



JAN 04, 2024

OPEN ACCESS



DOI:
dx.doi.org/10.17504/protocols.io.261ged1eov47/v1

Protocol Citation: Jayanthi Gangiredla, Mark Mammel, Zhihui Yang 2024. Norovirus genotyping and phylogeny analysis_ViroTrakr workflow 1_v.1. **protocols.io**
<https://dx.doi.org/10.17504/protocols.io.261ged1eov47/v1>

License: This is an open access protocol distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited

Protocol status: Working
 We use this protocol and it's working

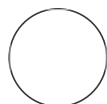
Created: Jan 04, 2024

Norovirus genotyping and phylogeny analysis_ViroTrakr workflow 1_v.1

Mark

Jayanthi Gangiredla¹, Mammel¹, Zhihui Yang¹

¹Office of Applied Research and Safety Assessment, Center for Food Safety and Applied Nutrition, U.S. Food and Drug Administration,



Zhihui Yang
 FDA

DISCLAIMER

Please note that this protocol is public domain, which supersedes the CC-BY license default used by protocols.io.

ABSTRACT

This workflow provides step-by-step instructions for norovirus analysis within the GalaxyTrakr platform. It includes the quality assessment for raw sequencing data (from most next-generation sequencing platforms), drafting de novo assemblies, and reporting the sequence genotype and phylogenetic results. This workflow was designed for norovirus, which is one of the major targets of our ViroTrakr database.

This protocol covers how to:

- Set up an account in Galaxy Trakr;
- Create a new history/workspace for a new submission;
- Upload raw data obtained from local folders or download from NCBI;
- Execute the ViroTrakr workflow 1 (norovirus);
- Interpret the results.

ViroTrakr:

[foodborne viruses \(ID 396739\) - BioProject - NCBI \(nih.gov\)](#)

Reference: Quality control assessment for microbial genomes: GalaxyTrakr MicroRunQC workflow V.5:

[Quality control assessment for microbial genomes: GalaxyTrakr MicroRunQC workflow \(protocols.io\)](#)

PROTOCOL integer ID:
92954

1 Log into your GalaxyTrakr account.

1.1 Create a GalaxyTrakr account if you are the first-time user:
[User Registration Form - Galaxy Genome Trakr \(galaxytrakr.org\)](#)

User Registration Form

Location

FDA MOD1

Add New Location

First Name

Enter First Name, Do not use characters: \[\]{};|'="+~*?<>@.

Last Name

Enter First Name, Do not use characters: \[\]{};|'="+~*?<>@.

Email

Email will be used for automated messages to include registration information!

Primary Phone

Please enter number with country code, without dashes, for example +17035456789

If possible please use a mobile number than can accept text messages, only used for support

Title

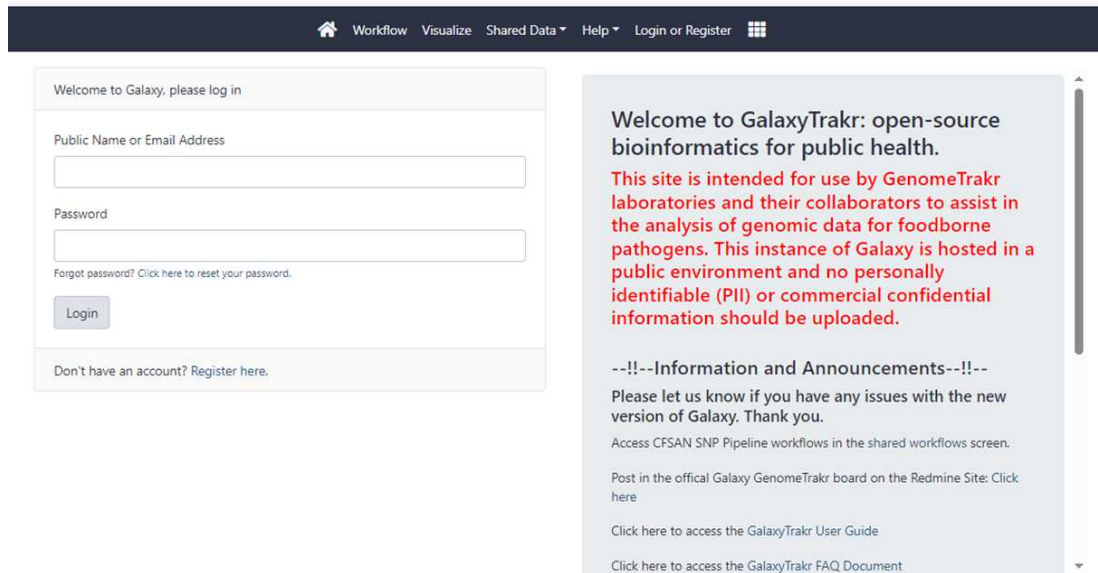
Requirements

Please annotate intended use of Galaxy and Analysis tools. List specific tools you would like to see deployed in Galaxy.

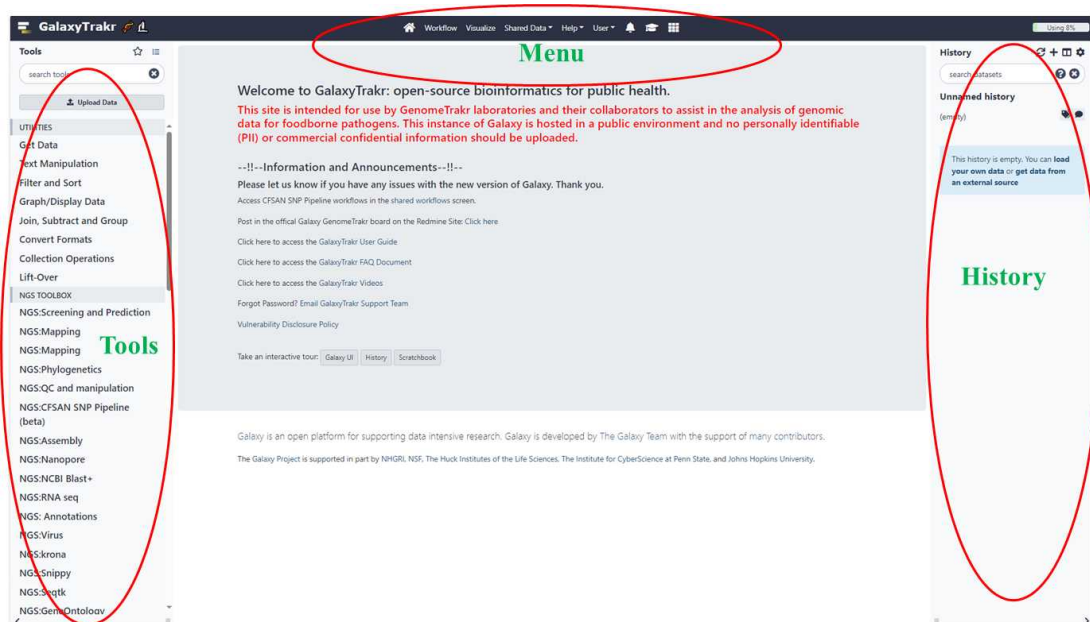
Register

© 2023 - Galaxy Genome Trakr Organization

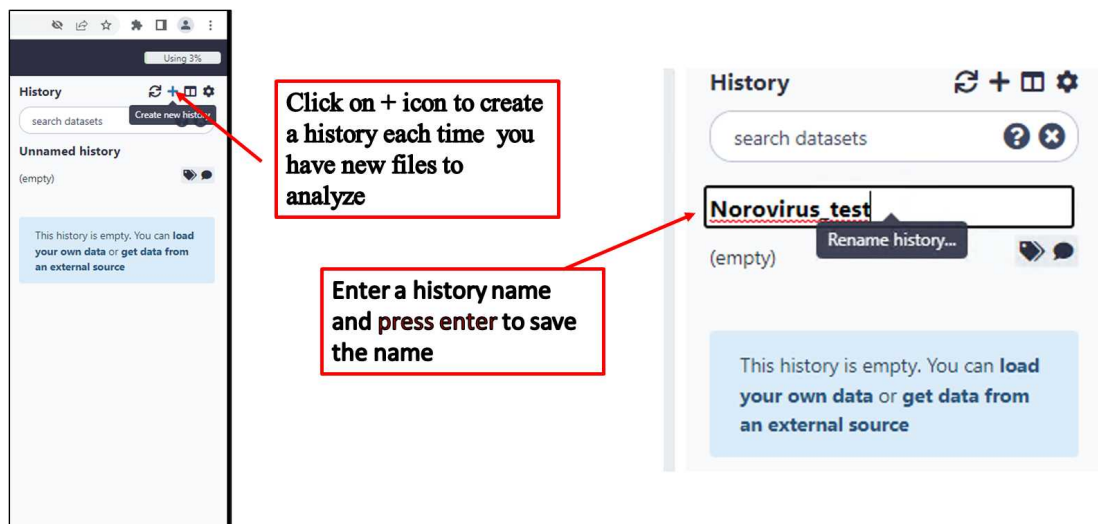
1.2 Log into your GalaxyTrakr account if you already have one:
[Galaxy \(galaxytrakr.org\)](#)



1.3 Get familiar with Galaxy components: Tools, Menu and History.



2 Create a new history.



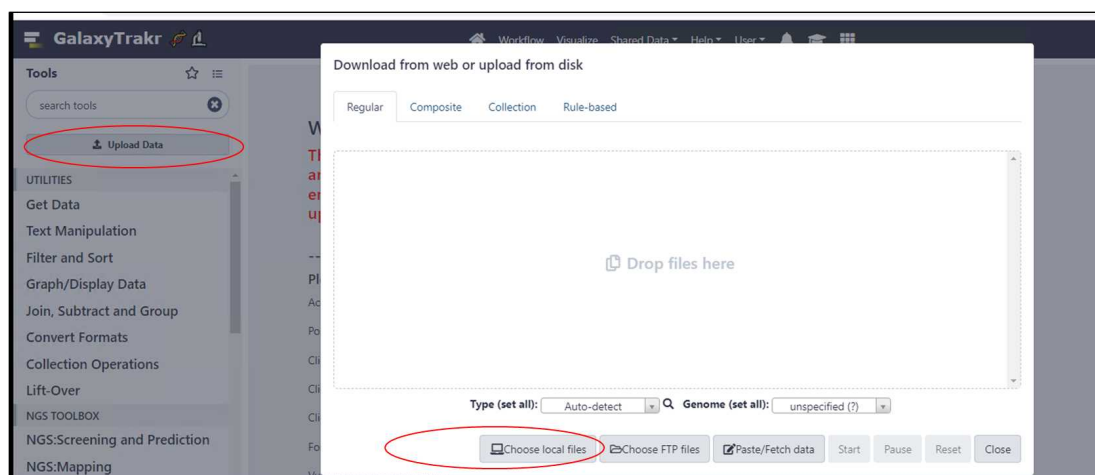
3 Upload data.

The raw sequencing data in fastq files can be imported into GalaxyTrakr directly from your local folder (instructions shown in 3.1); or downloaded from SRA (instructions shown as in 3.2) if the files have been already submitted to ViroTrakr in NCBI (Submission protocol: [NCBI submission protocol for foodborne virus surveillance \(protocols.io\)](https://www.protocols.io/submission-protocol-for-foodborne-virus-surveillance)).

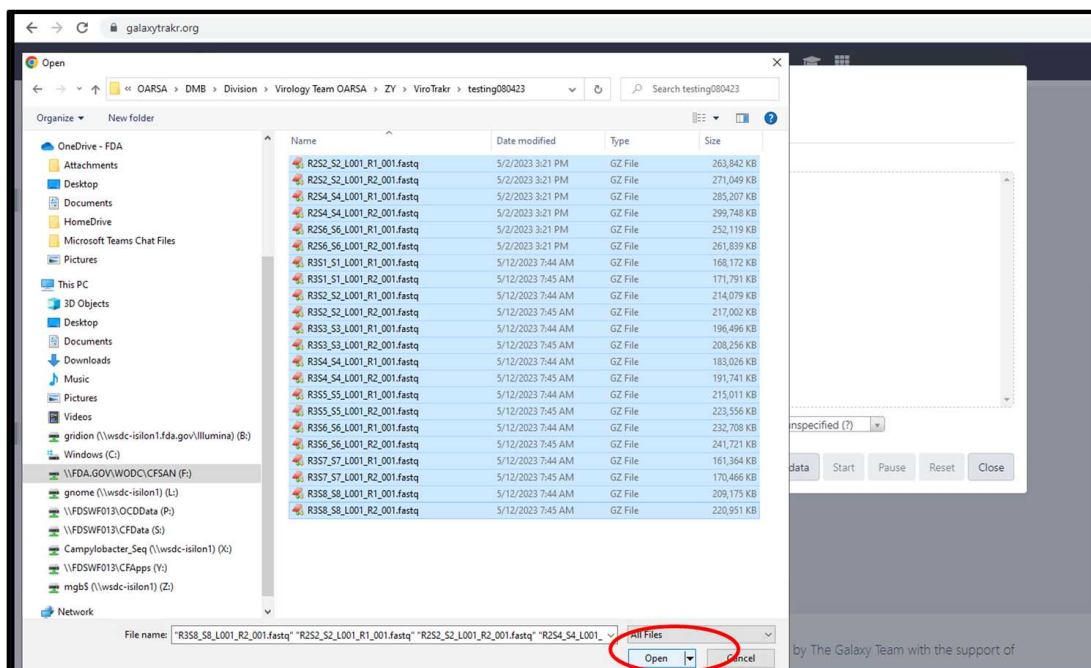
After being uploaded to GalaxyTrakr, the files will remain in your account until they are deleted.

3.1 Upload raw data from local folder.

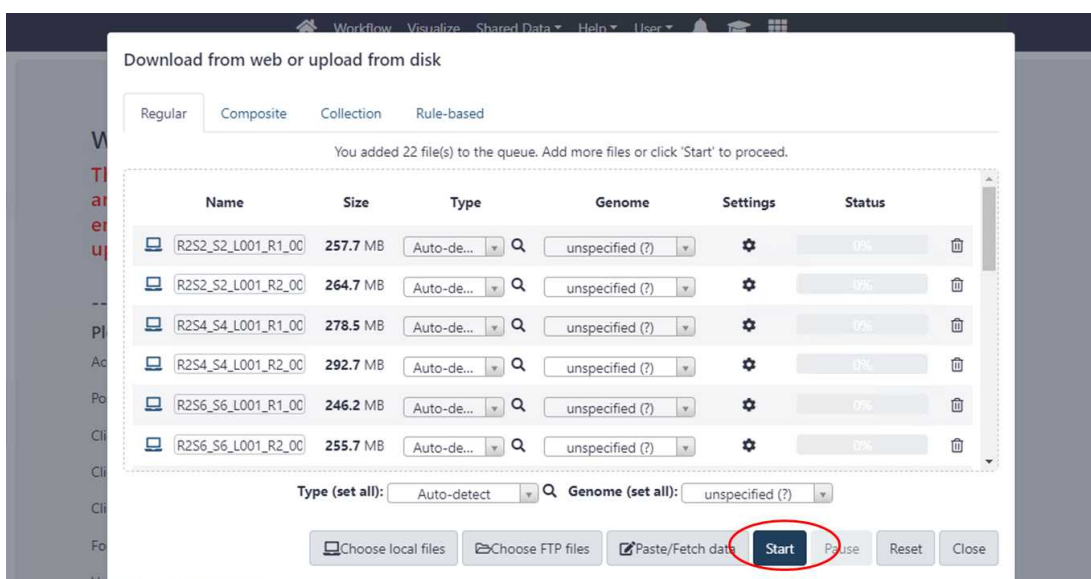
3.1.1. Click on the button "Upload Data", then "Choose local files".



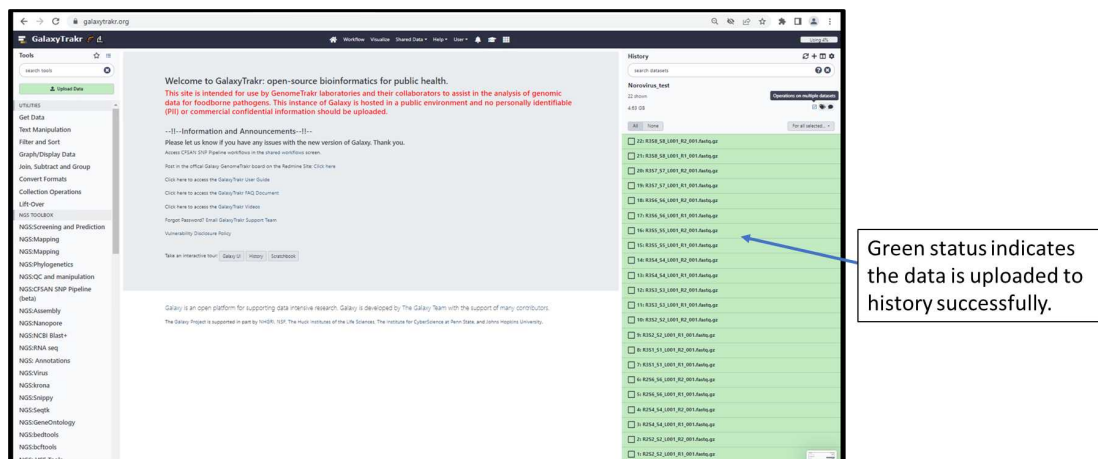
3.1.2. Select fastq files from your local folder.



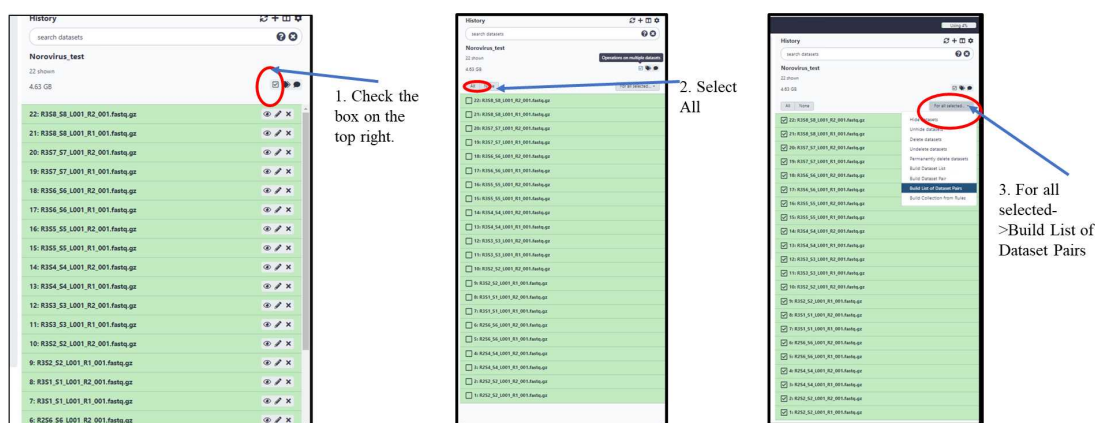
3.1.3. Select the files and click “Start” to upload.



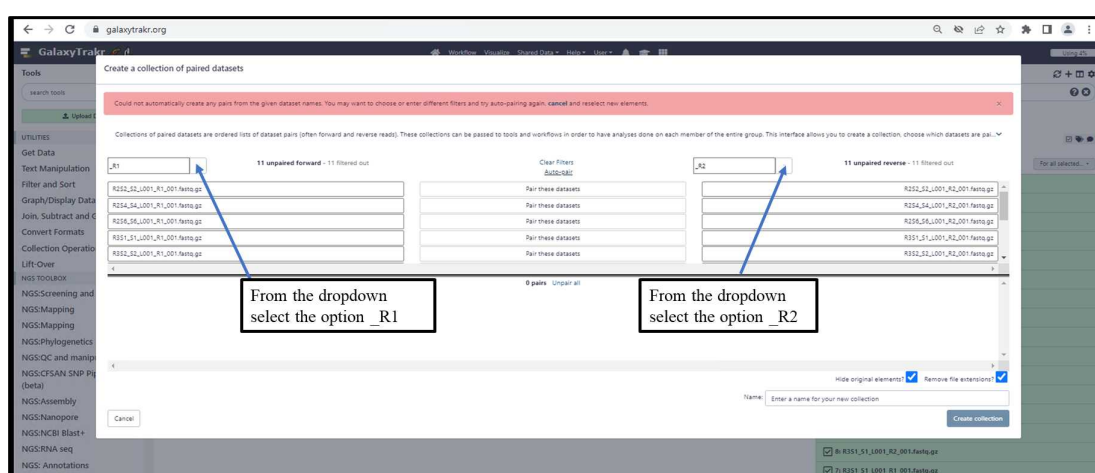
3.1.4. Check the status of data upload.



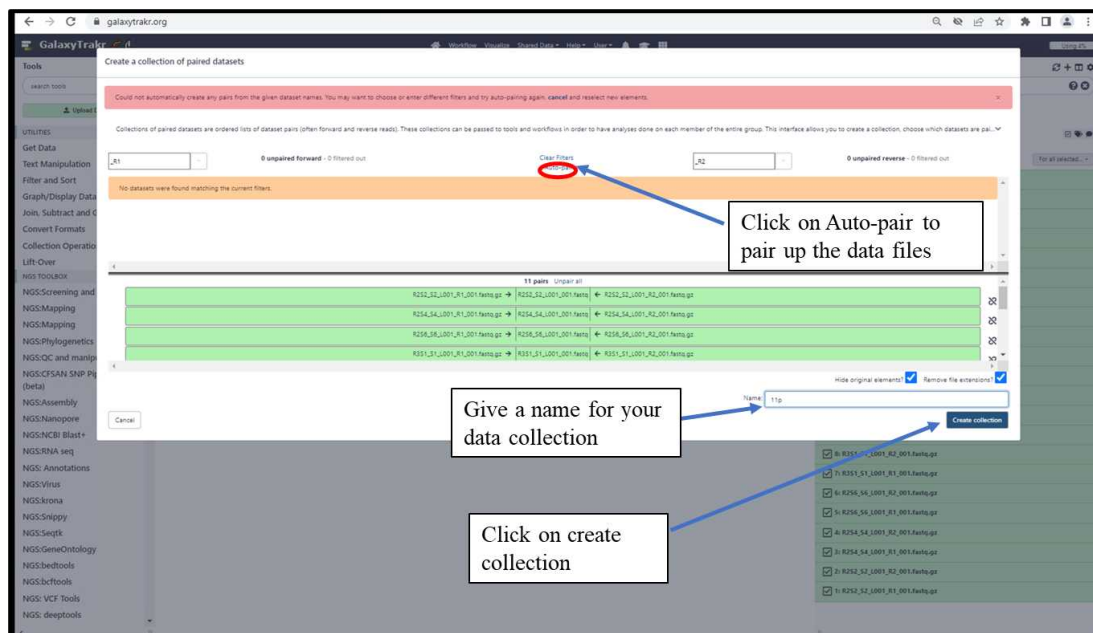
3.1.5. Build a list of Dataset Pairs (pairing the forward and reverse files into their respective samples for batch analysis (Follow steps1, 2 and 3).



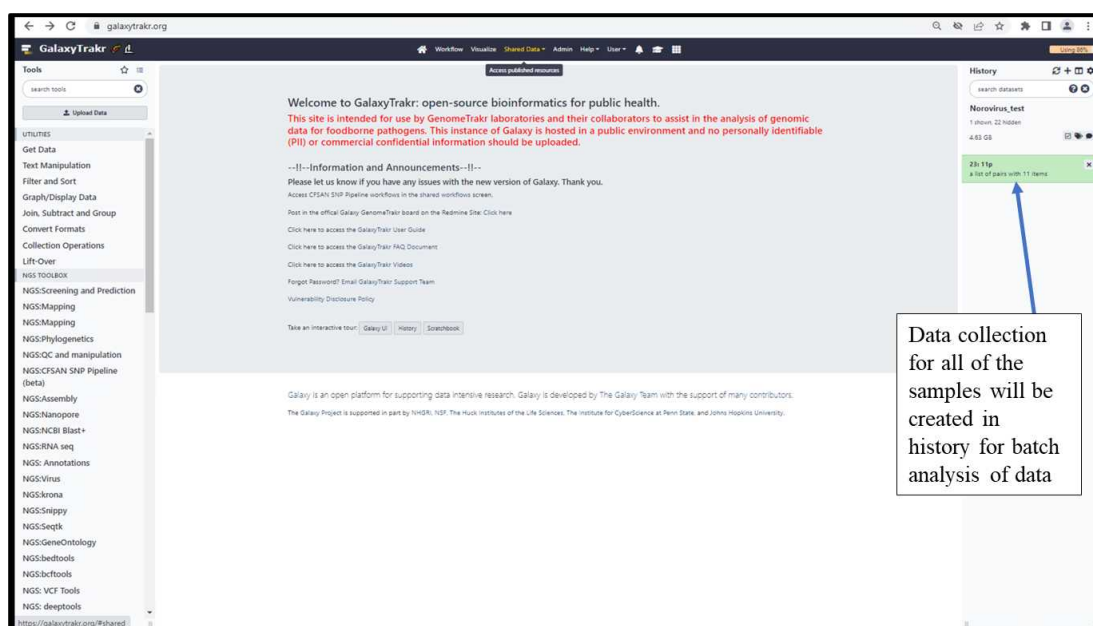
3.1.6. Create a collection of paired datasets.



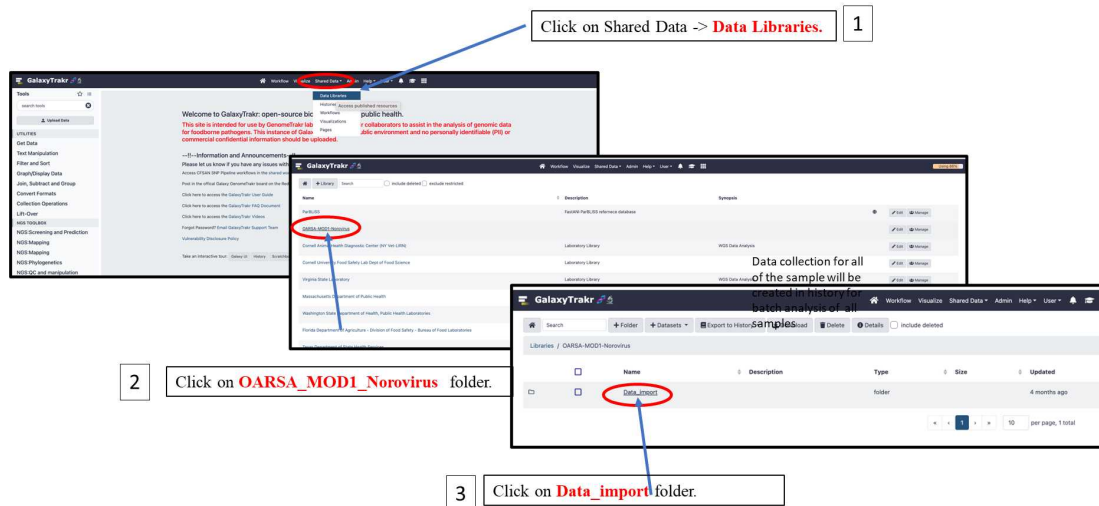
3.1.6. Create a collection of paired datasets - Cont.



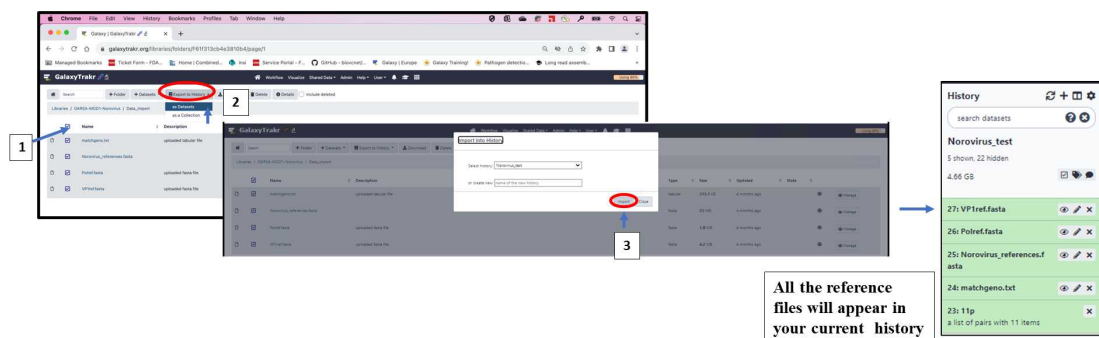
3.1.7. Data collection will be created in history.



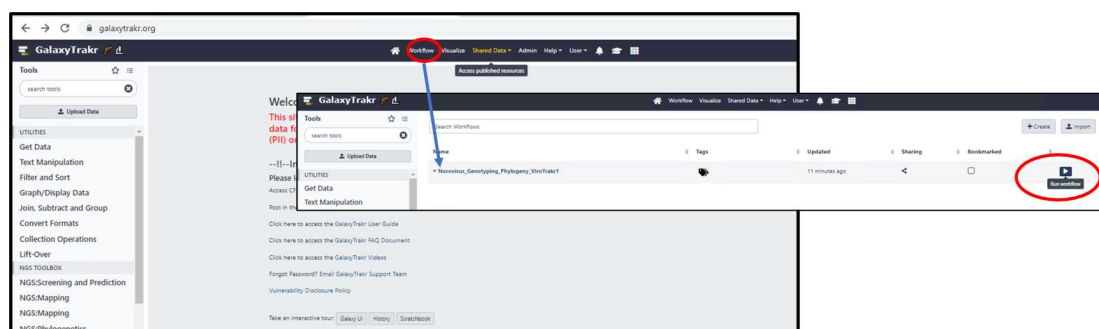
3.1.8. Import the reference data files from Shared Data folder following the steps 1-3 as shown below.



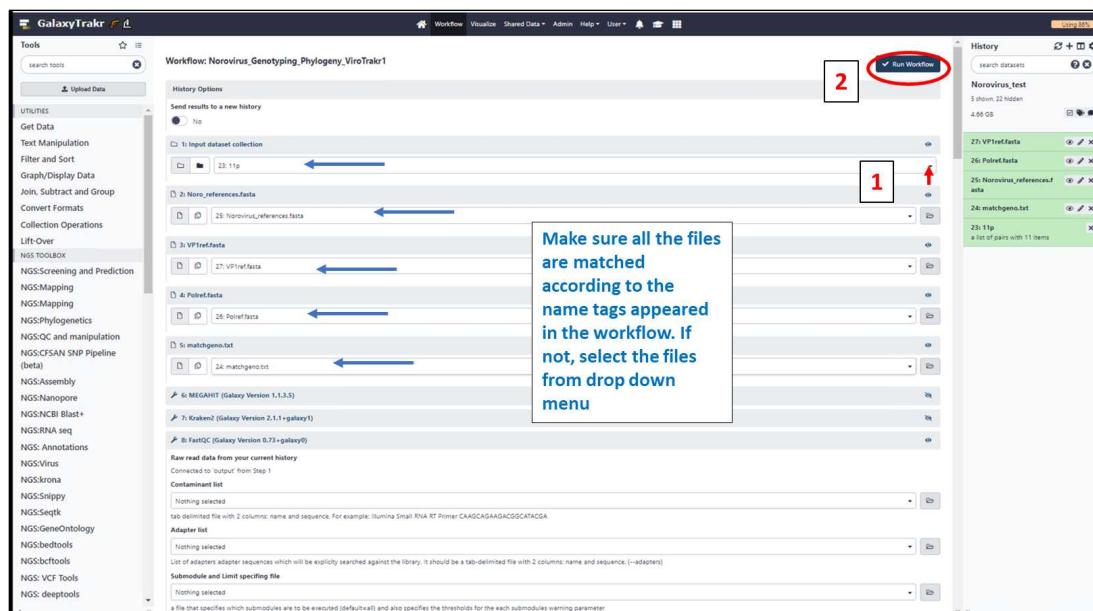
3.1.9. Select all files from the folder and export them as Datasets to your current history following the steps 1-3 as shown below.



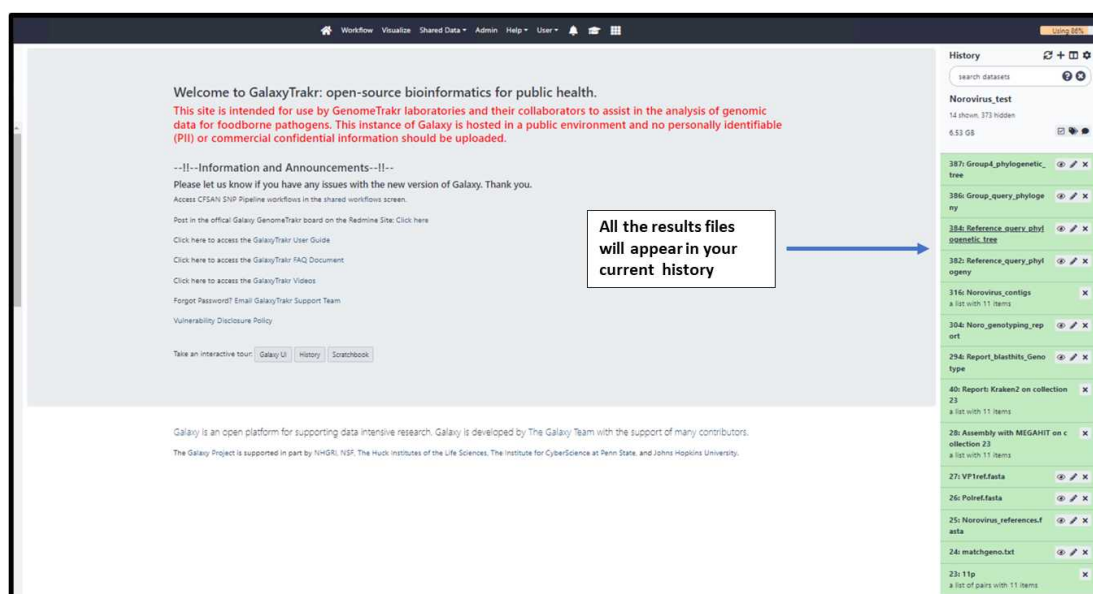
3.1.10. Click on WorkFlow tab from main menu, select and run the Norovirus_ Genotyping_Phylogeny workflow.



3.1.11. Select all the appropriate files from each dropdown menu and run workflow.



3.1.12. Once the workflow run is successful (Green status), results will appear in the history.



3.1.13. Select and view the result files in the middle panel.



3.1.14. Download the result files to your local folder.



Result files include:

- Assembly with MEGAHIT: Metagenomic assemblies
- Report: Kraken2: Kraken2 reports
- Report_blasthits_Genotype: Reporting Best BLAST Hits against reference sequences
- Noro_genotyping_report: Final report contains QC stats and genotyping results
- Norovirus_contigs: Norovirus specific contigs extracted from metagenomics assembly
- Reference_query_phylogenetic_tree: Phylogenetic tree represents the input genomes along with the reference genomes from all groups .png format and .txt format.
- Group4_phylogenetic_tree: Phylogenetic tree represents the input genomes along with the GroupII.4 reference genomes from all groups .png format and .txt format.

3.2 Download data from SRA database (if you submit your raw data to ViroTrakr before GalaxyTrakr analysis).

3.2.1. SRA database link to norovirus sequence files for bioprojectPRJNA490509:

- https://www.ncbi.nlm.nih.gov/Traces/study/?acc=SRP173043&o=acc_s%3Aa

3.2.2. From SRA database, (1) select SRR samples of your choice then (2) click on accession list.

The screenshot shows the SRA database interface. At the top, the 'Accession' field is set to 'SRP173043'. Below this, a 'Common Fields' section displays metadata for the selected study. The 'Select' section shows a table with columns: 'Runs', 'Bytes', 'Bases', 'Download', 'Accession List', 'Cloud Data Delivery', and 'Computing'. The 'Accession List' button is highlighted with a red circle and a red arrow labeled '2'. Below the 'Select' section, a table of 15 items is displayed, with the first item (SRR14119290) highlighted with a red circle and a red arrow labeled '1'.

| Run | BioSample | Assay Type | AvgSpotLen | Bases | BioSampleModel | Bytes | Center Name | collection_date | Experiment | geo_loc_name | Instrument | Isolation_source | Library Name | Library Layout | |
|-----|-------------|--------------|------------|-------|----------------|--------------------|-------------|---------------------------------|------------|--------------|---------------|------------------|--------------|----------------|--------|
| 1 | SRR14119290 | SAMN1857787 | WGSS | 1,141 | 66.22 k | Pathogen.cl | 138.05 kb | OARSA/CFSAN/FDA | 2016-02 | SRX10489631 | USA: Maryland | GridION | stool | G251 | SINGLE |
| 2 | SRR24719432 | SAMN35345295 | RNA-Seq | 196 | 16.50 M | One Health Enteric | 10.61 Mb | FDA, CFSAN, OARSA | 2014-02 | SRX20497484 | USA: Maryland | Illumina MiSeq | stool | E1653 | PAIRED |
| 3 | SRR24745610 | SAMN35429855 | WGSS | 218 | 691.34 M | One Health Enteric | 383.27 Mb | FDA, CFSAN, OARSA | 2017-04-06 | SRX20522747 | USA: CA | Illumina MiSeq | clinic | R258 | PAIRED |
| 4 | SRR26123209 | SAMN37480926 | WGSS | 182 | 447.47 M | One Health Enteric | 282.29 Mb | US FOOD AND DRUG ADMINISTRATION | 2017-03-17 | SRX21837013 | USA: CA | Illumina MiSeq | stool | R358 | PAIRED |
| 5 | SRR26123210 | SAMN37480925 | WGSS | 181 | 486.85 M | One Health Enteric | 307.85 Mb | US FOOD AND DRUG ADMINISTRATION | 2017-04-06 | SRX21837012 | USA: CA | Illumina MiSeq | stool | R356 | PAIRED |
| 6 | SRR26123211 | SAMN37480924 | WGSS | 181 | 395.36 M | One Health Enteric | 248.16 Mb | US FOOD AND DRUG ADMINISTRATION | 2017-03 | SRX21837011 | USA: CA | Illumina MiSeq | stool | R354 | PAIRED |
| 7 | SRR26123212 | SAMN37480923 | WGSS | 181 | 452.89 M | One Health Enteric | 278.13 Mb | US FOOD AND DRUG ADMINISTRATION | 2017-03-17 | SRX21837010 | USA: CA | Illumina MiSeq | stool | R352 | PAIRED |
| 8 | SRR26123214 | SAMN37480921 | WGSS | 253 | 538.79 M | One Health Enteric | 291.85 Mb | US FOOD AND DRUG ADMINISTRATION | 2014-01-02 | SRX21837008 | USA: CA | Illumina MiSeq | stool | R255 | PAIRED |
| 9 | SRR26123215 | SAMN37480920 | WGSS | 216 | 708.76 M | One Health Enteric | 401.13 Mb | US FOOD AND DRUG ADMINISTRATION | 2018-03-06 | SRX21837007 | USA: CA | Illumina MiSeq | stool | R254 | PAIRED |
| 10 | SRR26123216 | SAMN37480919 | WGSS | 255 | 602.05 M | One Health Enteric | 327.92 Mb | US FOOD AND DRUG ADMINISTRATION | 2013-12-31 | SRX21837006 | USA: CA | Illumina MiSeq | stool | R251 | PAIRED |
| 11 | SRR26123217 | SAMN37480918 | WGSS | 251 | 1.26 G | One Health Enteric | 800.42 Mb | US FOOD AND DRUG ADMINISTRATION | 2013-12-31 | SRX21837005 | USA: CA | Illumina MiSeq | stool | R154 | PAIRED |
| 12 | SRR26123218 | SAMN37480917 | WGSS | 257 | 946.81 M | One Health Enteric | 604.14 Mb | US FOOD AND DRUG ADMINISTRATION | 2017-12-29 | SRX21837004 | USA: CA | Illumina MiSeq | stool | R152 | PAIRED |
| 13 | SRR8298615 | SAMN10060452 | WGSS | 192 | 329.73 M | Pathogen.cl | 89.25 Mb | OARSA/CFSAN/FDA | 2014-12 | SRX5113117 | USA | Illumina MiSeq | stool | E2254r | PAIRED |
| 14 | SRR8298616 | SAMN10060453 | WGSS | 195 | 554.36 M | Pathogen.cl | 143.12 Mb | OARSA/CFSAN/FDA | 2013-01-06 | SRX5113116 | USA | Illumina MiSeq | stool | E1956 | PAIRED |
| 15 | SRR8349645 | SAMN10611964 | WGSS | 190 | 917.20 M | Pathogen.cl | 238.88 Mb | OARSA/CFSAN/FDA | 2018-04-18 | SRX5160644 | USA | Illumina MiSeq | stool | E3957r | PAIRED |

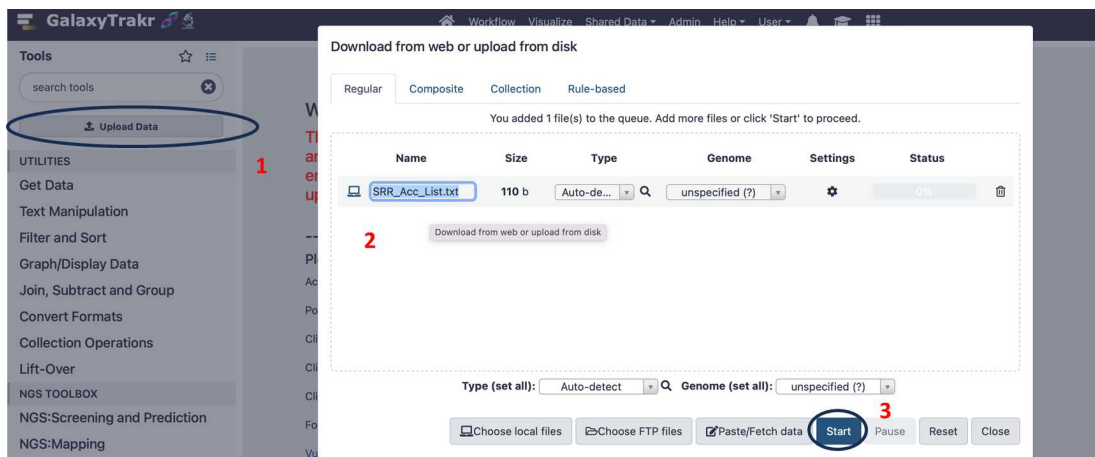
3.2.3.

The text files will be downloaded with SRR accessions in a format as SRR_Acc_List.txt:

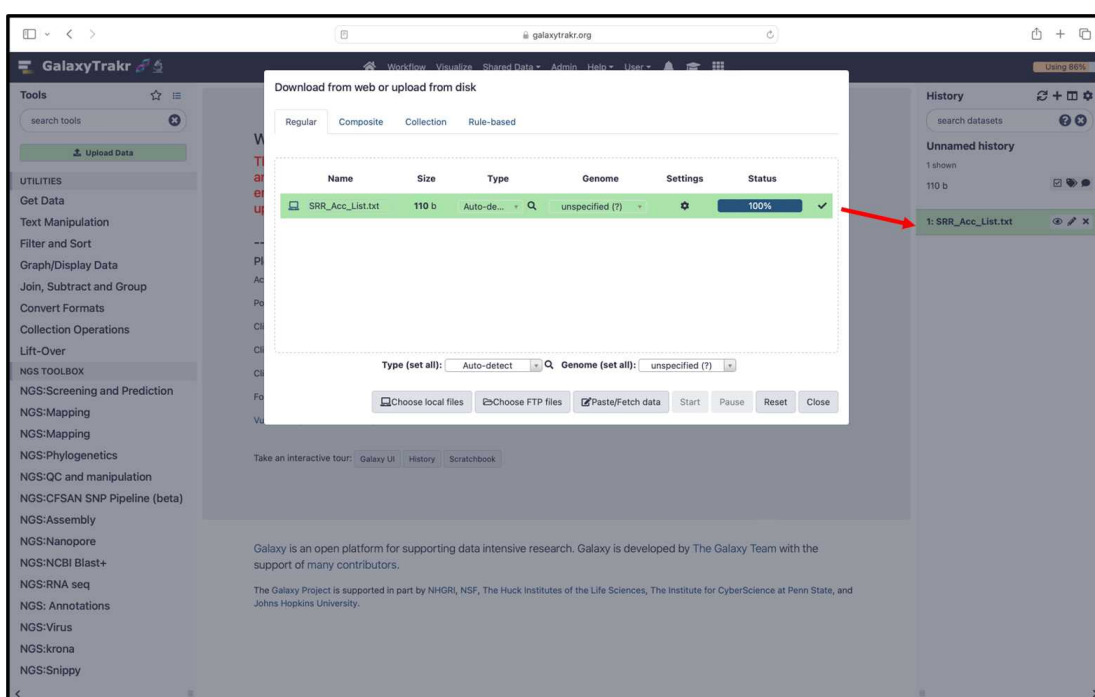
For example:

SRR26123209
SRR26123210
SRR26123211
SRR26123212
SRR26123214
SRR26123215
SRR26123216
SRR26123217
SRR26123218

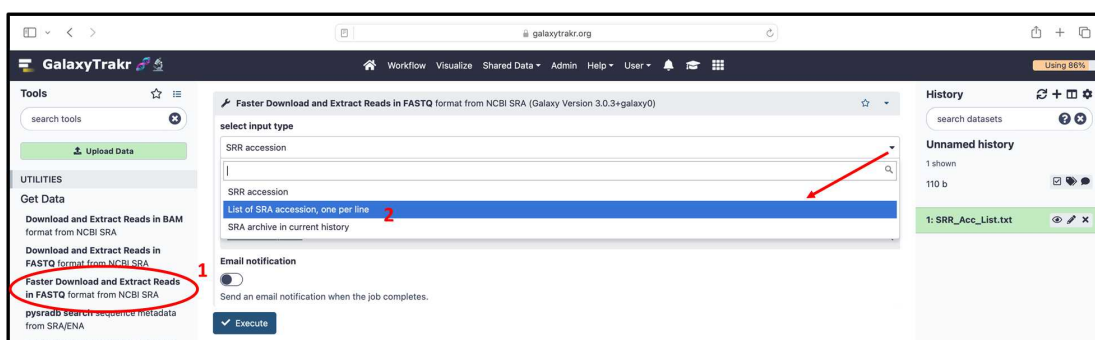
3.2.4. To "Upload" the data, select the SRR_Acc_list.txt and click the "Start" button:



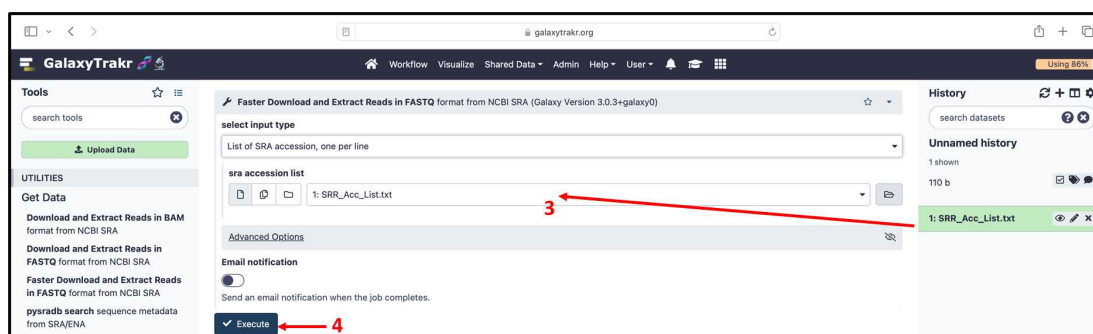
3.2.5. Once the download is completed, the text file will be added in history.



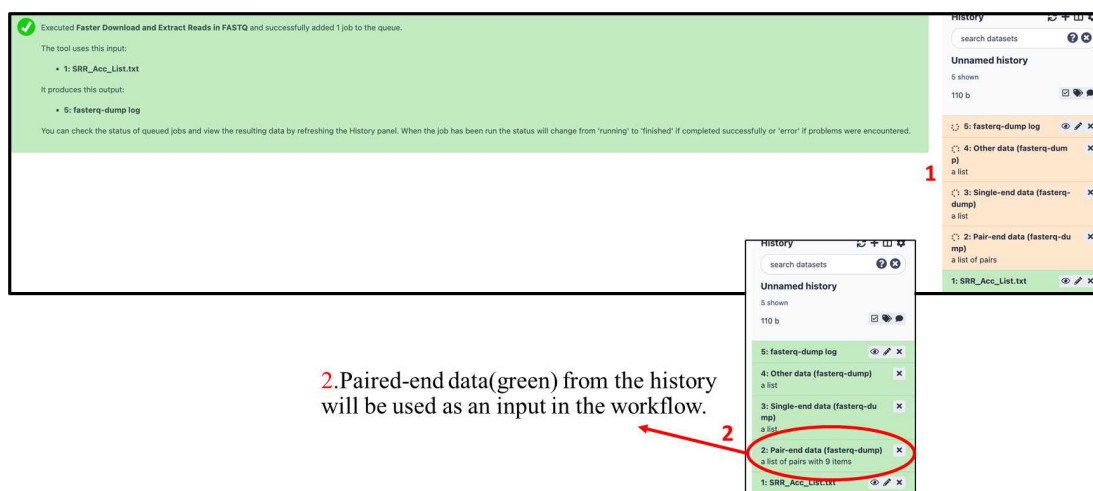
3.2.6. From "Get Data" on the left menu, select "Faster Download and Extract reads in FASTQ from NCBI SRA", and select option by clicking on drop down menu "List of SRA accessions, one per line".



3.2.6. (cont.) Select the file SRR_Acc_List.txt and Click the Execute” button.



3.2.7. Data files will be downloading from NCBI SRA database. (Download time varies, depending on the number of files downloading and the NCBI server status).



3.2.8. Follow the steps from 3.1.8. to 3.1.14 to run the workflow and collect the results.