Nov 04, 2024 Version 2

Hepatitis A virus genotyping and phylogeny analysis_ViroTrakr workflow 2_v.2 V.2

DOI
https://dx.doi.org/10.17504/protocols.io.5qpvokj1dl4o/v2

Jayanthi Gangiredla[1], Mark Mammel[1], Zhihui Yang[1]

[1]FDA, HFP, OLOAS, OAMT

GenomeTrakr ViroTrakr

Zhihui Yang
FDA

DOI: https://dx.doi.org/10.17504/protocols.io.5qpvokj1dl4o/v2

Protocol Citation: Jayanthi Gangiredla, Mark Mammel, Zhihui Yang 2024. Hepatitis A virus genotyping and phylogeny analysis_ViroTrakr workflow 2_v.2. **protocols.io** https://dx.doi.org/10.17504/protocols.io.5qpvokj1dl4o/v2Version created by Zhihui Yang

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

Created: November 04, 2024

Last Modified: November 04, 2024

Protocol Integer ID: 111523

Keywords: virotrakr database, galaxytrakrmicrorunqc workflow, virotrakr, sequencing platform, analysis within the galaxytrakr, sequencing data, quality control assessment for microbial genome, galaxy trakr, phylogenetic result, drafting de novo assembly, de novo assembly, sequence genotype, account in galaxy trakr, galaxytrakr, microbial genome, hepatitis, workflow, virus, assembled sequence, sequences from local folder, foodborne virus, raw data

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 hepatitis A virus (HAV) analysis within the GalaxyTrakr platform. It includes the quality assessment for raw sequencing data (from most next-generation sequencing platforms), drafting de novo assemblies, executing the workflow either from the raw sequencing data or assembled sequences, and reporting the sequence genotype and phylogenetic results. This workflow was designed for HAV, which is one of the major targets of our ViroTrakr database.

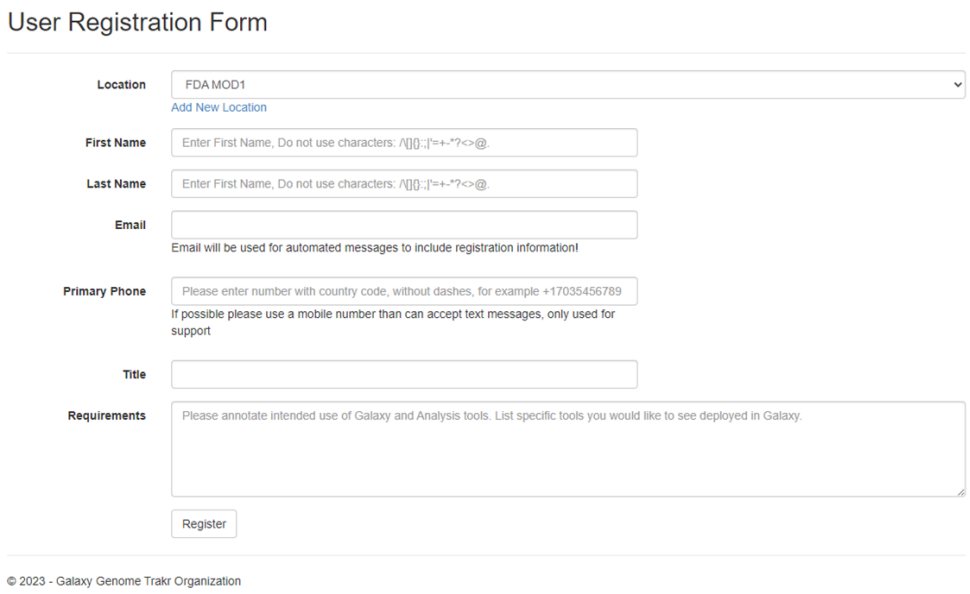
This protocol covers how to:
-Set up an account in Galaxy Trakr (Item 1);
-Create a new history/workspace for a new submission (Item 2);
-Upload raw data obtained from local folders (Item 3.1) or download from NCBI (Item 3.2);
-Upload assembled sequences from local folders (Item 4);
-Execute the ViroTrakr workflow with either raw data or assembled sequences (Item 3.1.8-3.1.12);
-Interpret the results (Item 3.1.13-3.1.14).

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

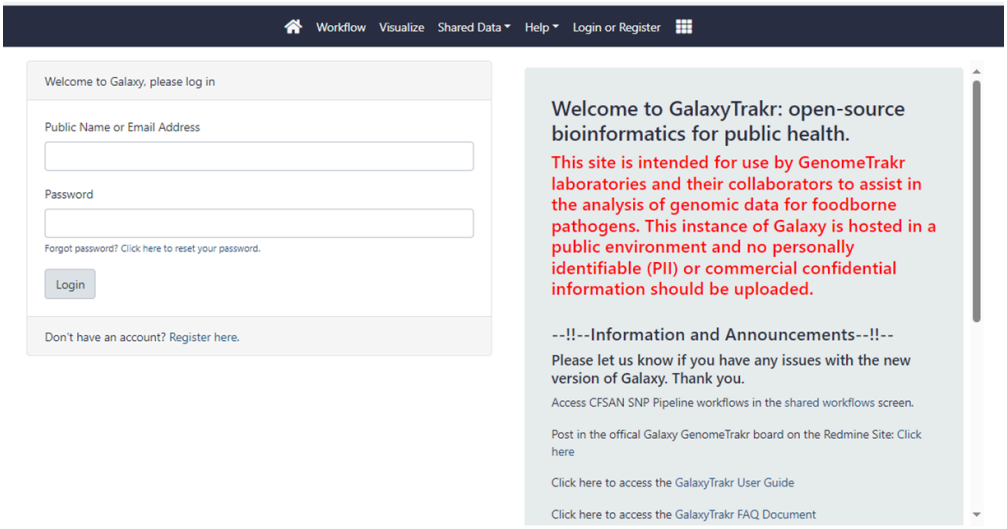
-Reference: Quality control assessment for microbial genomes: GalaxyTrakrMicroRunQC workflow V.5: **Quality control assessment for microbial genomes: GalaxyTrakrMicroRunQC workflow (protocols.io)**

1 Log into your GalaxyTrakr account

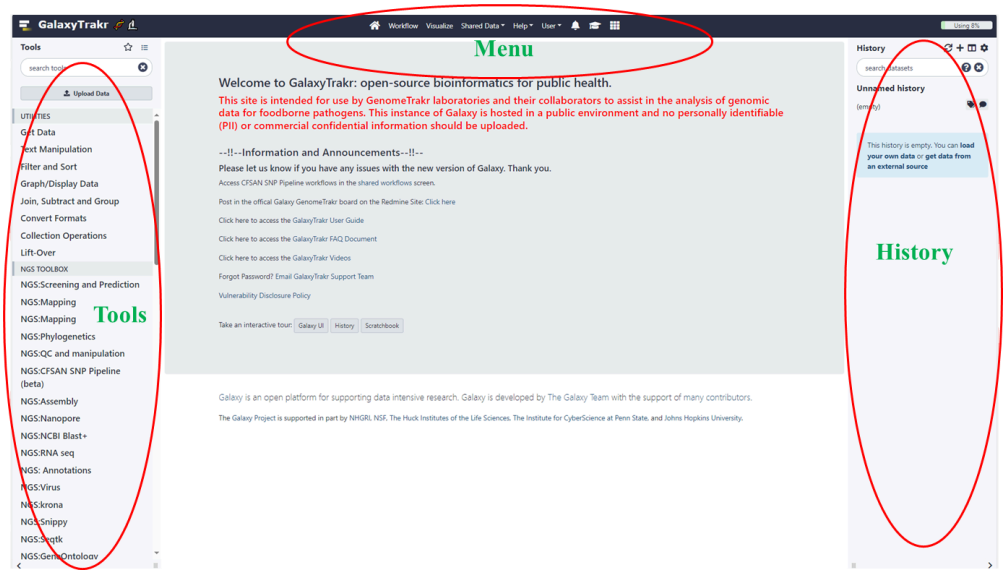
1.1. Create a GalaxyTrakr account if you are the first-time user:
User Registration Form - Galaxy Genome Trakr (galaxytrakr.org)



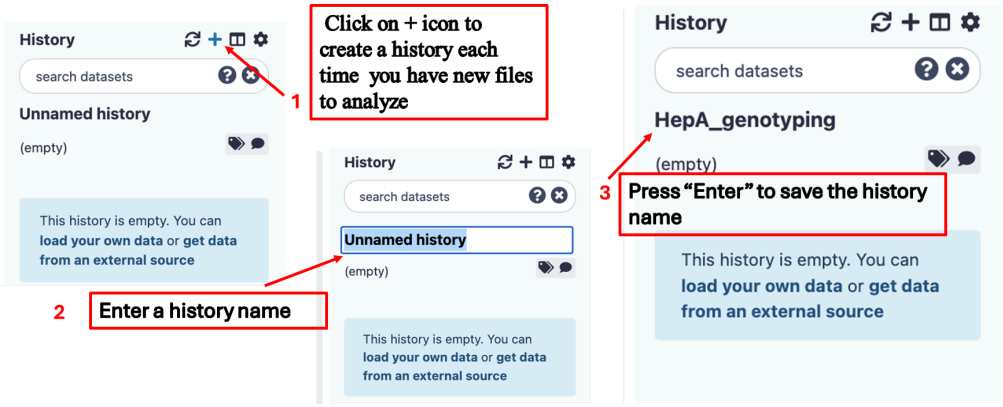
1.2. Log into your GalaxyTrakr account if you already have one:
Galaxy (galaxytrakr.org)



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



2 Create a new history.



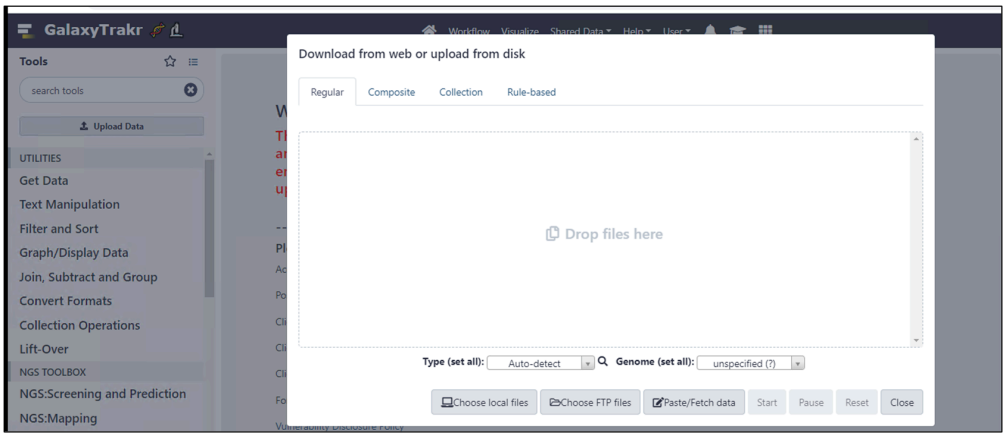
3 Upload raw sequencing data.

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)**).

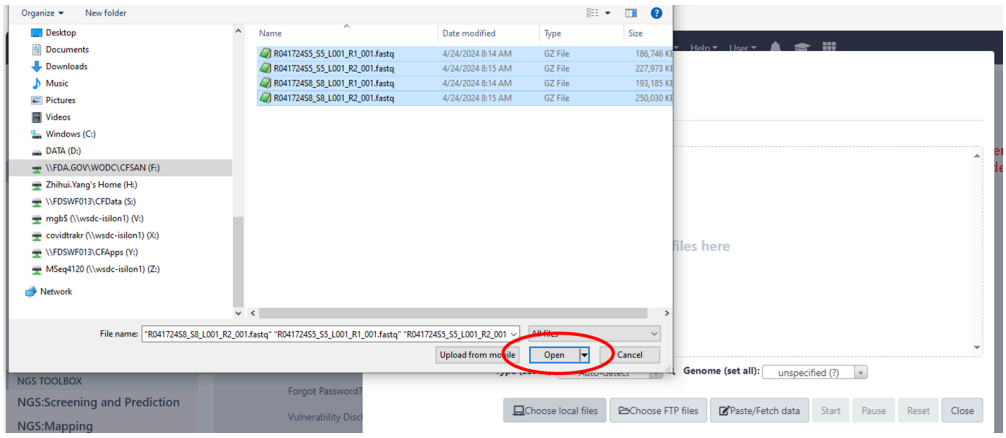
After being uploaded to GalaxyTrakr, the files will remain in your account until they are deleted.

3.1. Upload raw data from local folder.

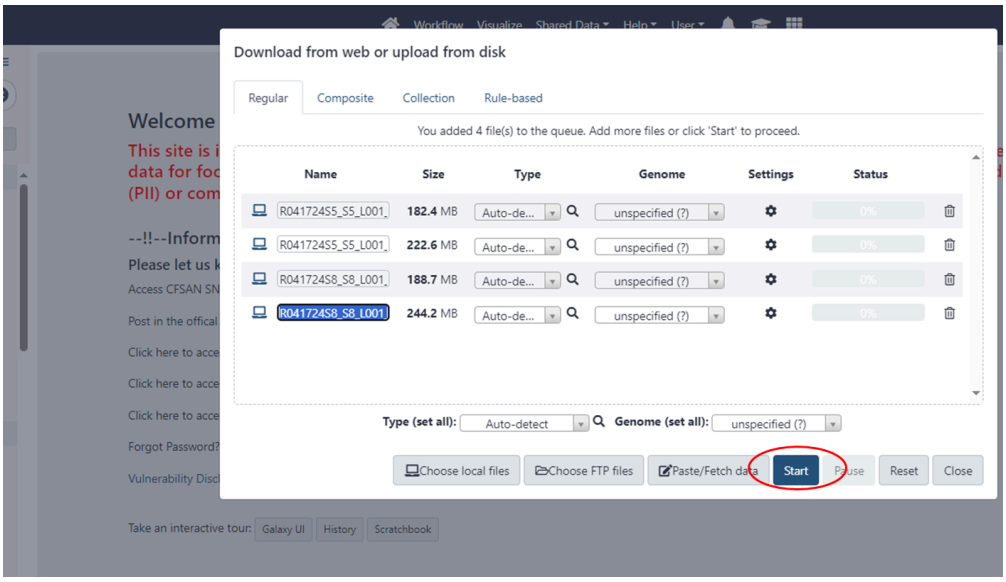
3.1.1. Click on the button “Upload Data”, then “Choose local files”.



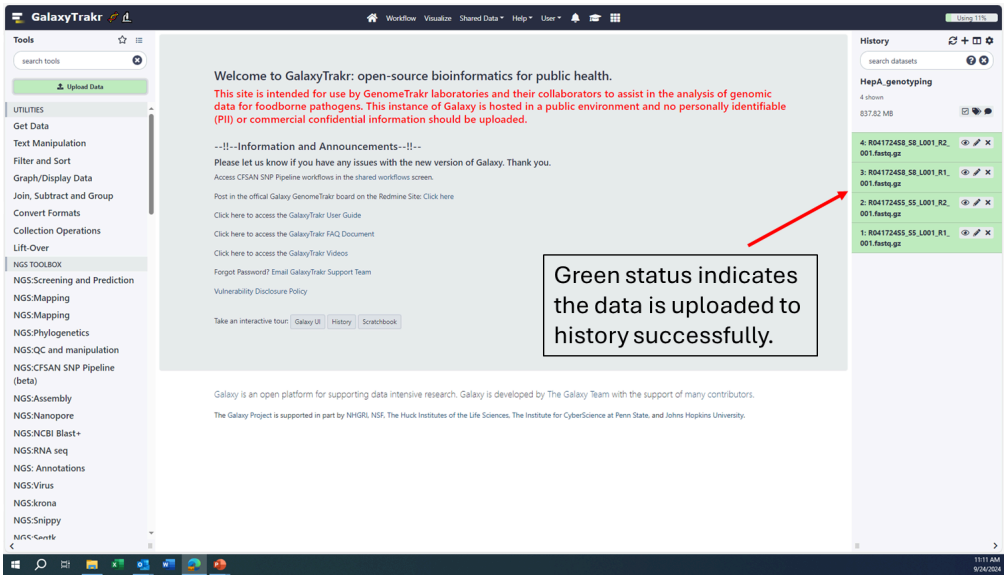
3.1.2. Select fastq files from your local folder.



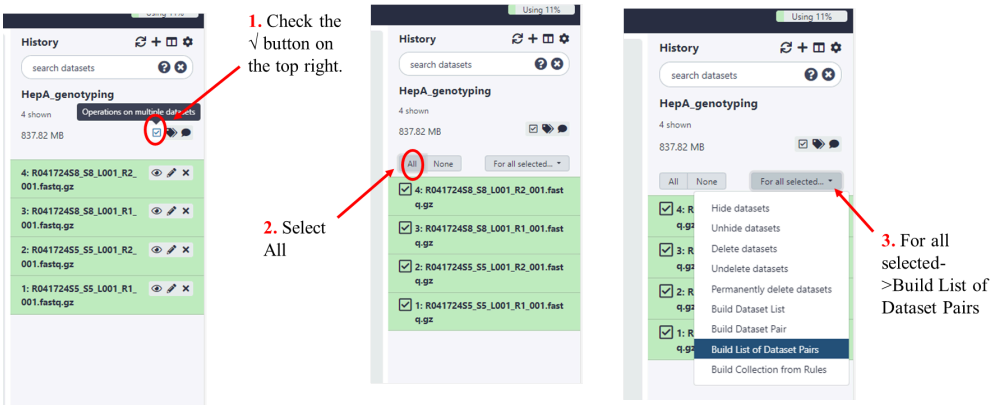
3.1.3. Select the files and click “Start” to upload.



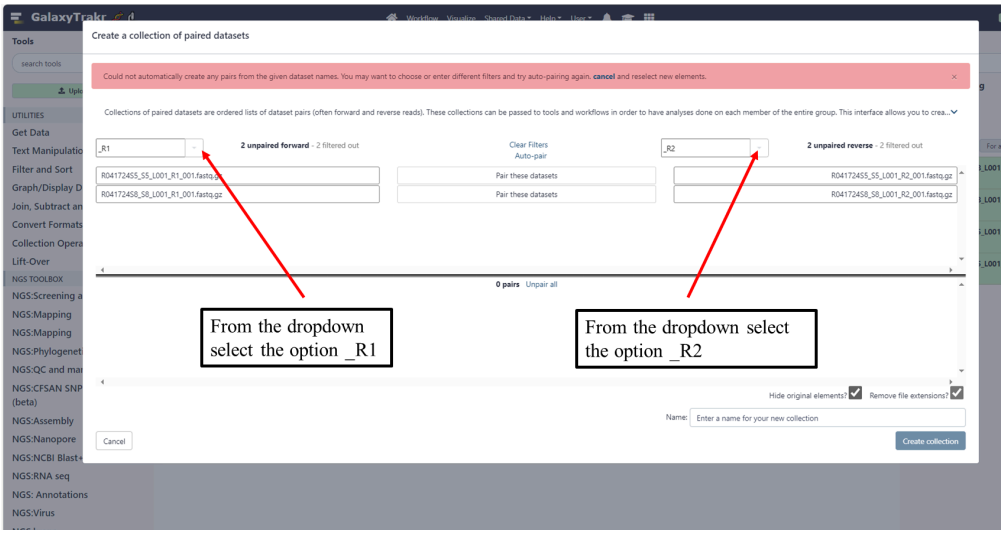
3.1.4. Check the status of data upload.



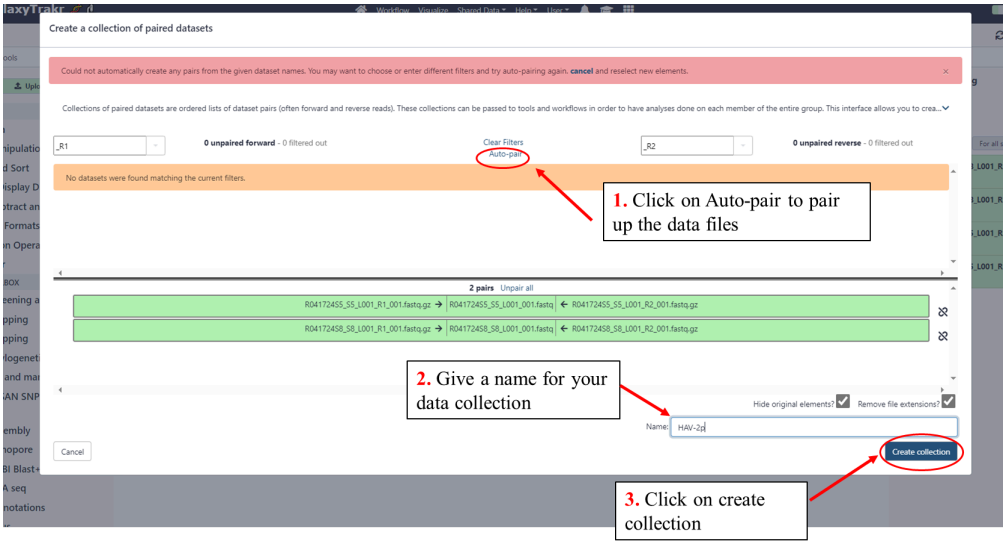
3.1.5. Build a list of Dataset Pairs (pairing the forward and reverse files into their respective samples for batch analysis (Follow steps1, 2 and 3).



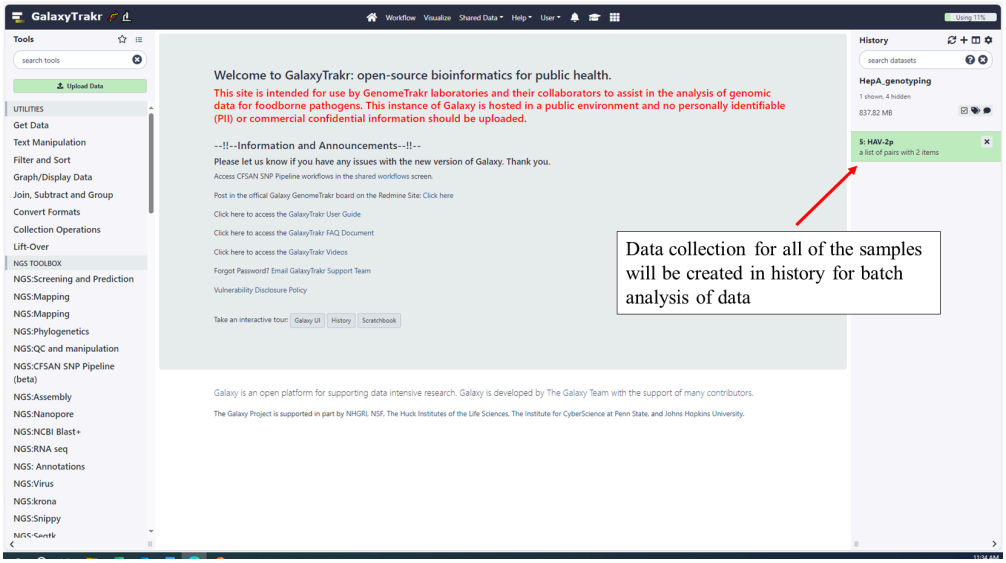
3.1.6. Create a collection of paired datasets.



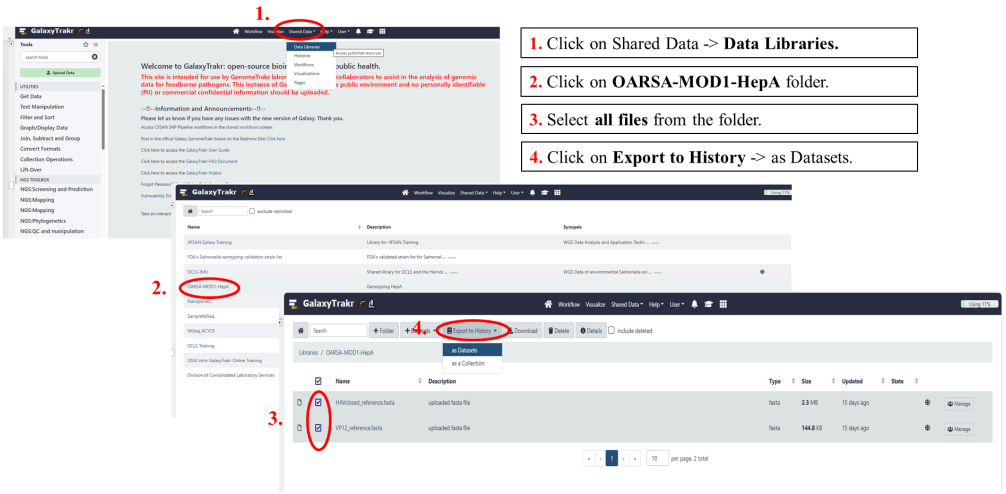
3.1.6. Create a collection of paired datasets - Cont.



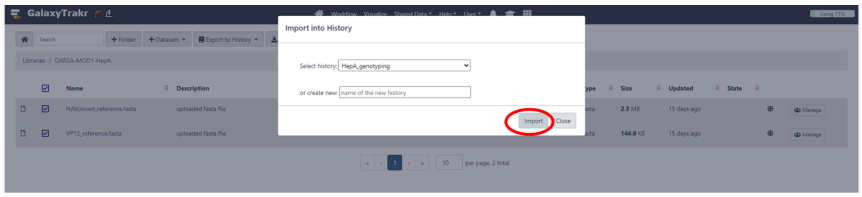
3.1.7. Data collection will be created in history.

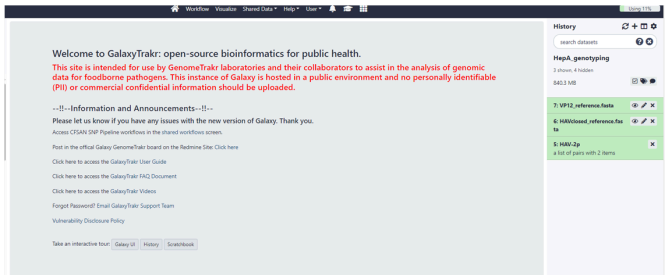


3.1.8. Import the reference data files from Shared Data folder following the steps 1-4 as shown below.

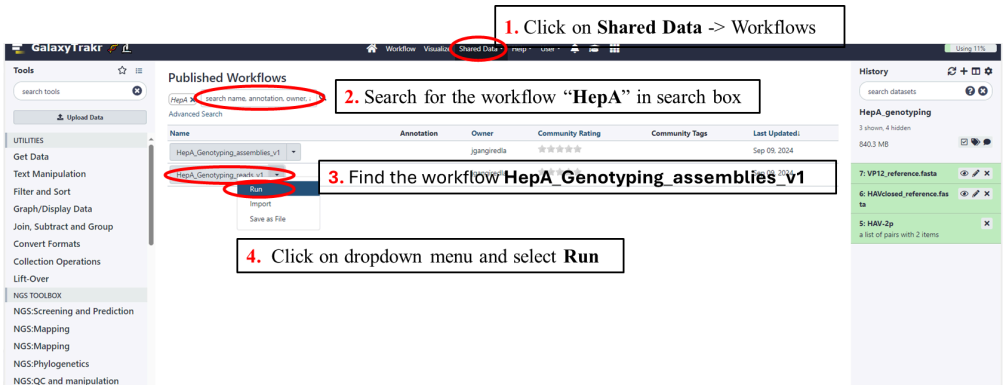


3.1.9. Select all files from the folder and Import them as Datasets to your current history following the steps 1-3 as shown below.

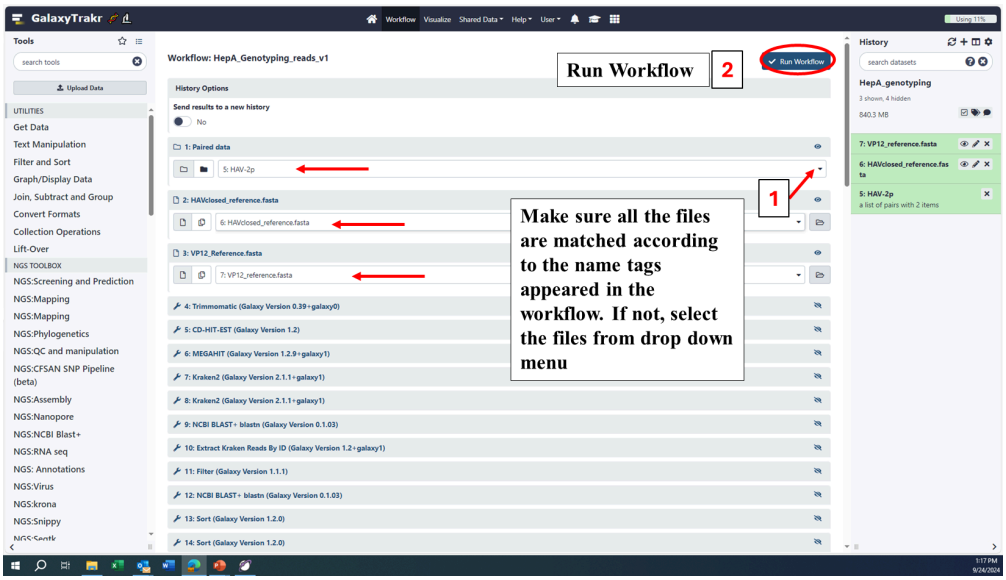




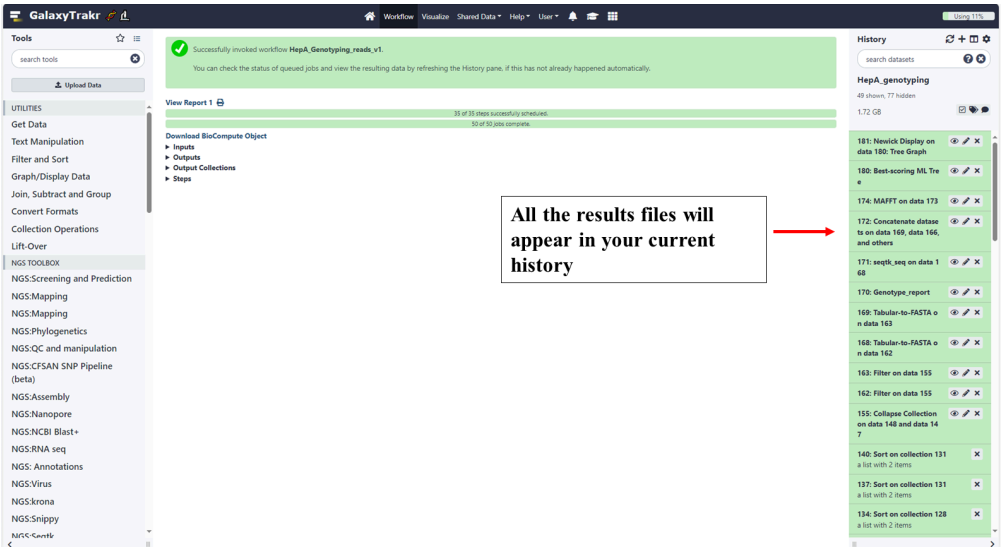
3.1.10. Click on WorkFlow tab from main menu, select and run the HepA_Genotyping_Reads workflow.



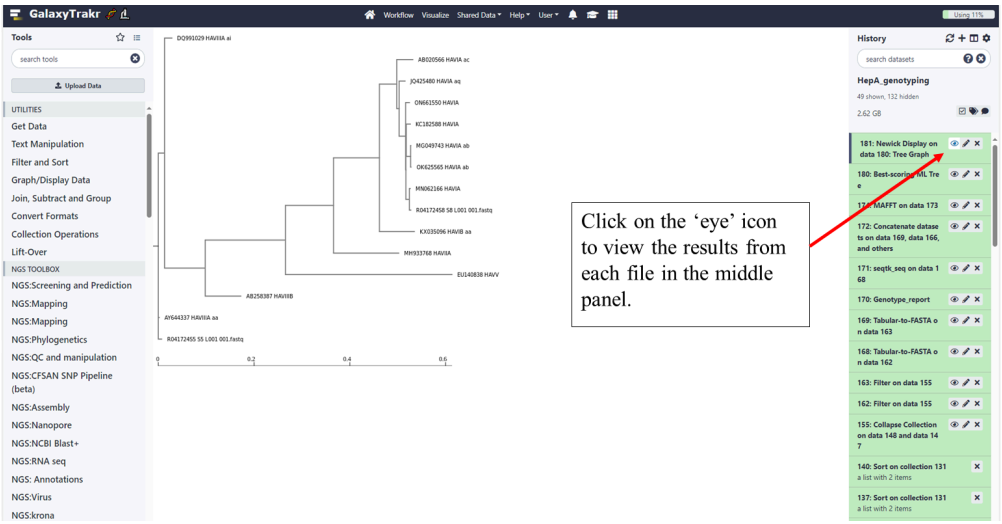
3.1.11. Select all the appropriate files from each dropdown menu and run workflow.



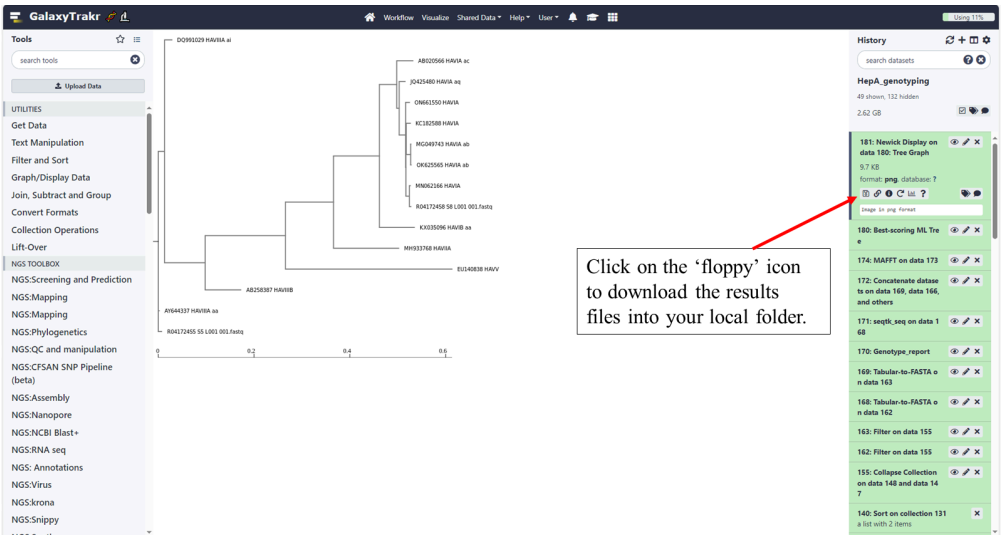
3.1.12. Once the workflow run is successful (Green status), results will appear in the history. Note: workflow takes approximately 5 – 10 minutes for the processing to complete.



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



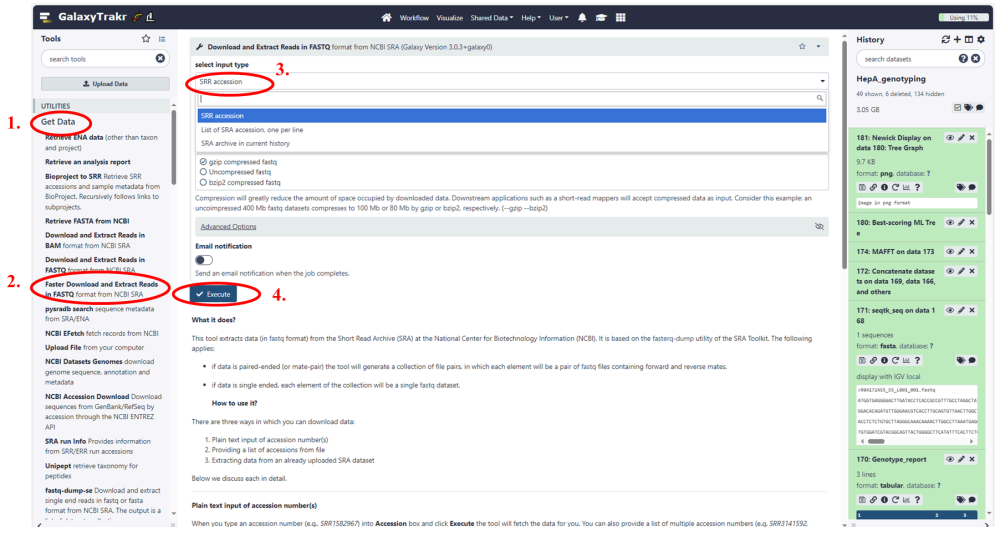
3.1.14. Download the result files to your local folder.



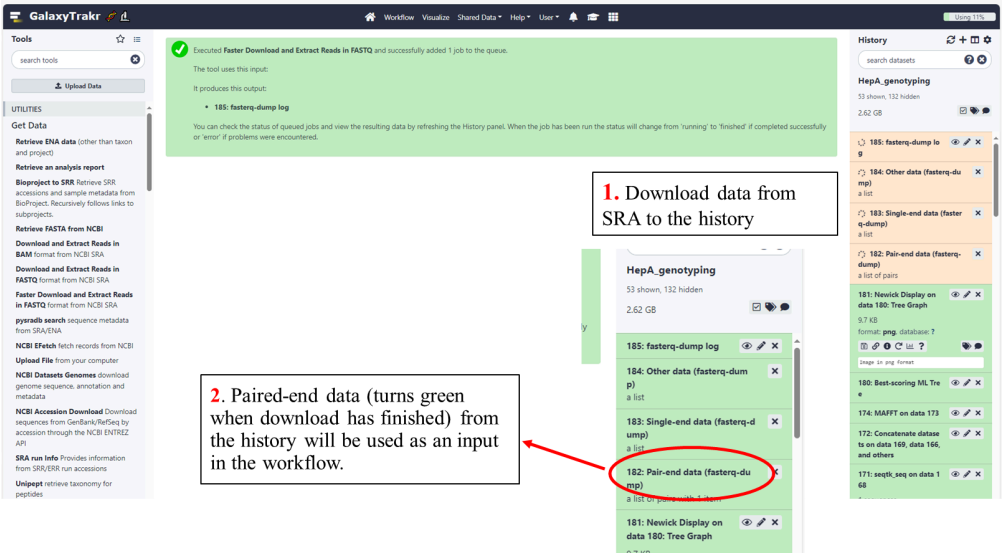
Results files include:
•Assembly with MEGAHIT: Metagenomic assemblies
•Report: Kraken2: Kraken2 reports
•Report_blasthits_Genotype: Reporting Best BLAST Hits against reference sequences
•HAV_genotyping_report: Final report contains QC stats and genotyping results
•HAV_contigs: HAV specific contigs extracted 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.

3.2. Download data from SRA database.
You may also download raw data from SRA if you already submit it to the ViroTrakr, or from other Bioprojects in NCB. You need to get the SRR# or SRA accession# ready for the download.

3.2.1. 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, three options available: (1) “SRR accession” number(s); (2) “List of SRA accessions, one per line”; or (3) SRA archive in current history”. Click button “Execute” once the information is put in.



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

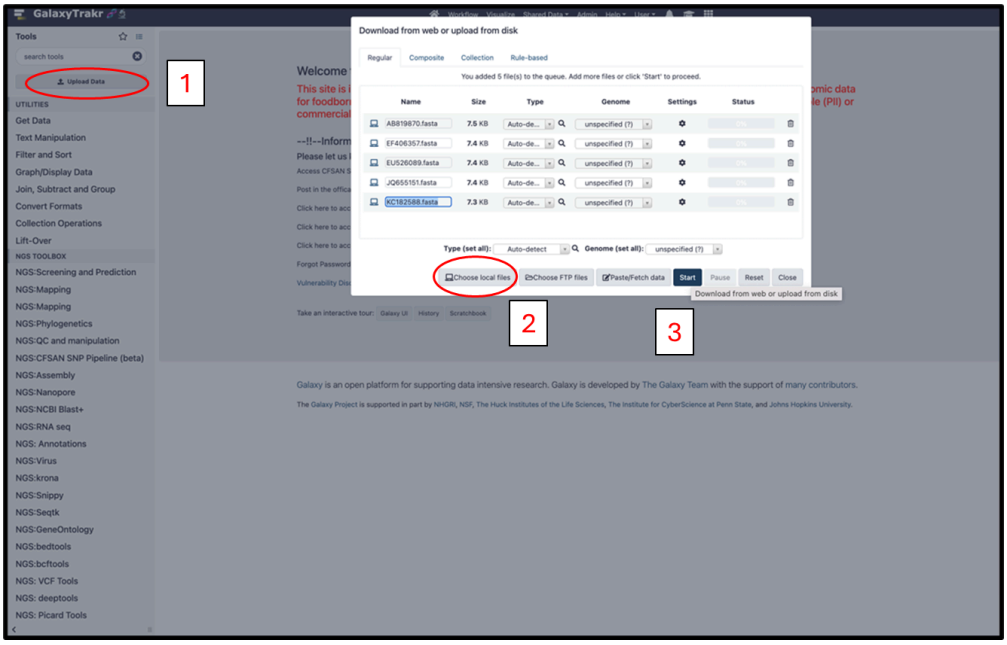


3.2.3. Follow the steps from 3.1.8. to 3.1.14 to run the workflow and collect the results.

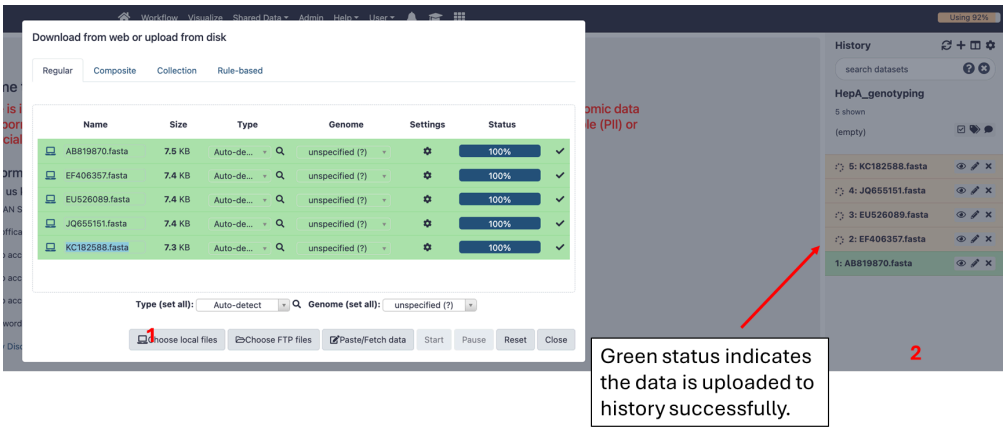
4 Upload assembled sequences.
This ViroTrakr workflow can be used with raw sequencing data, as well as assembled sequences
that you may already obtained with another platform and stored in your local folder.

4.1. Upload assembled sequences from local folder.

4.1.1. Click on the button “Upload Data”, “Choose local files”, then click the button “Start”.

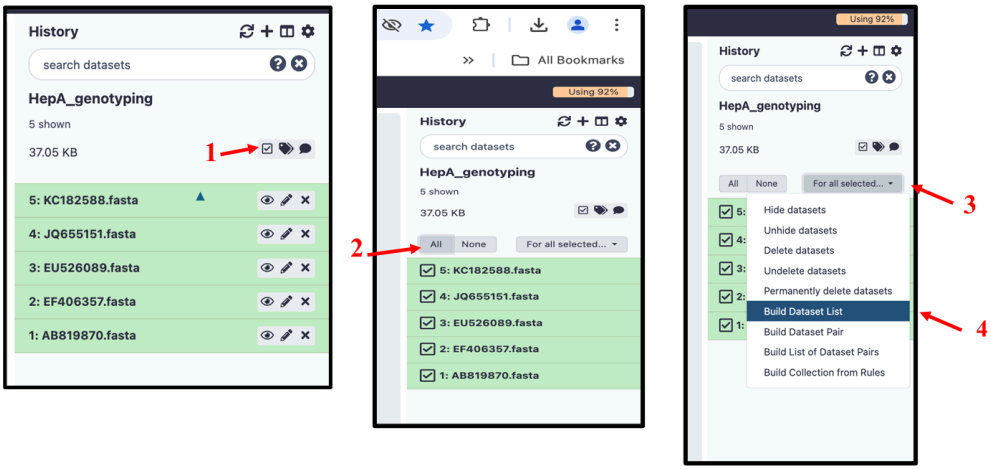


4.1.2. Check the status of data upload.

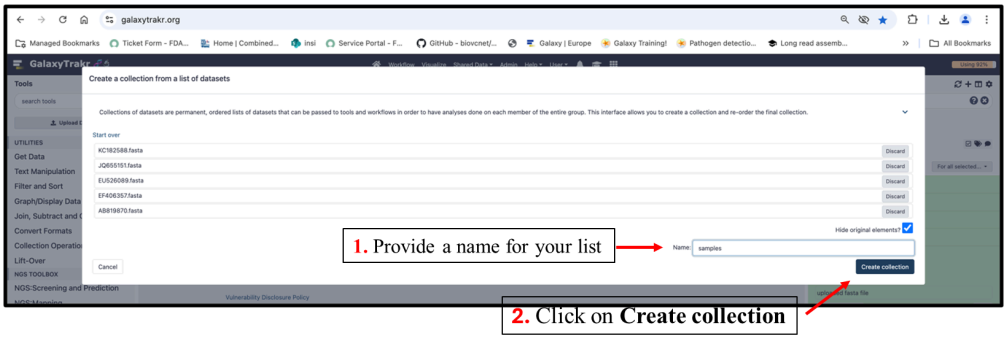


4.1.3. Create a dataset list.

1.click on Checkbox; 2. Click on All to select all the assemblies; 3. click on Drop down on for all selected. 4. Select Build Dataset List.



4.1.4. Build Dataset list on selected samples.



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