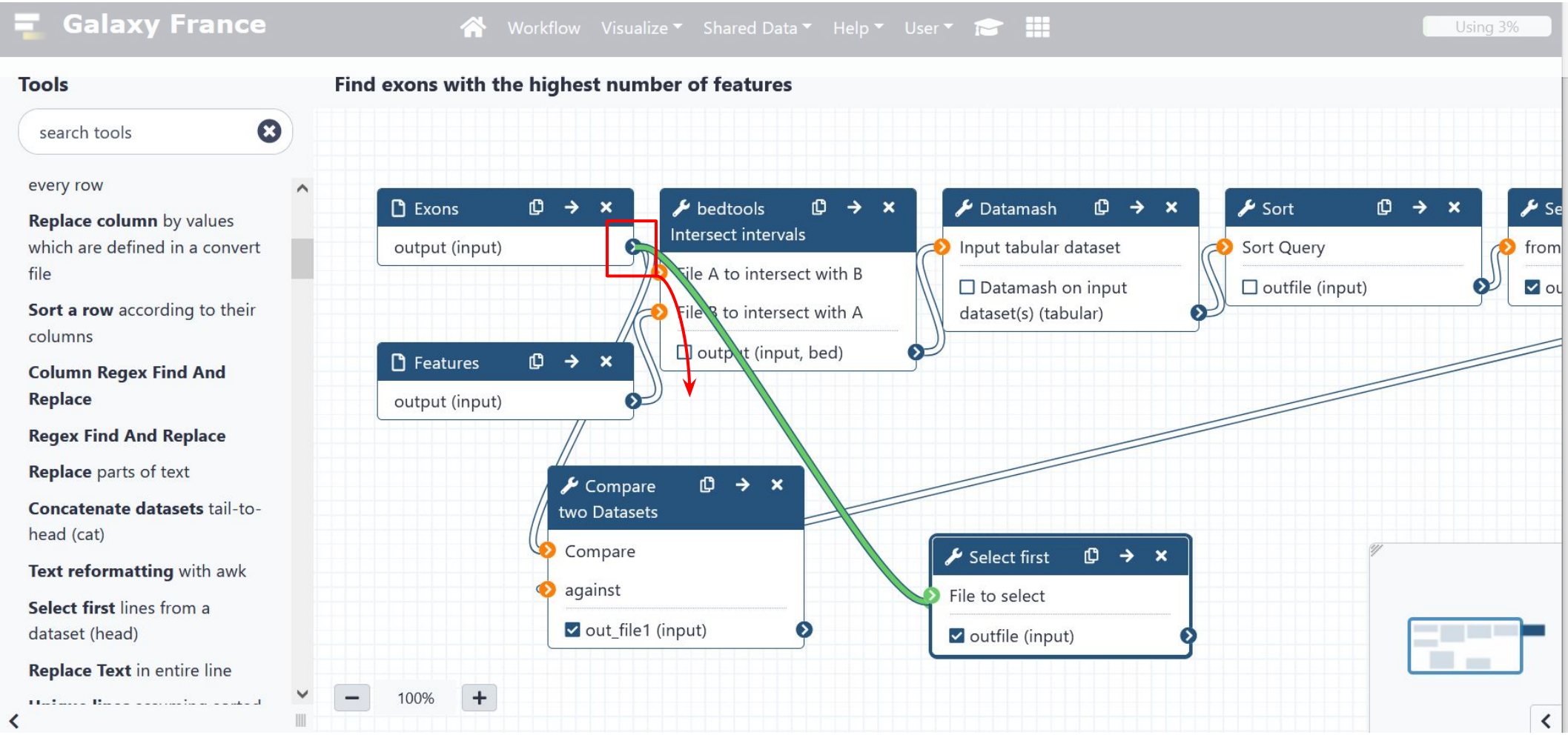
Workflow

Edit a workflow: add a tool

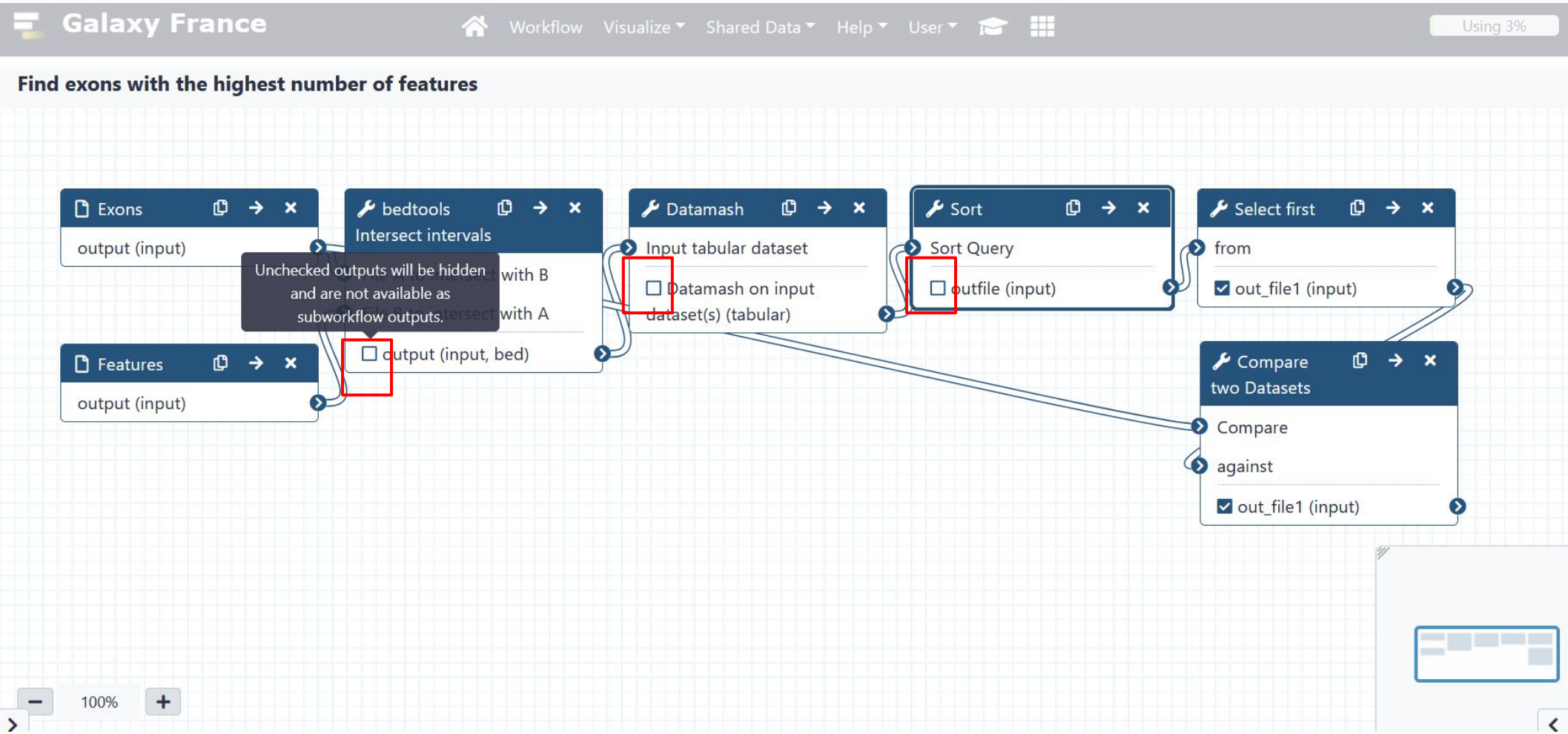


Workflow

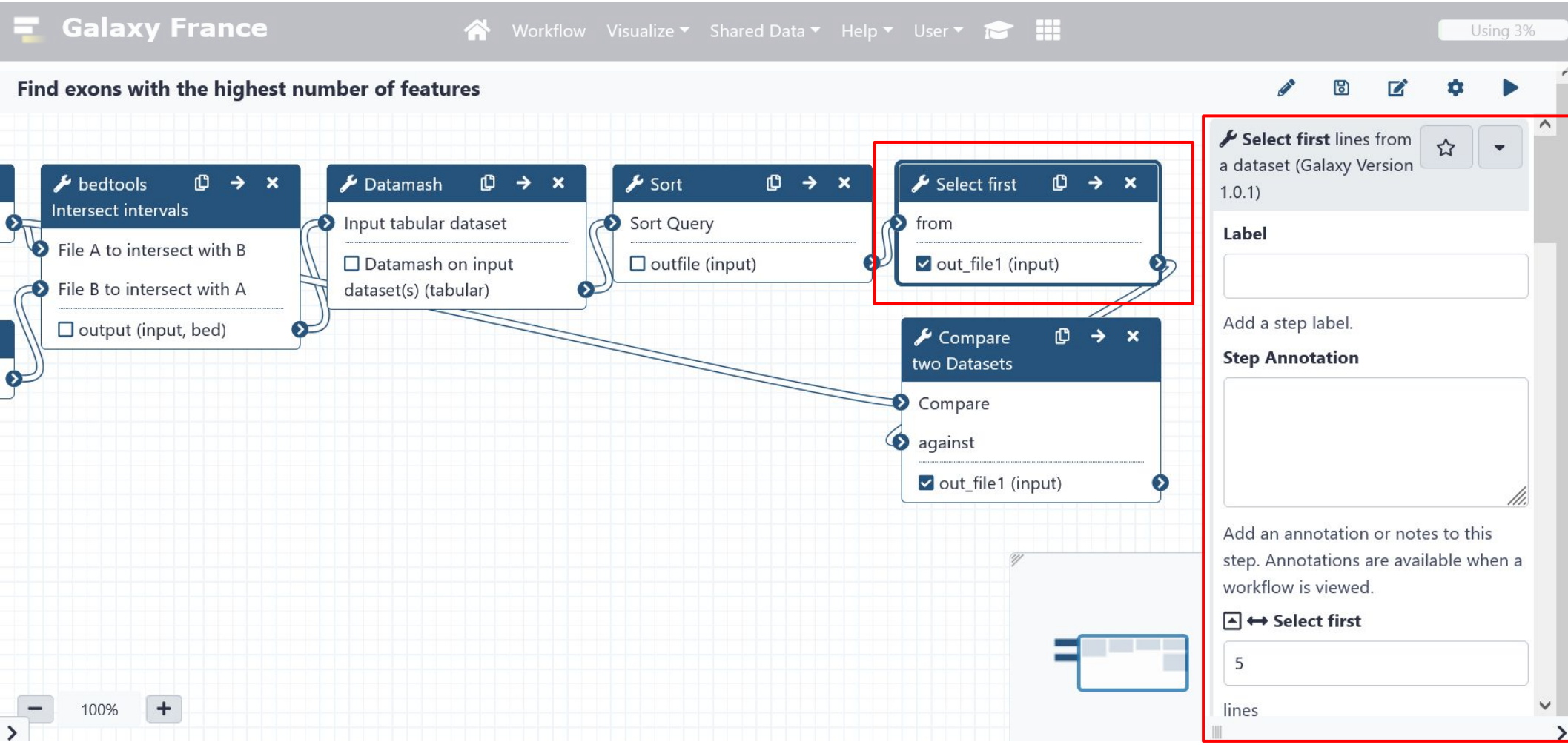
Edit a workflow: add a noodle



Edit a workflow: hide intermediate steps

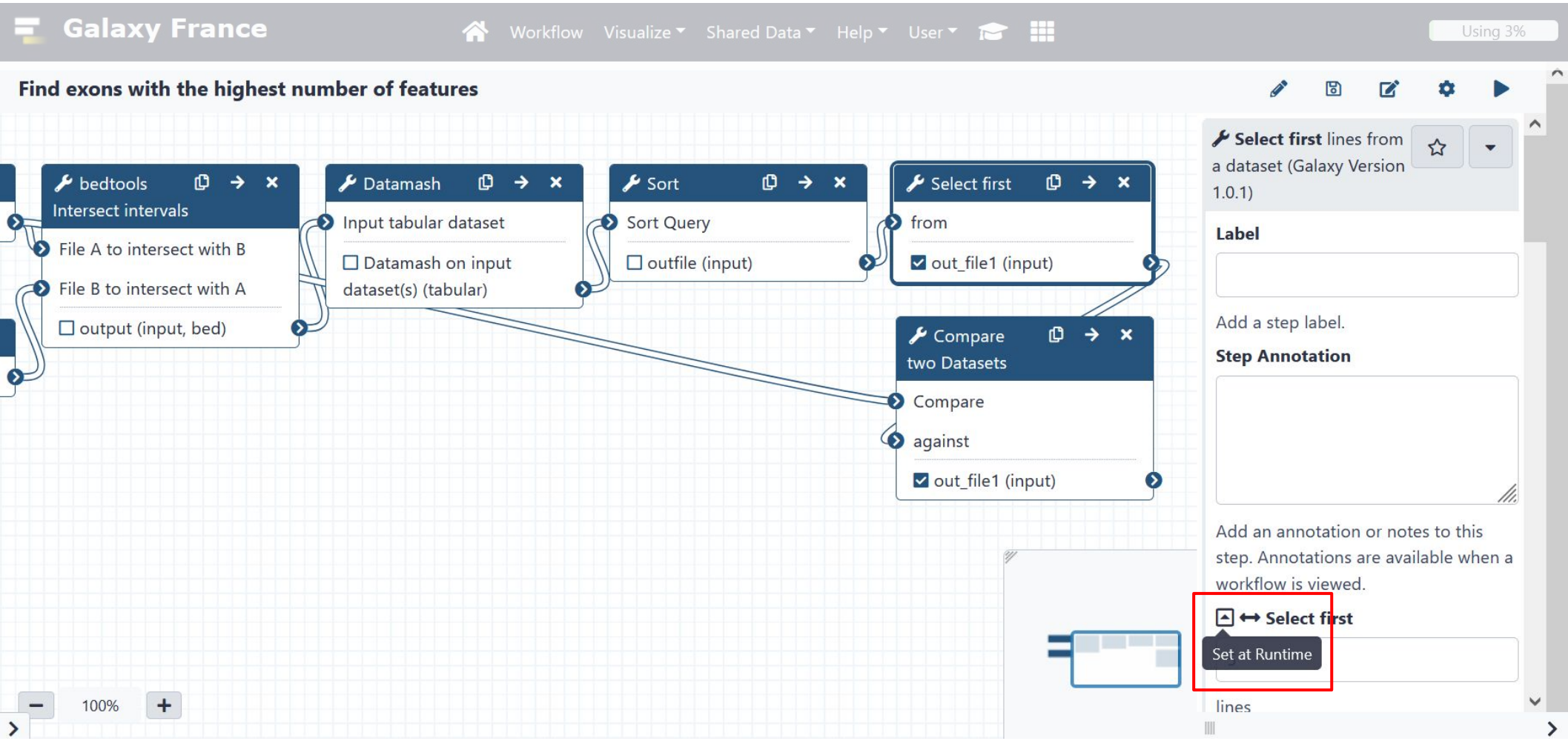


Edit a workflow: set or release a parameter

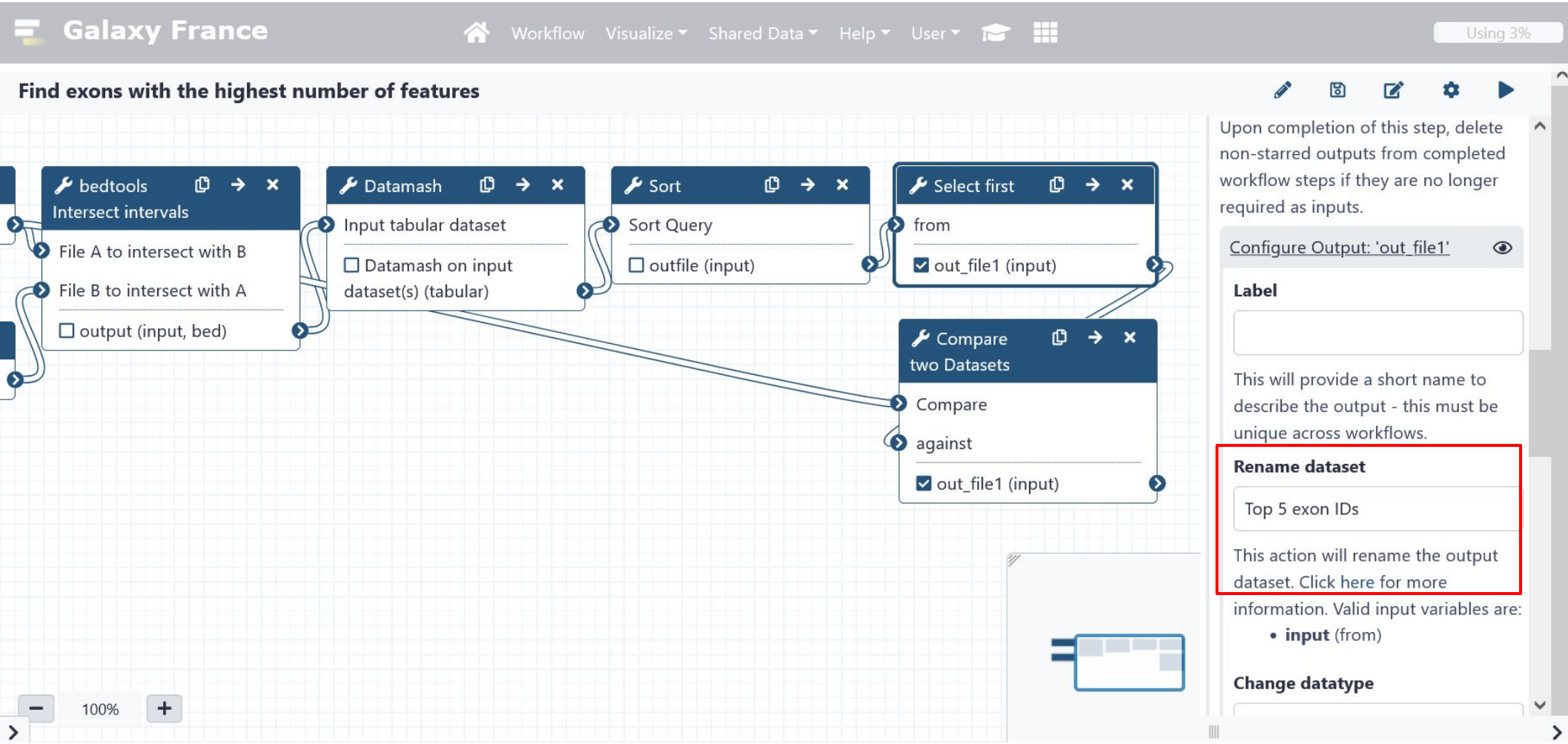


Workflow

Edit a workflow: set or release a parameter

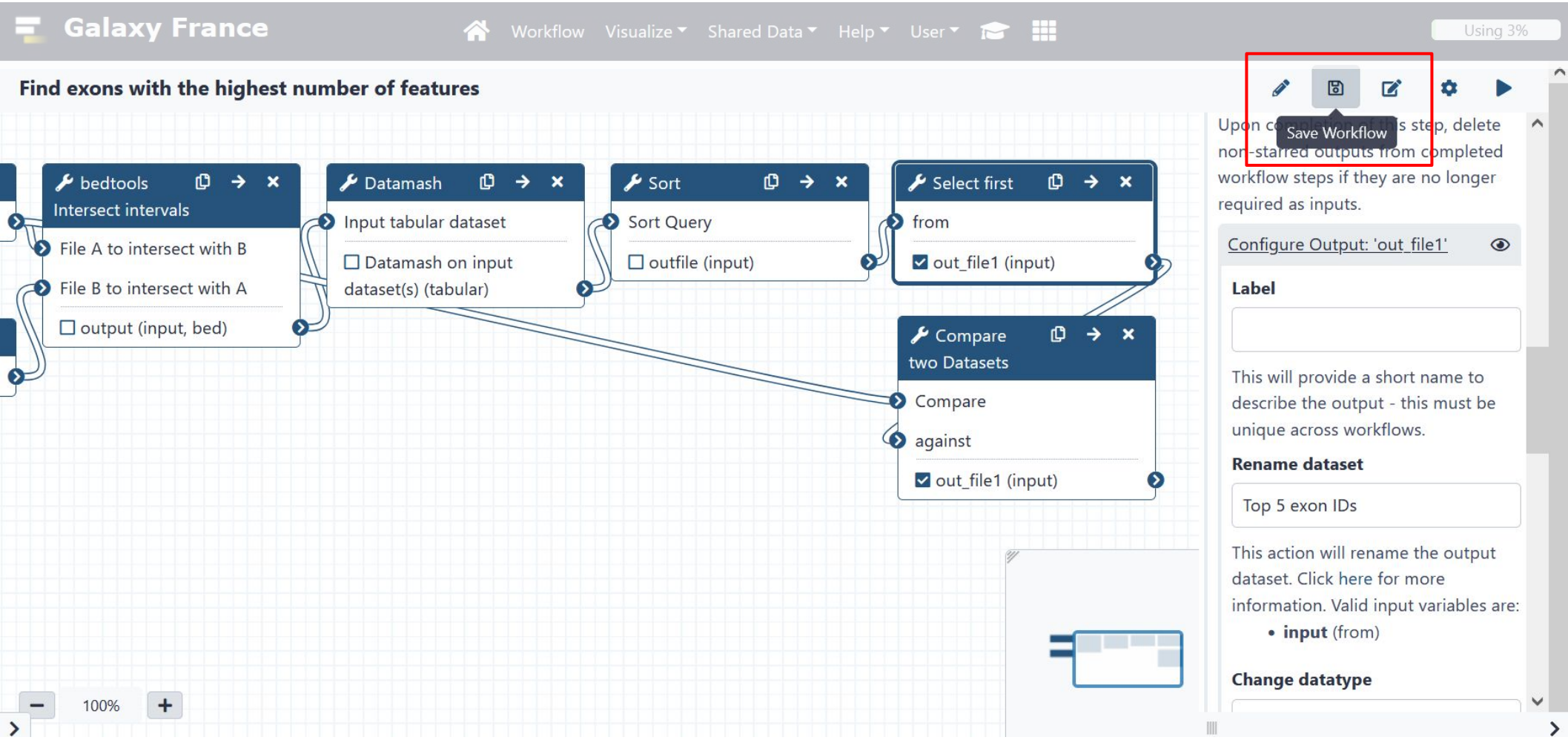


Edit a workflow: rename the outputs



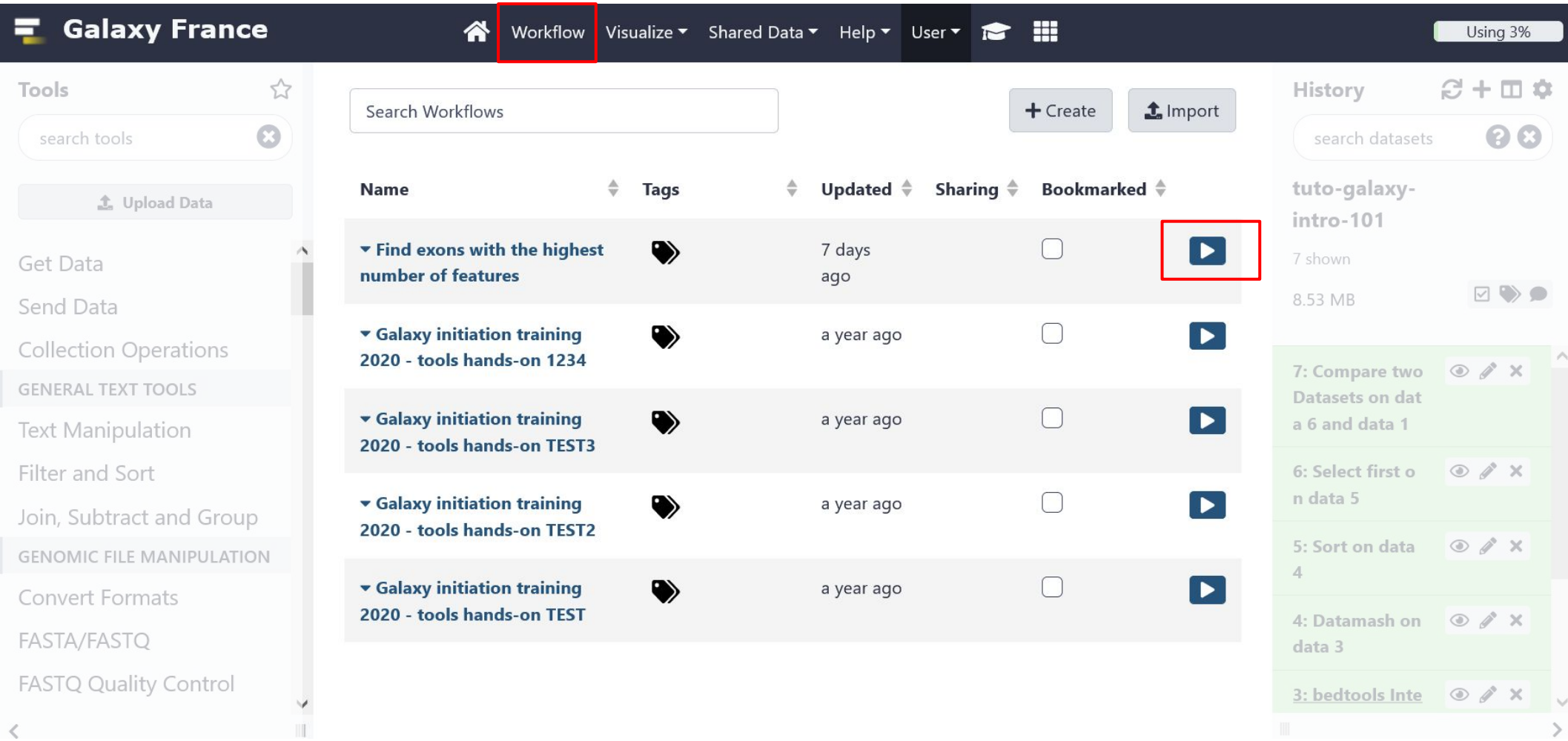
Workflow

Save

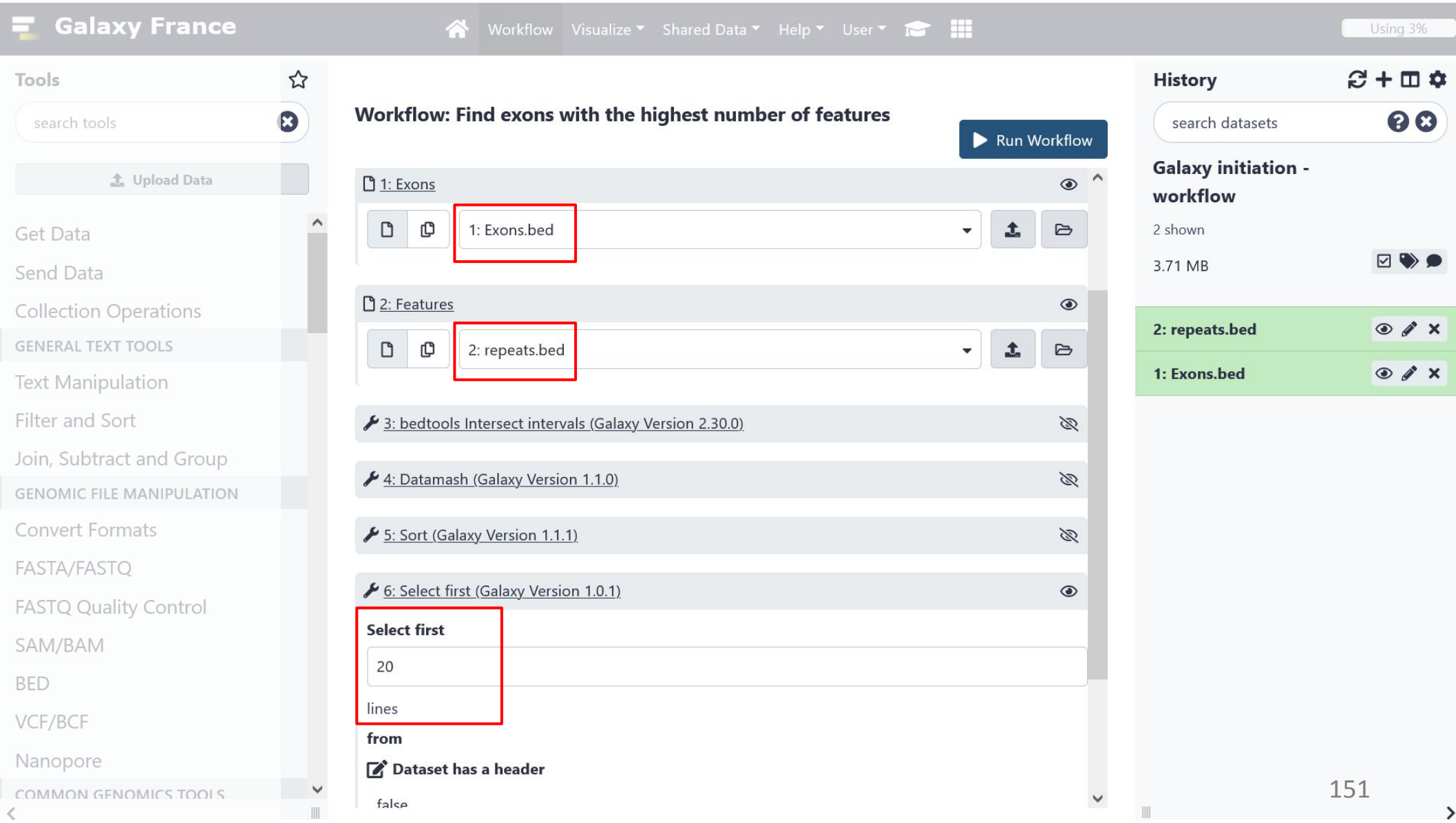


Workflow

Run a workflow

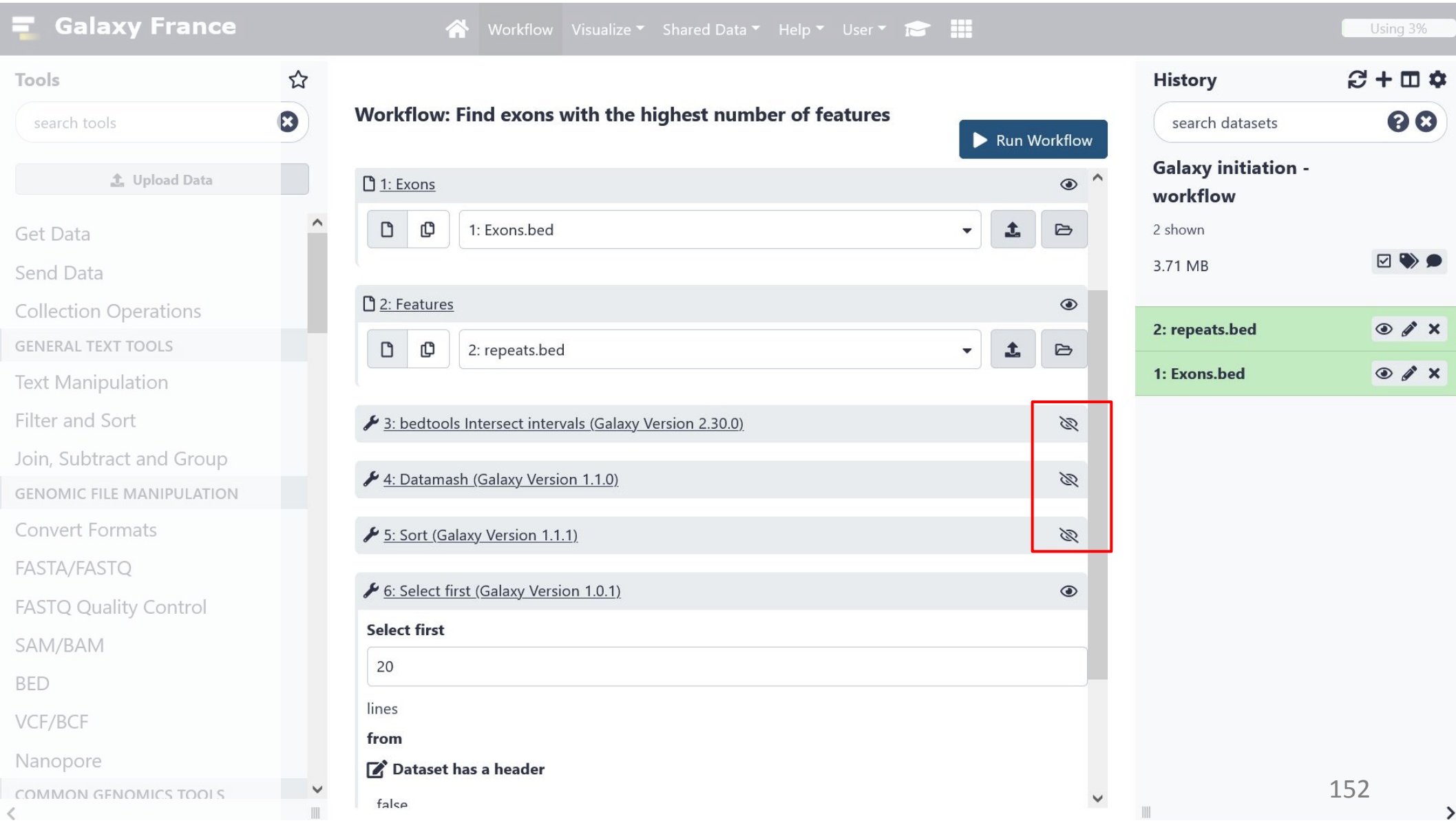


Run a workflow



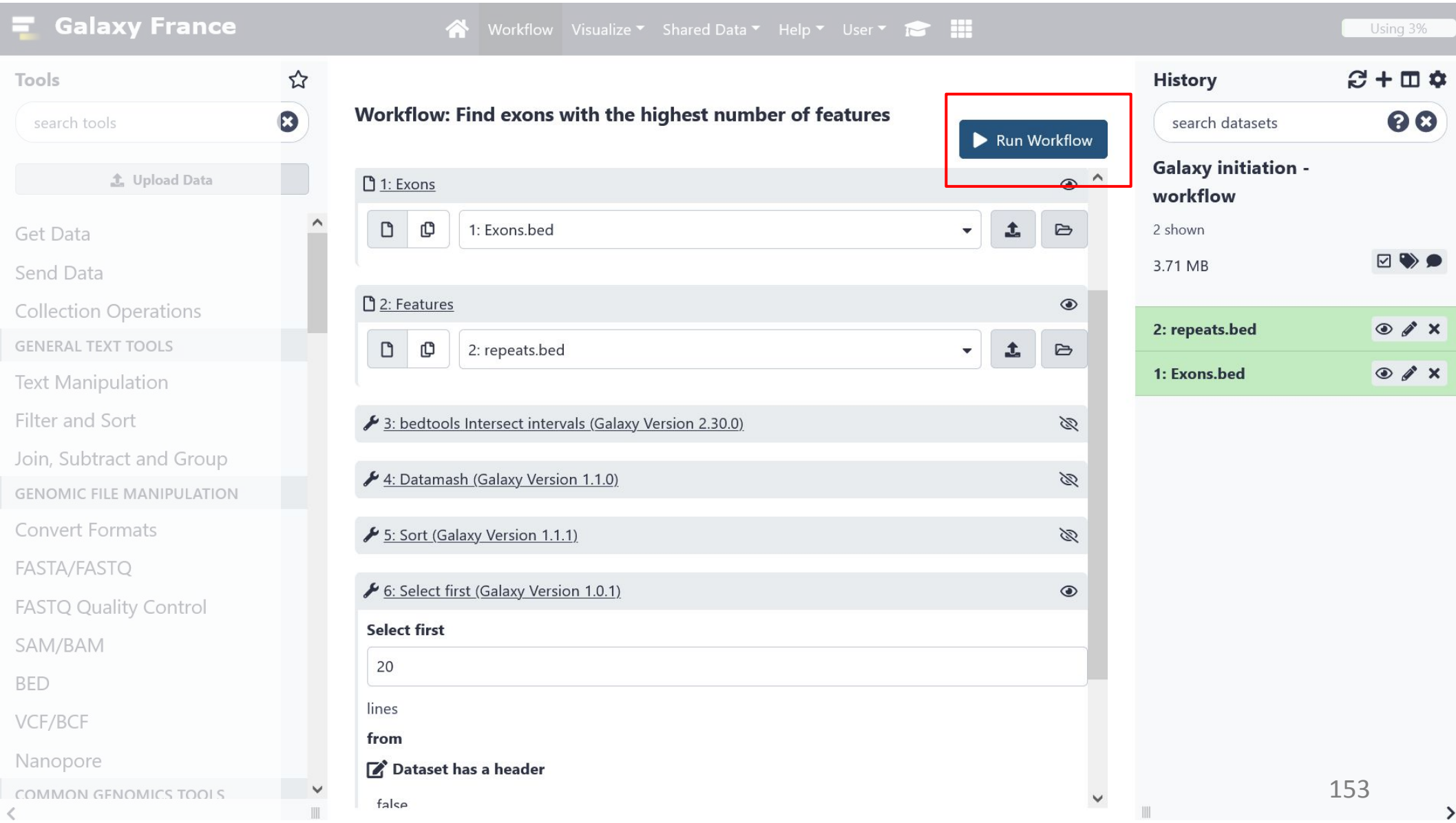
Workflow

Run a workflow



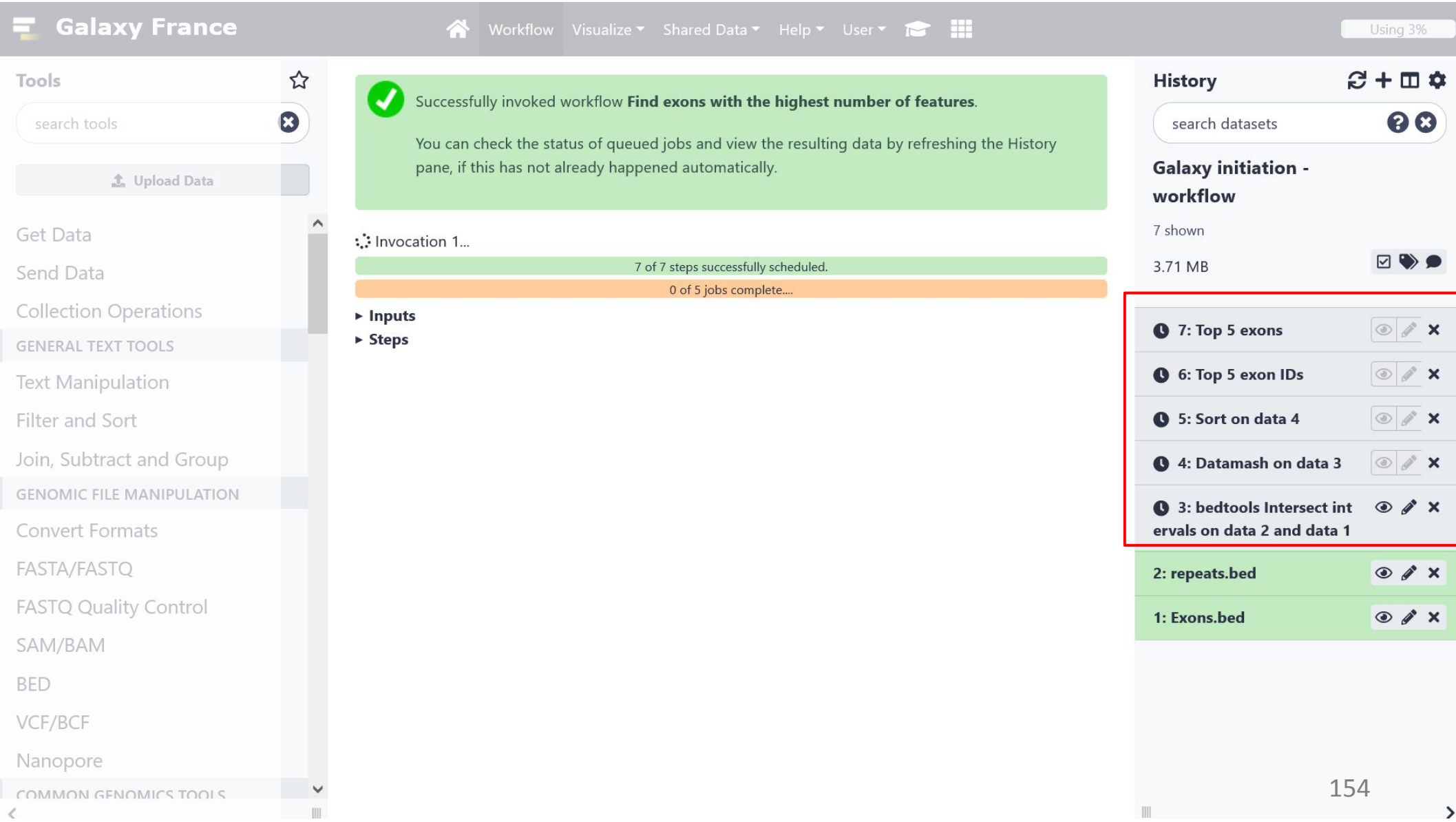
Workflow

Run a workflow

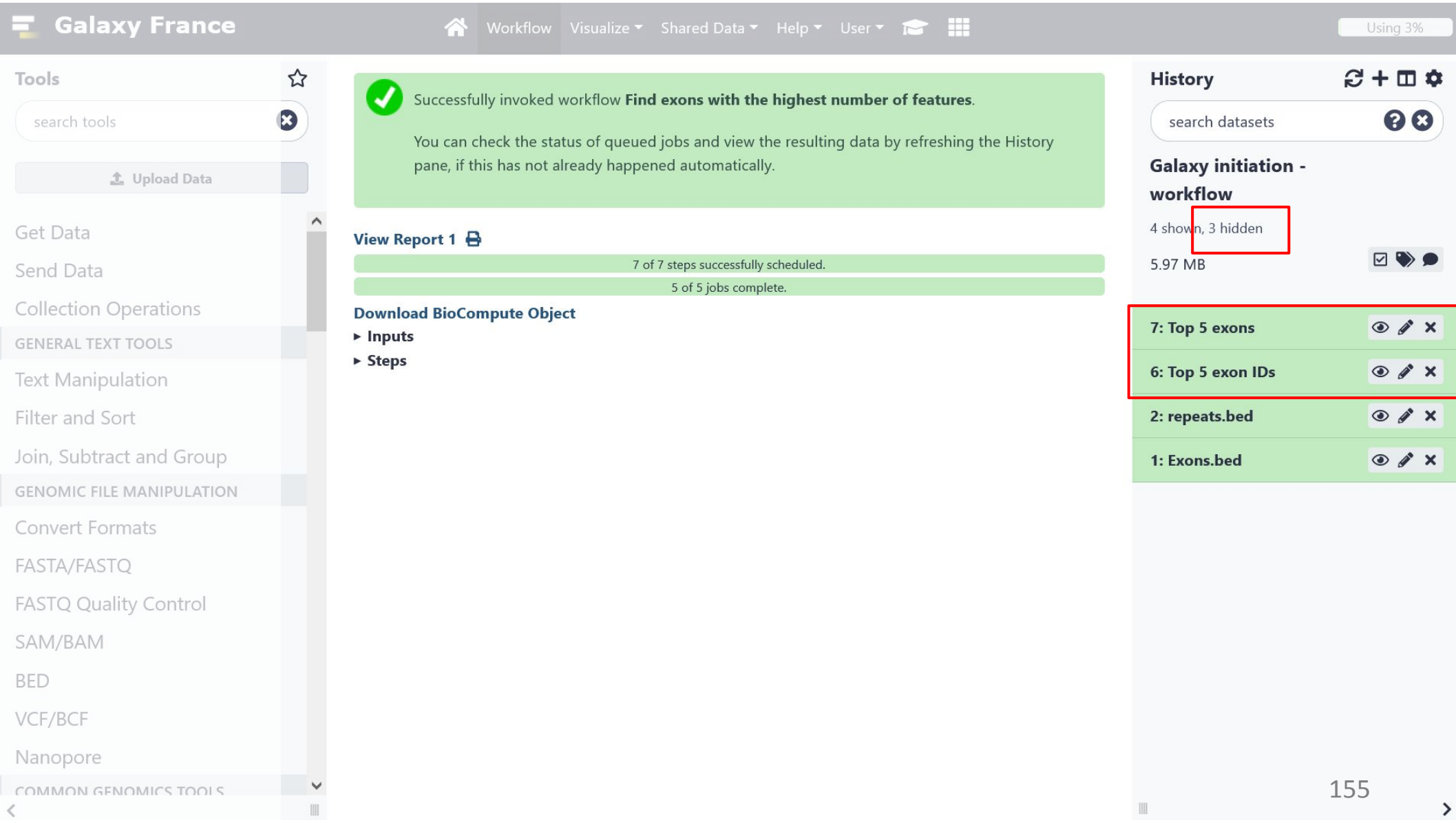


Workflow

Run a workflow

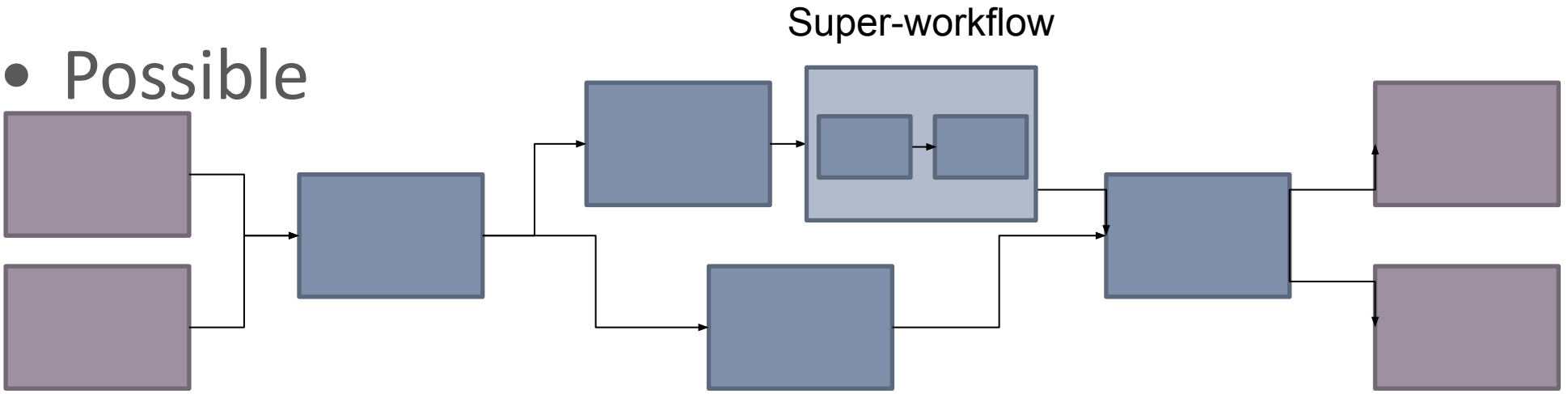


Run a workflow

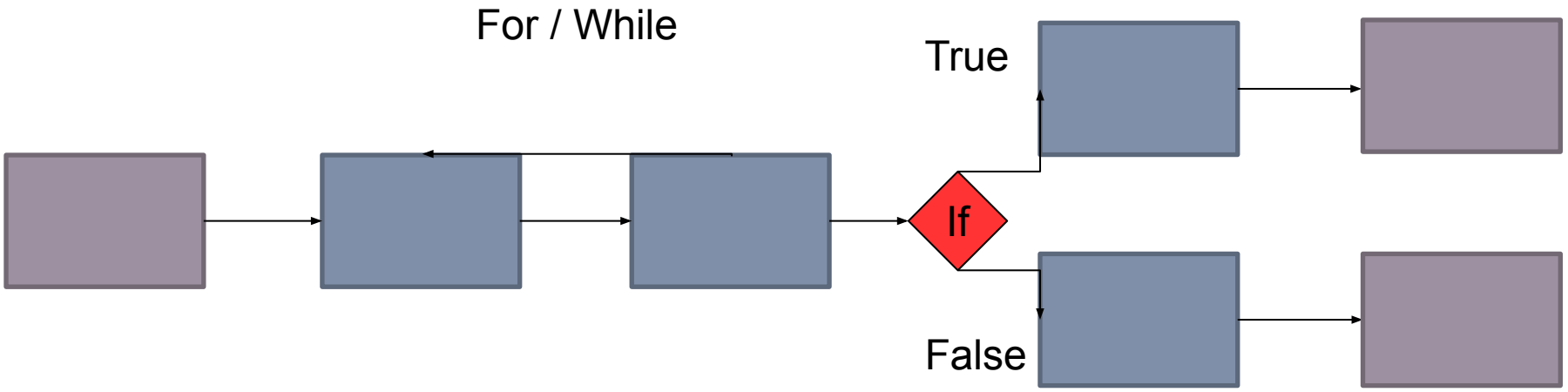


Workflow

- Possible



- Impossible (until now)



Last but not least!

SHARE

biologist ↔ biologist

- Sharing histories or datasets
  - With or without linked workflow

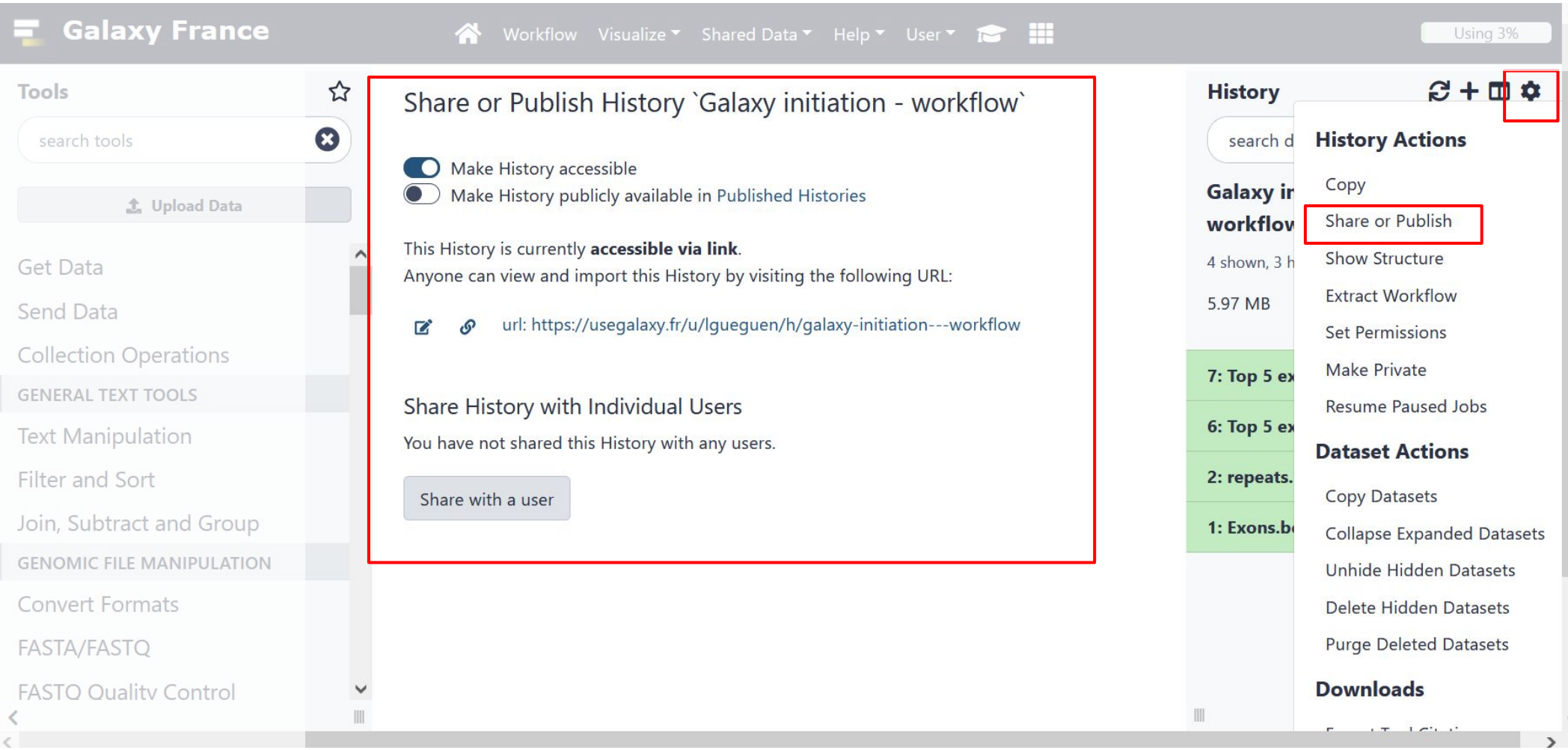
bioanalyst ↔ biologist

- Sharing workflows
  - Pre-configured parameters
  - With or without release parameters (set at runtime)
    - According to the user-end knowledge

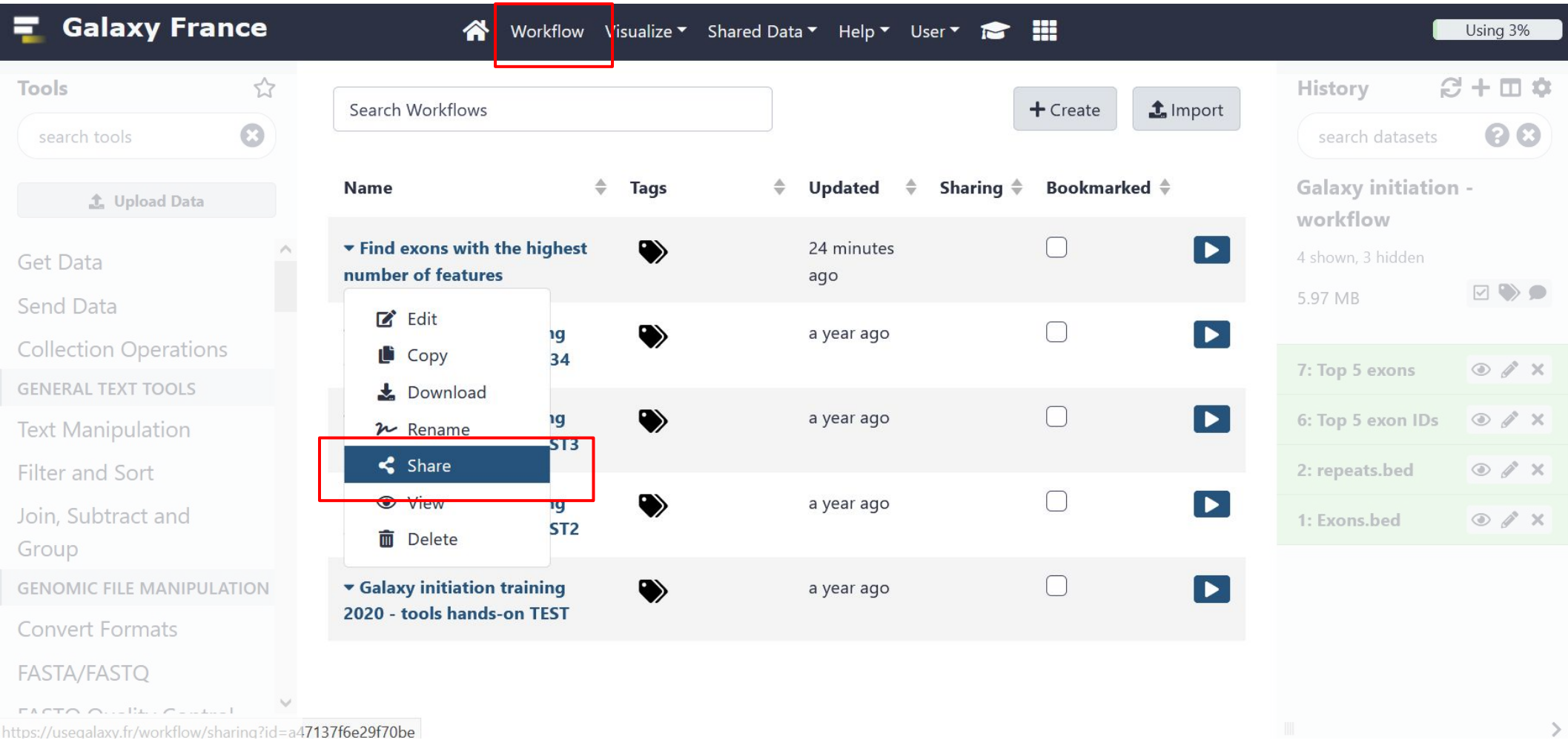
bioinformatician ↔ bioinformatician

- Sharing tools ,scripts and wrappers
  - Toolshed

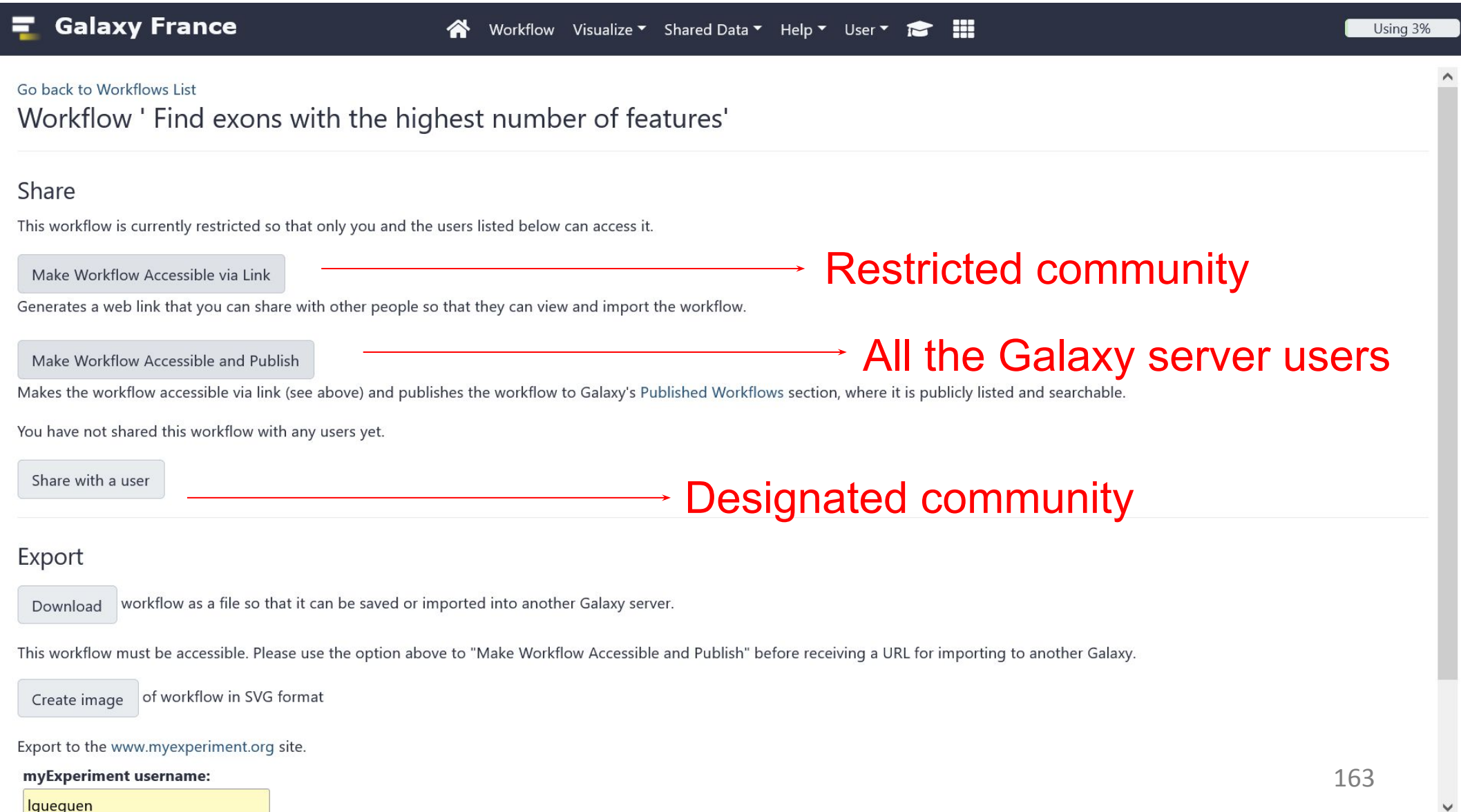
History


Galaxy France
Workflow
Visualize
Shared Data
Help
User
Using 3%
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
Share or Publish History `Galaxy initiation - workflow`
Make History accessible
Make History publicly available in Published Histories
This History is currently accessible via link.
Anyone can view and import this History by visiting the following URL:
url: https://usegalaxy.fr/u/lgueguen/h/galaxy-initiation---workflow
Share History with Individual Users
You have not shared this History with any users.
Share with a user
History
search d
Galaxy ir
workflow
4 shown, 3 h
5.97 MB
7: Top 5 ex
6: Top 5 ex
2: repeats.
1: Exons.b
History Actions
Copy
Share or Publish
Show Structure
Extract Workflow
Set Permissions
Make Private
Resume Paused Jobs
Dataset Actions
Copy Datasets
Collapse Expanded Datasets
Unhide Hidden Datasets
Delete Hidden Datasets
Purge Deleted Datasets
Downloads

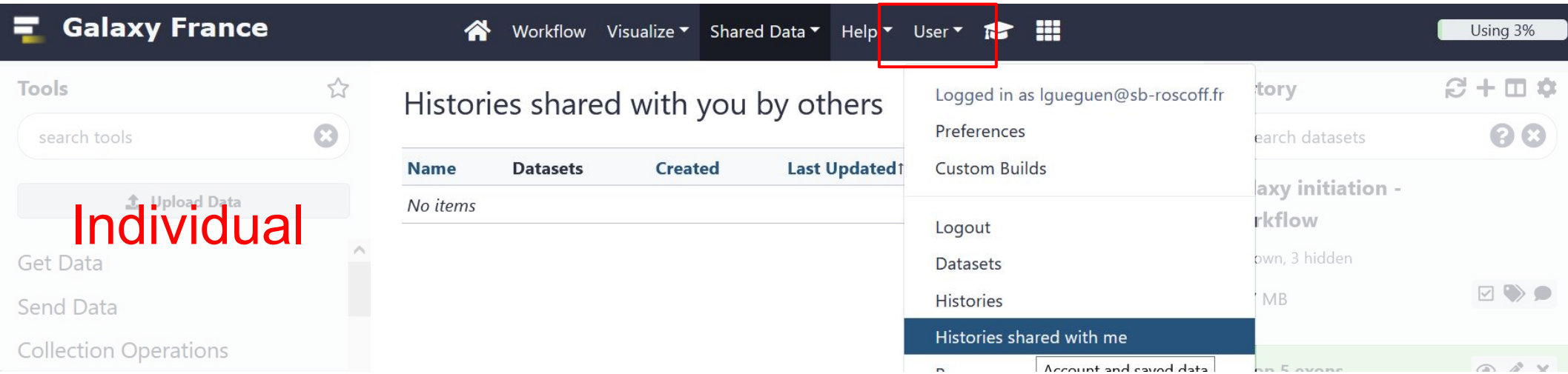
Workflow

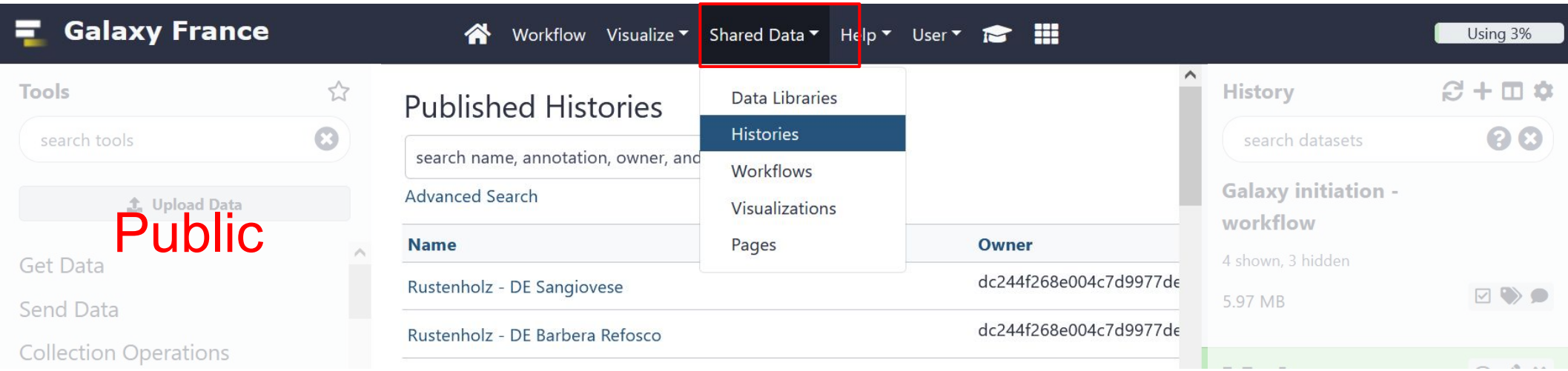

Galaxy France
Workflow
Visualize
Shared Data
Help
User
Using 3%
Tools
search tools
Upload Data
Get Data
Send Data
Collection Operations
GENERAL TEXT TOOLS
Text Manipulation
Filter and Sort
Join, Subtract and Group
GENOMIC FILE MANIPULATION
Convert Formats
FASTA/FASTQ
Search Workflows
+ Create
Import
Name
Tags
Updated
Sharing
Bookmarked
Find exons with the highest number of features
24 minutes ago
Edit
Copy
Download
Rename
Share
View
Delete
a year ago
a year ago
a year ago
Galaxy initiation training 2020 - tools hands-on TEST
a year ago
History
search datasets
Galaxy initiation - workflow
4 shown, 3 hidden
5.97 MB
7: Top 5 exons
6: Top 5 exon IDs
2: repeats.bed
1: Exons.bed
https://useqalaxy.fr/workflow/sharing?id=a47137f6e29f70be

Workflow: mode



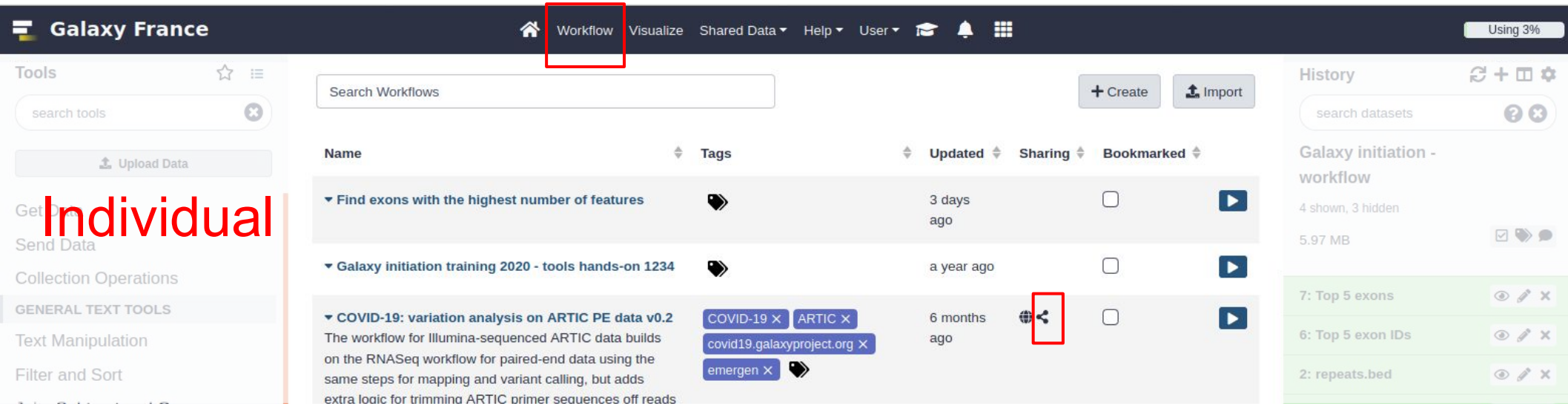
- Get shared histories

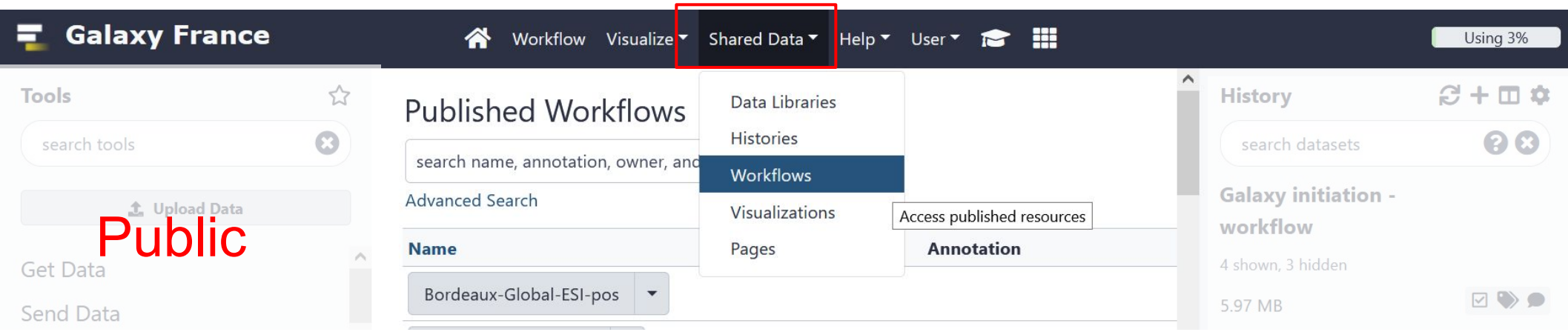




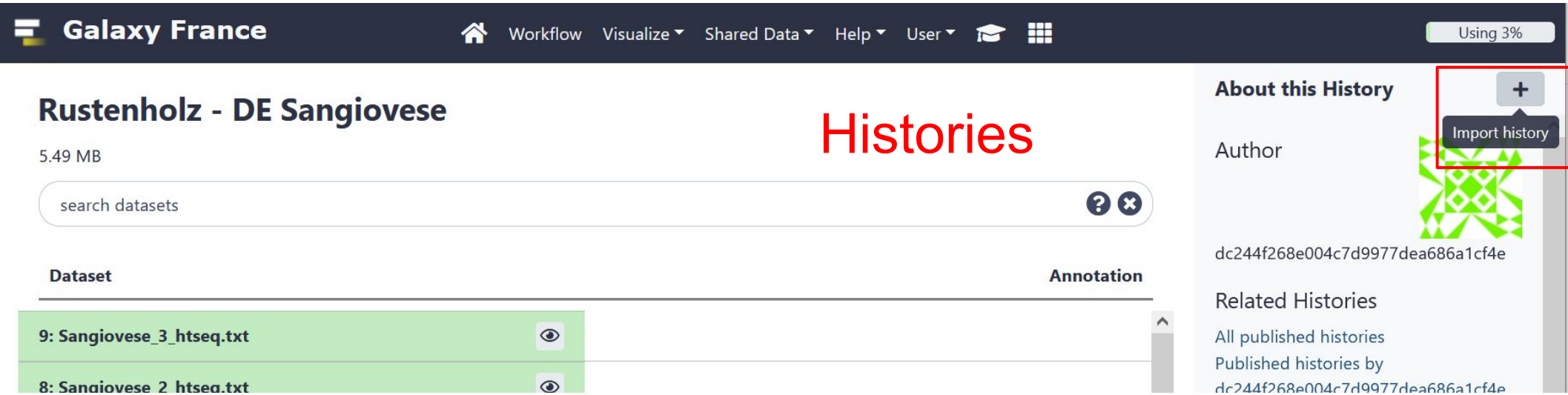
Share

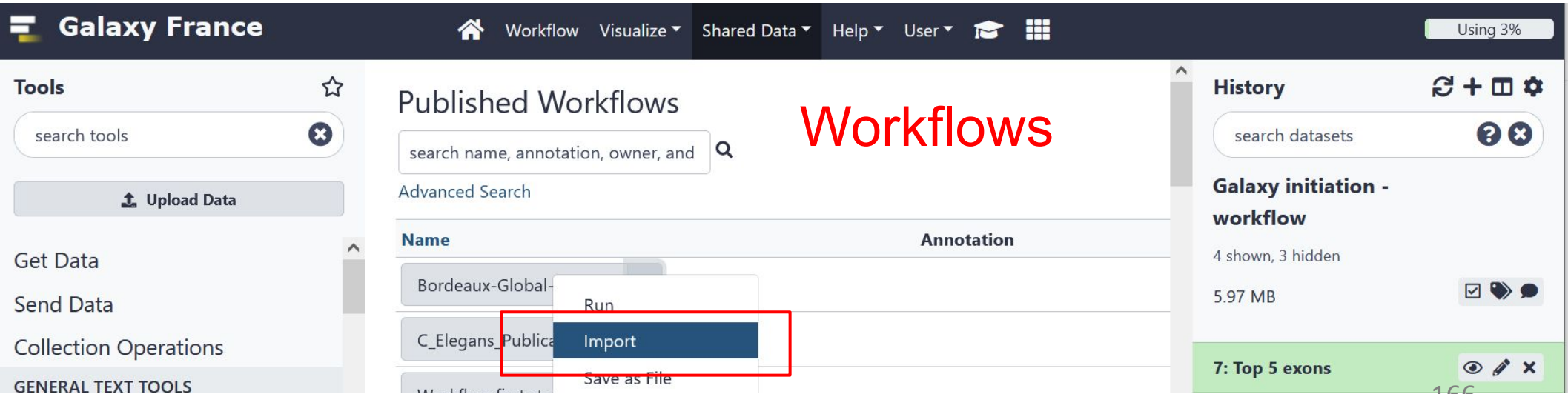
- Get shared workflows





- Import shared





166

Share

Level 5

- Share tools and descriptions in the ToolShed

Level 4

- Launch tools autonomously
- Use advanced parameters
- Provide workflow for colleagues Level 1-3

Level 3

- Launch tools autonomously
- Use workflow more or less preset

Level 2

- Use preset workflow

Level 1

- Share his data to colleagues Level 2-5

Hands-on

WORKFLOW AND SHARE

Galaxy 101

Authors: Saskia Hiltemann Nicola Soranzo Clemens Blank Anton Nekrutenko Björn Grüning Anne Pajon Helena Rasche Add Contributions!

Overview

Questions:

- Which coding exon has the highest number of single nucleotide polymorphisms (SNPs) on human chromosome 22?

Objectives:

- Familiarize yourself with the basics of Galaxy
- Learn how to obtain data from external sources
- Learn how to run tools
- Learn how histories work
- Learn how to create a workflow
- Learn how to share your work

Time estimation: 1 hour

Level: Introductory

Supporting Materials:

Topic Overview slides Datasets Workflows GTN Video Library Available on these Galaxies

Last modification: Nov 17, 2021

Introduction

This tutorial aims to familiarize you with the Galaxy user interface. It will teach you how to perform basic tasks such as importing data, running tools, working with histories, creating workflows, and sharing your work.

Results may vary

Your results may be slightly different from the ones presented in this tutorial due to differing versions of tools, reference data, external databases, or because of stochastic processes in the algorithms.

Agenda

https://training.galaxyproject.org/training-material/topics/introduction/tutorials/galaxy-intro-101/tutorial.html

CONCLUSION

Key points on Galaxy

- Easy-to-use graphical user interface for often complex command-line tools
- Keeps a full record of your analysis in a history
- Workflows enable you to repeat your analysis on different data
- Galaxy can connect to external sources for data import and visualization purposes
- Galaxy provides ways to share your results and methods with others

Best practices

- Regularly free space with "Purge Deleted Datasets "or" Delete Permanently "
- On usegalaxy.fr, if you need support (issue, request for a tool…), please open a subject on https://community.france-bioinformatique.fr/c/galaxy/8

Now, choose your favorite Galaxy!

130+ platforms for using Galaxy: https://galaxyproject.org/use
[Servers, clouds, deployable resources]

GENERAL PURPOSE:

usegalaxy.fr, usegalaxy.org, usegalaxy.eu, usegalaxy.org.au

DOMAIN SPECIFIC:

Metabarcoding:

Metabolomics:

Proteomics:

SBR-tools:

ChIP-seq:

Catalog of French Galaxy tools:
http://galaxycat.france-bioinformatique.fr

To go further

- On your own:
  - Training materials:

    https://galaxyproject.github.io/training-material
  - Interactive tours of Galaxy:

    https://usegalaxy.fr/tours

- Training courses:

Training	What ?	Where ?	When ?
RNAseq analysis with Galaxy	RNAseq	Roscoff, France	?
Galaxy Community Conference (GCC)	General purpose (data-intensive biology and Galaxy)	Minneapolis, USA	July 2022
Workflow4Experimenters	Metabolomics	Europe	2022 (closed), 2023

Please complete the evaluation questionnaire!

END