

Jan 04, 2024

## Norovirus genotyping and phylogeny analysis\_ViroTrakr workflow 1\_v.1

DOI

<https://dx.doi.org/10.17504/protocols.io.261ged1eov47/v1>

Jayanthi Gangiredla<sup>1</sup>, Mark Mammel<sup>1</sup>, Zhihui Yang<sup>1</sup>

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

GenomeTrakr

ViroTrakr



Zhihui Yang

FDA

### Create & collaborate more with a free account

Edit and publish protocols, collaborate in communities, share insights through comments, and track progress with run records.

Create free account

OPEN  ACCESS



**DOI:** <https://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:** January 04, 2024

**Last Modified:** January 04, 2024

**Protocol Integer ID:** 92954

**Keywords:** step instructions for norovirus analysis, norovirus analysis, virotrakt database, norovirus, quality control assessment for microbial genome, virotrakt, foodborne virus, galaxytrakt microrunqc workflow, microbial genome, sequencing data, sequencing platform, galaxytrakt platform, virus, phylogenetic result, galaxytrakt, galaxy trakt, sequencing, sequence genotype, microrunqc workflow, microrunqc, account in galaxy trakt, de novo assembly

## 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 GalaxyTrakt 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 ViroTrakt database.

This protocol covers how to:

Set up an account in Galaxy Trakt;

Create a new history/workspace for a new submission;

Upload raw data obtained from local folders or download from NCBI;

Execute the ViroTrakt workflow 1 (norovirus);

Interpret the results.

ViroTrakt:

**foodborne viruses (ID 396739) - BioProject - NCBI (nih.gov)**

Reference: Quality control assessment for microbial genomes: GalaxyTrakt MicroRunQC workflow V.5:  
**Quality control assessment for microbial genomes: GalaxyTrakt MicroRunQC workflow (protocols.io)**



## 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](https://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](https://galaxytrakr.org))**

Welcome to Galaxy. please log in

Public Name or Email Address

Password

Forgot password? Click here to reset your password.

Login

Don't have an account? Register here.

**Welcome to GalaxyTrakr: open-source bioinformatics for public health.**

**This site is intended for use by GenomeTrakr laboratories and their collaborators to assist in the analysis of genomic data for foodborne pathogens. This instance of Galaxy is hosted in a public environment and no personally identifiable (PII) or commercial confidential information should be uploaded.**

--!!--Information and Announcements--!!--

Please let us know if you have any issues with the new version of Galaxy. Thank you.

Access CFSAN SNP Pipeline workflows in the shared workflows screen.

Post in the official Galaxy GenomeTrakr board on the Redmine Site: Click here

Click here to access the GalaxyTrakr User Guide

Click here to access the GalaxyTrakr FAQ Document

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

**GalaxyTrakr**

Workflow Visualize Shared Data Help User

**Tools**

search tools

Upload Data

UTILITIES

Get Data

Text Manipulation

Filter and Sort

Graph/Display Data

Join, Subtract and Group

Convert Formats

Collection Operations

Lift-Over

NGS TOOLBOX

NGS:Screening and Prediction

NGS:Mapping

NGS:Mapping

NGS:Phylogenetics

NGS:QC and manipulation

NGS:CFSAN SNP Pipeline (beta)

NGS:Assembly

NGS:Nanopore

NGS:NCBI Blast+

NGS:RNA seq

NGS: Annotations

NGS:Virus

NGS:Skrona

NGS:Snippy

NGS:Sniff

NGS:Genome Ontology

**Menu**

Welcome to GalaxyTrakr: open-source bioinformatics for public health.

**This site is intended for use by GenomeTrakr laboratories and their collaborators to assist in the analysis of genomic data for foodborne pathogens. This instance of Galaxy is hosted in a public environment and no personally identifiable (PII) or commercial confidential information should be uploaded.**

--!!--Information and Announcements--!!--

Please let us know if you have any issues with the new version of Galaxy. Thank you.

Access CFSAN SNP Pipeline workflows in the shared workflows screen.

Post in the official Galaxy GenomeTrakr board on the Redmine Site: Click here

Click here to access the GalaxyTrakr User Guide

Click here to access the GalaxyTrakr FAQ Document

Click here to access the GalaxyTrakr Videos

Forgot Password? Email GalaxyTrakr Support Team

Vulnerability Disclosure Policy

Take an interactive tour: Galaxy UI History Scratchbook

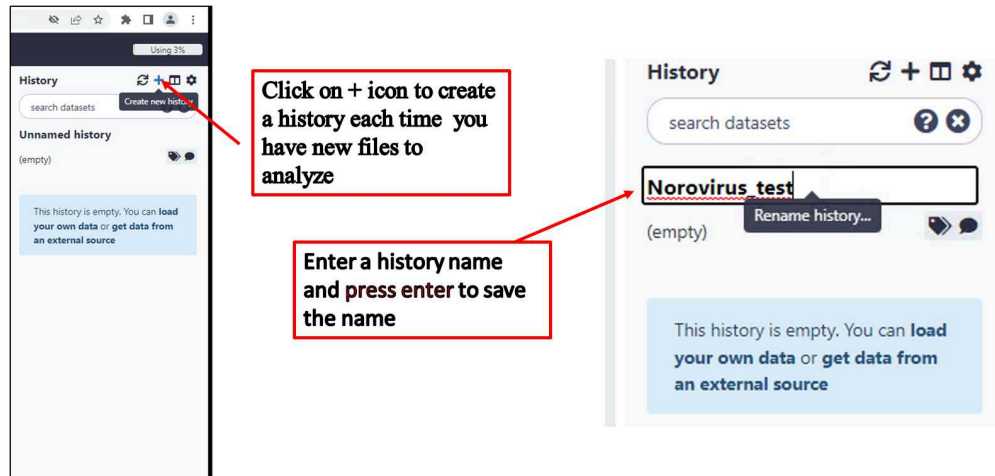
**History**

search history

Unnamed history (empty)

This history is empty. You can load your own data or get data from an external source

### 2 Create a new history.



### 3 Upload data.

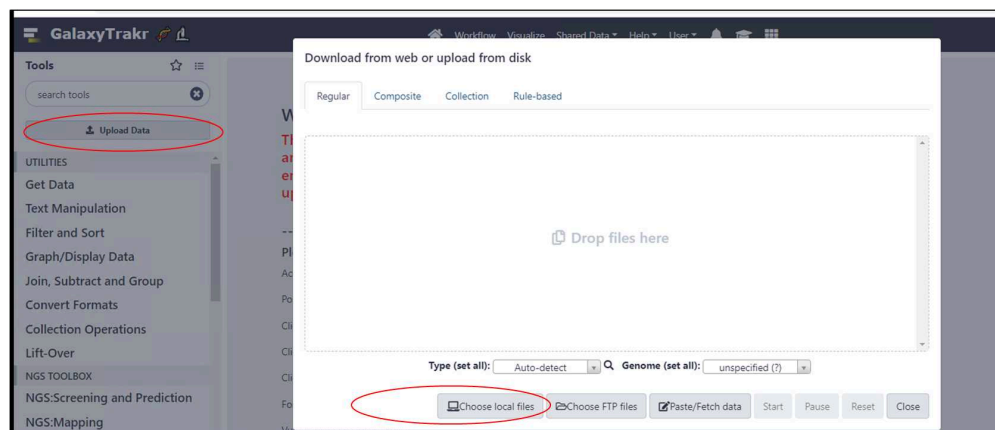
The raw sequencing data in fastq files can be imported into GalaxyTrakr directly from your local folder (instructions shown in 3.1); or downloaded from SRA (instructions shown as in 3.2) if the

files have been already submitted to ViroTrakr in NCBI (Submission protocol: **NCBI submission protocol for foodborne virus surveillance (protocols.io)**).

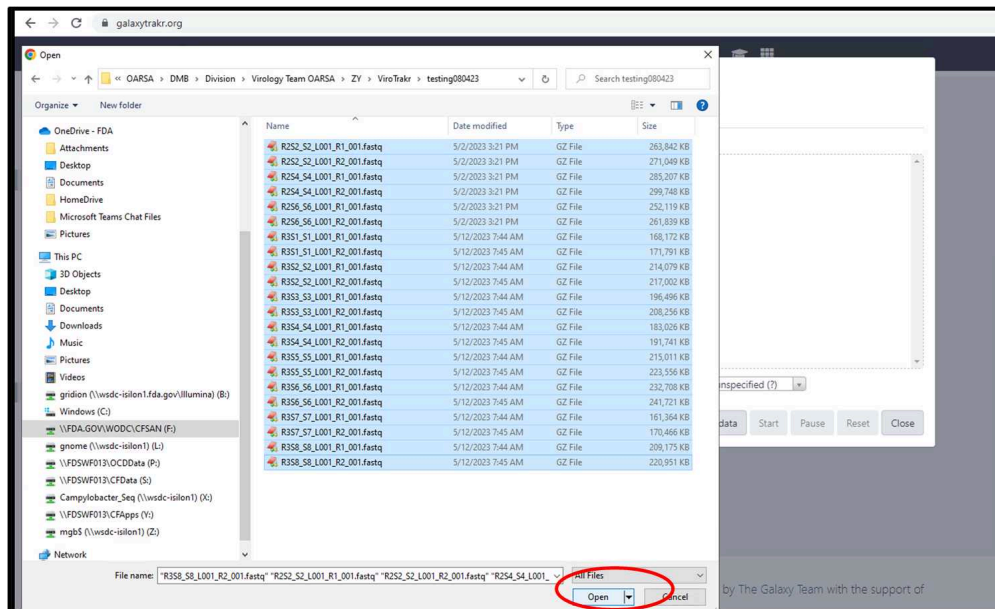
After being uploaded to GalaxyTrakr, the files will remain in your account until they are deleted.

#### 3.1 Upload raw data from local folder.

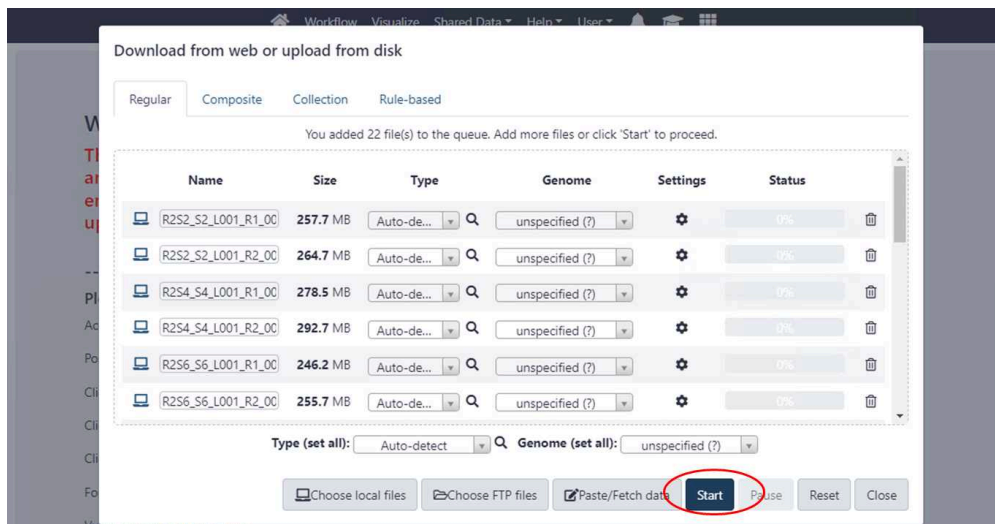
3.1.1. Click on the button "Upload Data", then "Choose local files".



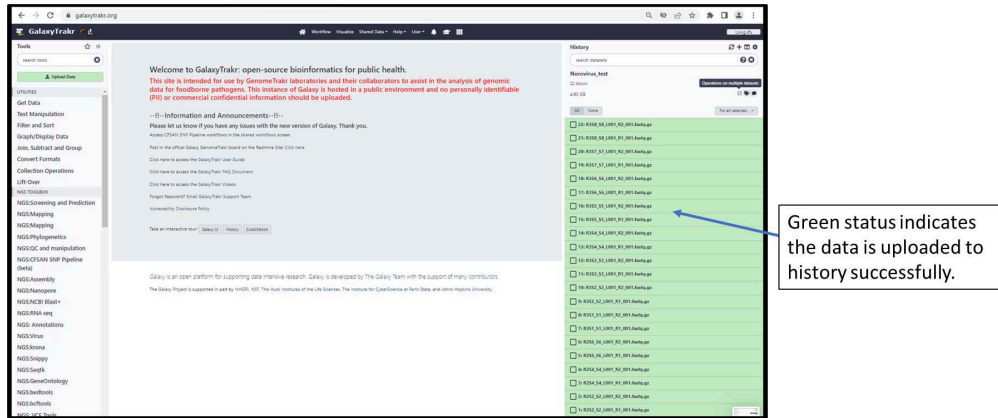
### 3.1.2. Select fastq files from your local folder.



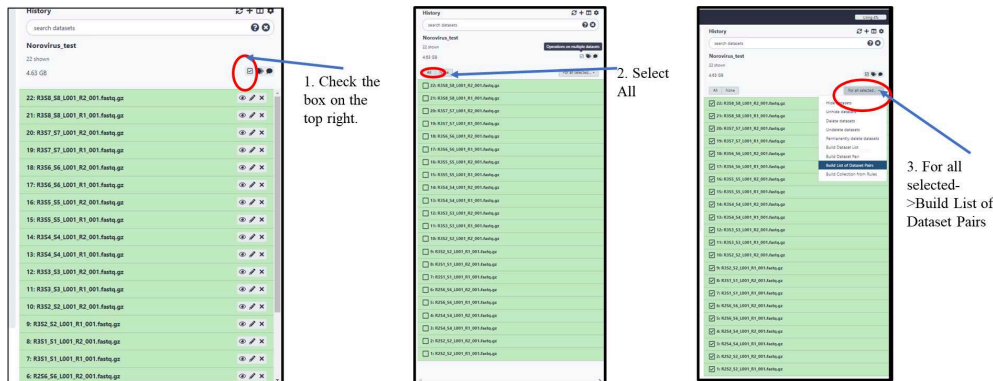
### 3.1.3. Select the files and click "Start" to upload.



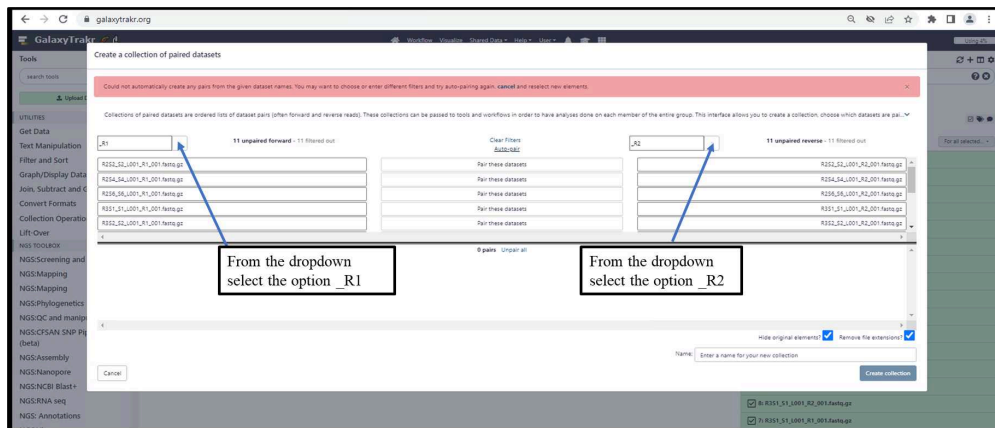
### 3.1.4. Check the status of data upload.



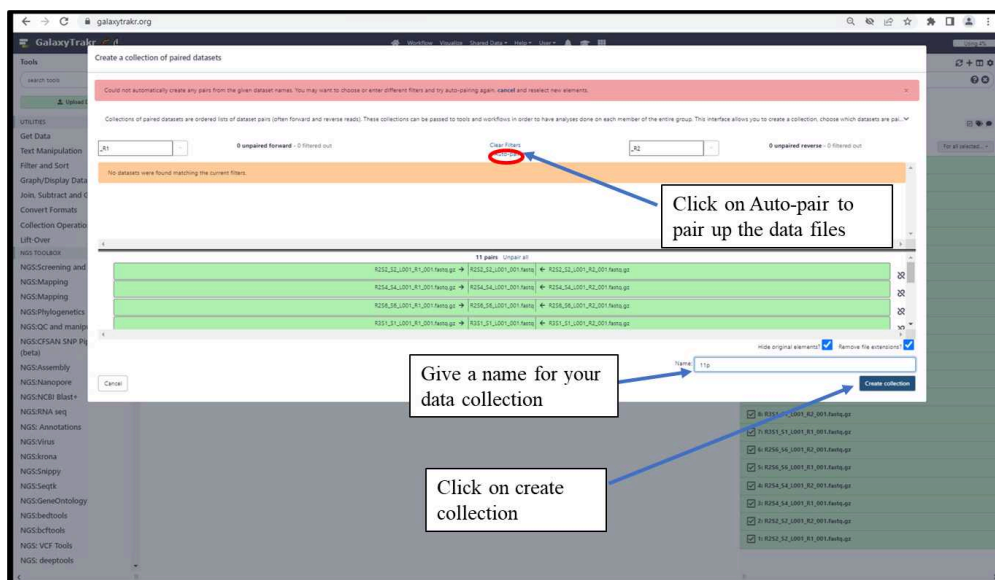
3.1.5. Build a list of Dataset Pairs (pairing the forward and reverse files into their respective samples for batch analysis (Follow steps1, 2 and 3).



3.1.6. Create a collection of paired datasets.

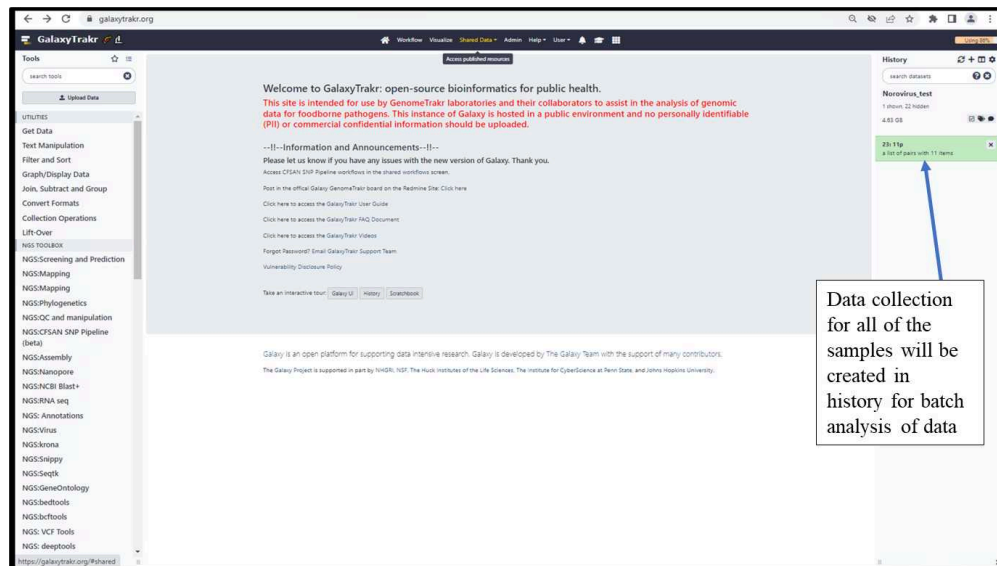


### 3.1.6. Create a collection of paired datasets - Cont.

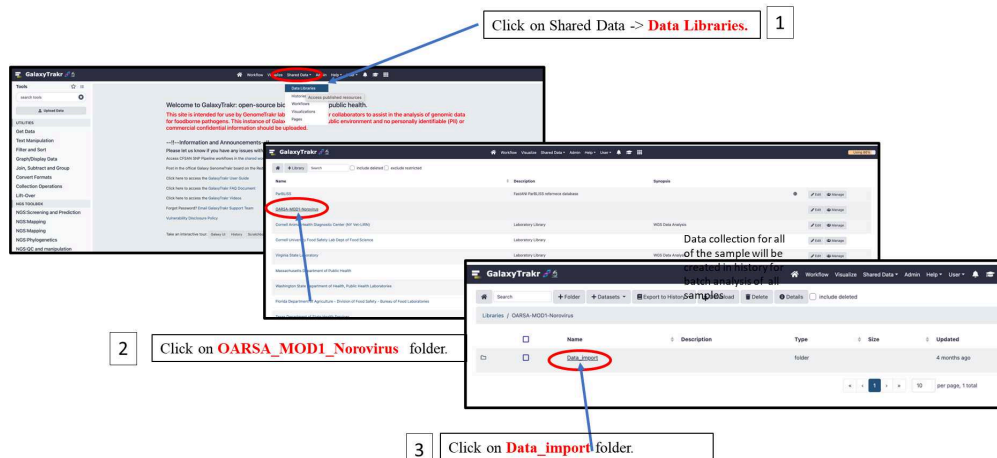


### 3.1.7. Data collection will be created in history.

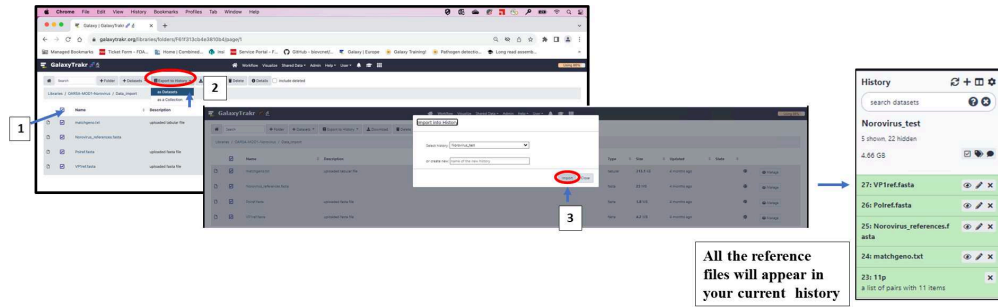




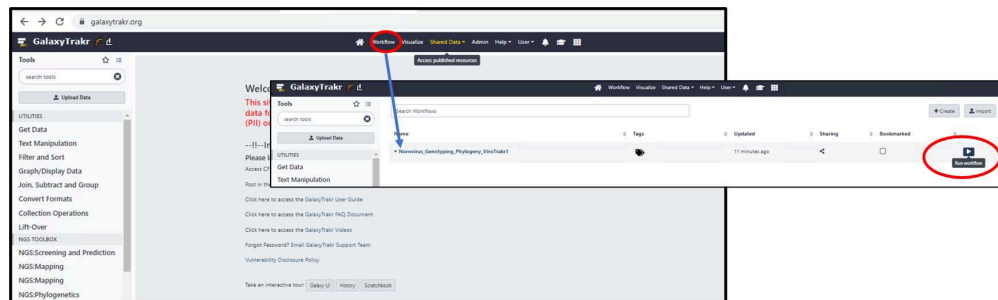
3.1.8. Import the reference data files from Shared Data folder following the steps 1-3 as shown below.



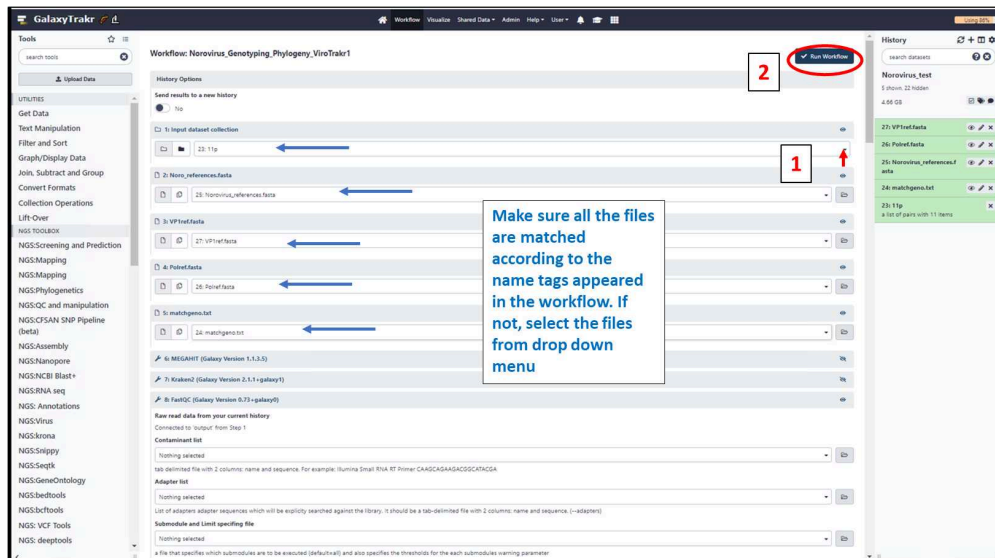
3.1.9. Select all files from the folder and export them as Datasets to your current history following the steps 1-3 as shown below.



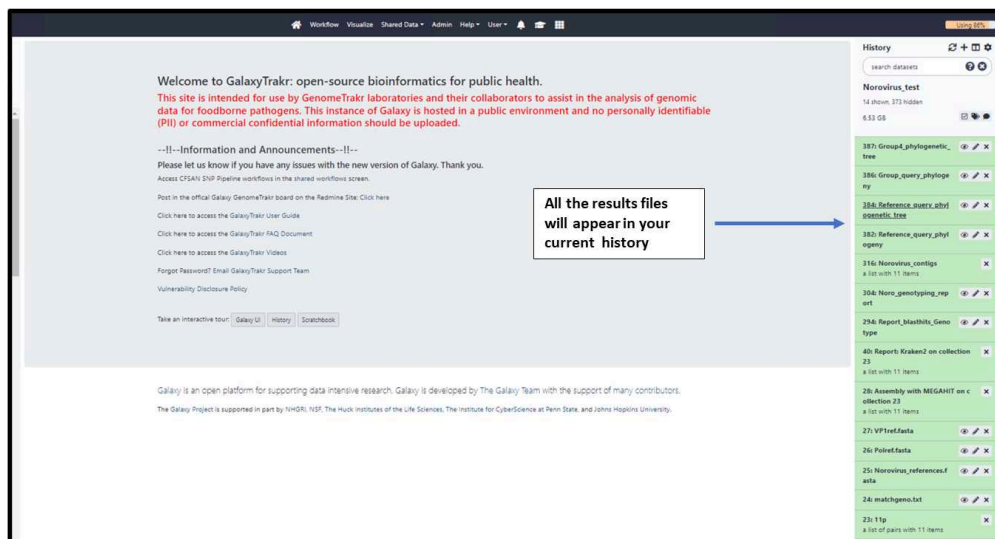
3.1.10. Click on Workflow tab from main menu, select and run the Norovirus\_Genotyping\_Phylogeny workflow.



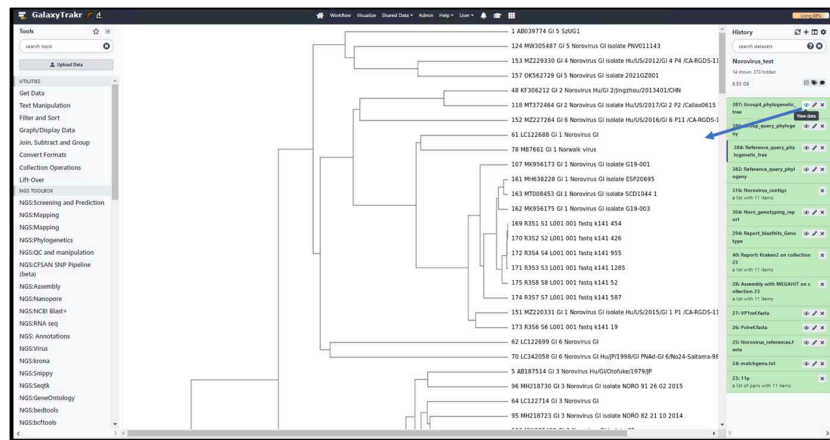
3.1.11. Select all the appropriate files from each dropdown menu and run workflow.



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

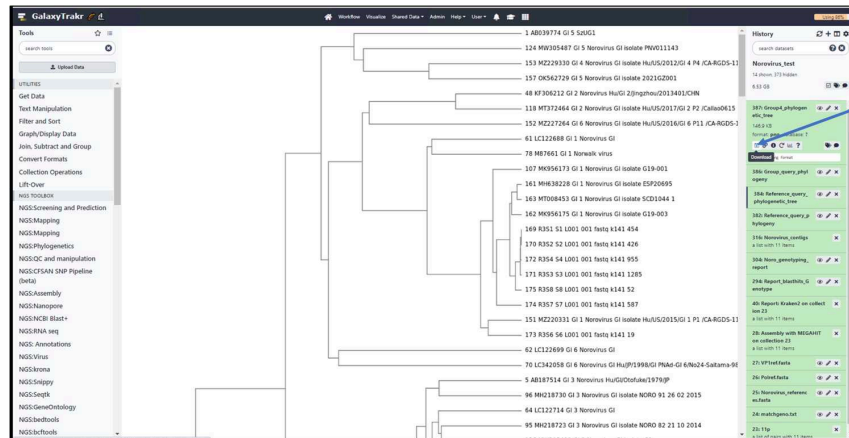


3.1.13. Select and view the result files in the middle panel.



Click on the 'eye' icon to view the results from each file in the middle panel.

### 3.1.14. Download the result files to your local folder.



Click on the 'floppy' icon to download the results files into 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 Group1.4 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](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 NCBI SRA database interface. At the top, there is a search bar with 'SRP173043' entered. Below the search bar, there is a 'Common Fields' section with various metadata fields. The main section is a table of SRR samples. The table has columns for Run, Bytes, Bases, Download, Metadata, and Accession List. The 'Accession List' column is highlighted with a red circle and a red '2' next to it. The table lists 15 SRR samples, each with a Run ID, Bytes, Bases, and a link to the Accession List.

### 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

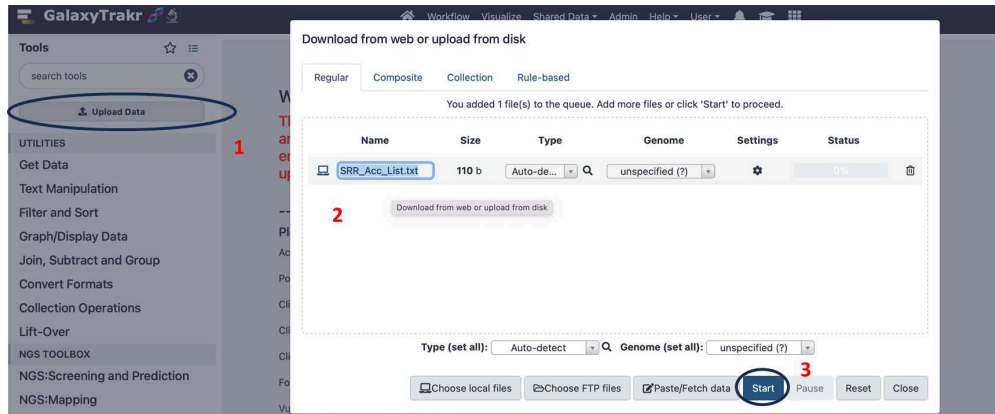
SRR26123215

SRR26123216

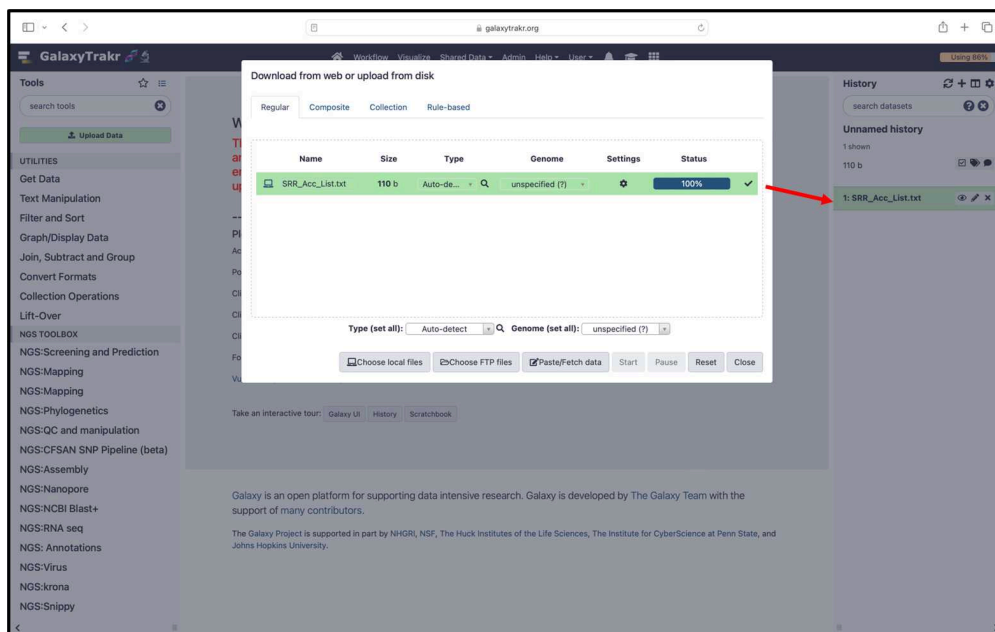
SRR26123217

SRR26123218

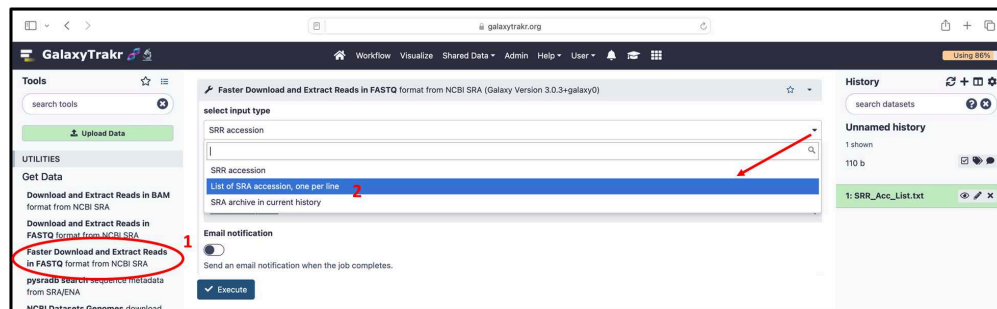
### 3.2.4. To "Upload" the data, select the SRR\_Acc\_list.txt and click the "Start" button:



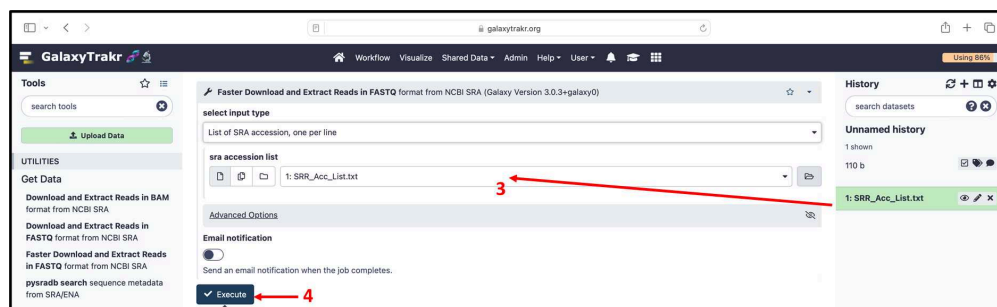
3.2.5. Once the download is completed, the text file will be added in history.



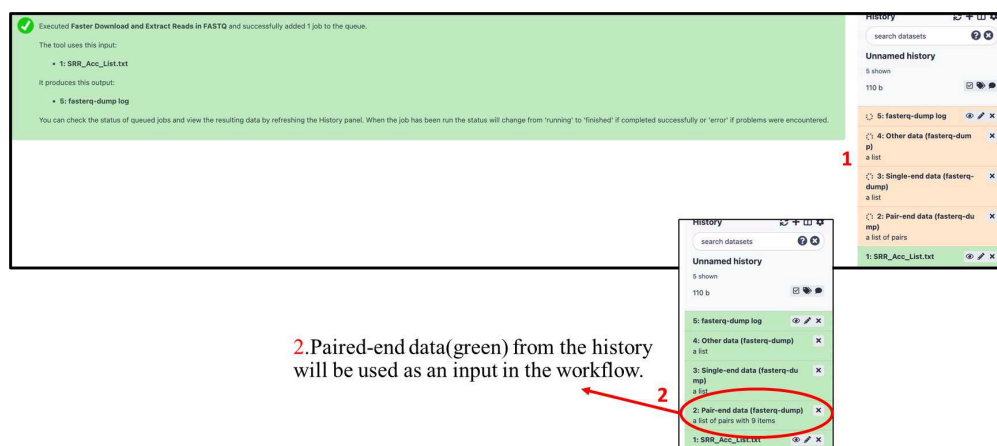
3.2.6. From "Get Data" on the left menu, select "Faster Download and Extract reads in FASTQ from NCBI SRA", and select option by clicking on drop down menu "List of SRA accessions, one per line".



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.