Nov 05, 2021

Salmonella serotype prediction using the GalaxyTrakr SeqSero2 workflow

Paul Morin[1], Ruth Timme[1], Michelle Moore[1], Shauna Madson[1], Evelyn Ladines[1], Julia Manetas[1], Karen Jinneman[1]

[1]US Food and Drug Administration

1

dx.doi.org/10.17504/protocols.io.bybfpsjn

GenomeTrakr
Tech. support email: **genomeTrakr@fda.hhs.gov**

Ruth Timme
US Food and Drug Administration

Salmonella serotypes are defined by two surface structures, O antigen and two H antigens. Traditional serotype determination is performed with the *Salmonella* serological somatic (O) and flagellar (H) tests and paired with biochemical confirmation. More than 2,600 *Salmonella* serotypes have been described in the White-Kauffmann-Le Minor scheme. Molecular methods for serotype determination have been developed based on genes responsible for serotype antigens. These genes are encoded in the *rfb* gene cluster, *fliC*, and *fljB*. SeqSero2 is a bioinformatic pipeline that uses whole genome sequence (WGS) data from pure-culture isolates to perform *in silico* analysis to determine the antigenic formula, including somatic (O) antigens and both flagellar (H) antigens. This provides continuity with the well-established scheme for phenotypic *Salmonella* serotypes.

PURPOSE:
This document outlines the steps required to run SeqSero2 v1.1.1 on a collection of isolates in the GalaxyTrakr environment. This is performed by utilizing a custom workflow called “SeqSero2 v1.1.1 collection workflow” and downloading the resulting table.

SCOPE: This protocol covers the following tasks:
1. set up an account in GalaxyTrakr
2. Create a new history/workspace
3. Upload data
4. Execute the SeqSero2 workflow
5. Download the results

Citation: Paul Morin, Ruth Timme, Michelle Moore, Shauna Madson, Evelyn Ladines, Julia Manetas, Karen Jinneman Salmonella serotype prediction using the GalaxyTrakr SeqSero2 workflow https://dx.doi.org/10.17504/protocols.io.bybfpsjn

DOI

dx.doi.org/10.17504/protocols.io.bybfpsjn

Paul Morin, Ruth Timme, Michelle Moore, Shauna Madson, Evelyn Ladines, Julia Manetas, Karen Jinneman 2021. Salmonella serotype prediction using the GalaxyTrakr SeqSero2 workflow. **protocols.io** https://dx.doi.org/10.17504/protocols.io.bybfpsjn

protocol

Gangiredla, J., Rand, H., Benisatto, D. et al. GalaxyTrakr: a distributed analysis tool for public health whole genome sequence data accessible to non-bioinformaticians. BMC Genomics 22, 114 (2021).

salmonella, genomic serotyping, seqsero2, Galaxy

protocol ,

Sep 16, 2021

Nov 10, 2021

53319

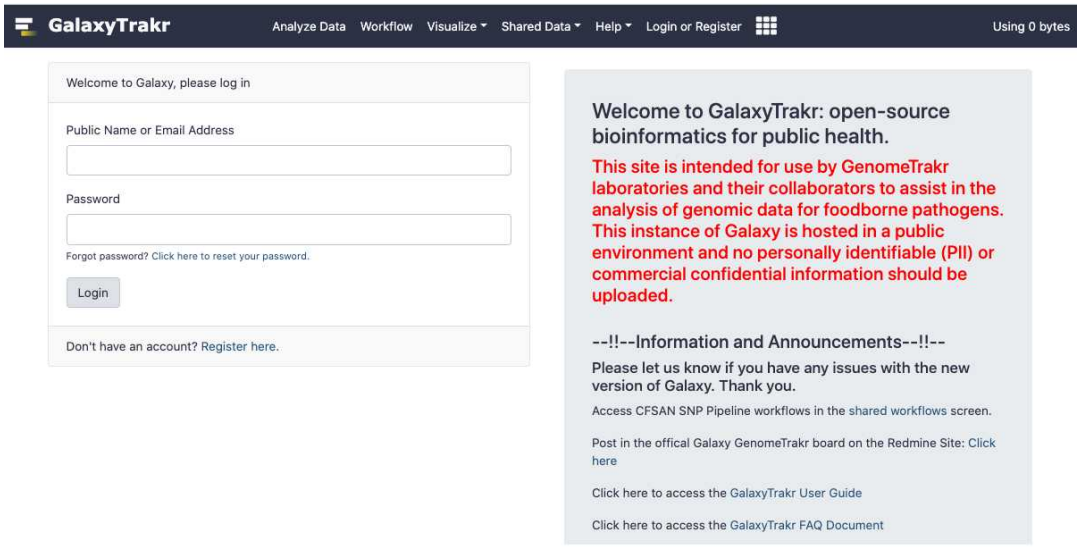
Salmonella WGS fastq files or SRA accessions

:

When using GalaxyTrakr, it is recommended to use Google Chrome for optimal browser experience although Microsoft Edge and Safari are also compatible browsers. Internet Explorer and Mozilla FireFox are NOT compatible with GalaxyTrakr.

Login and import workflow

1 Log into GalaxyTrakr 1909 (https://galaxytrakr.org/root/login)

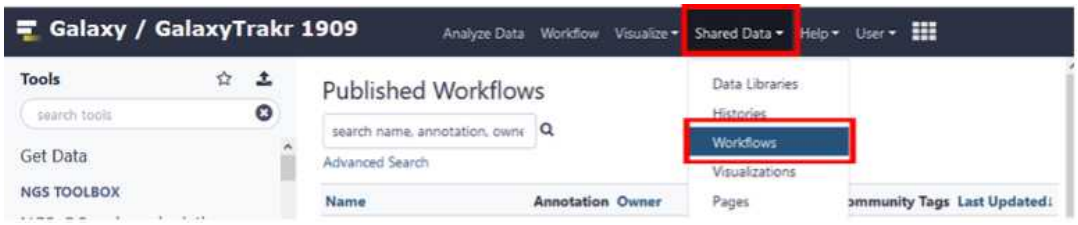


Link to create a new GalaxyTrakr account:
https://account.galaxytrakr.org/Account/Register

1 **Import** the SeqSero2 workflow into the Tools Panel

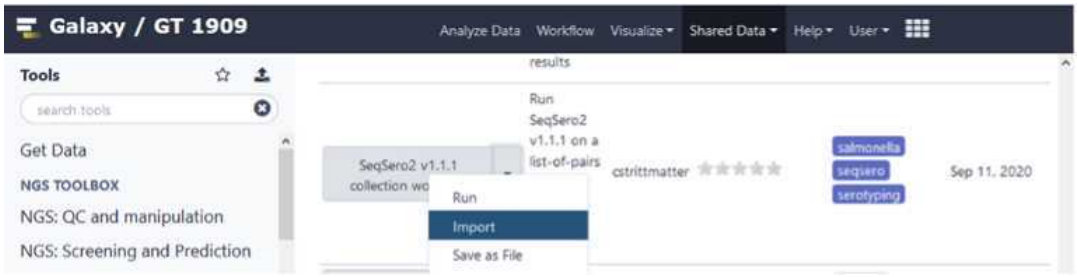
Step 2 only needs to be done once. After this workflow is imported it will be available for use in your Tools Panel.

1.1 Click on **Shared Data** and then **Workflows** from the dropdown menu.



1.2 Select the shared workflow: **SeqSero2 v1.1.1 collection workflow (cstrittmatter 9/11/2020)** and select **import** from the dropdown arrow.

Citation: Paul Morin, Ruth Timme, Michelle Moore, Shauna Madson, Evelyn Ladines, Julia Manetas, Karen Jinneman Salmonella serotype prediction using the GalaxyTrakr SeqSero2 workflow https://dx.doi.org/10.17504/protocols.io.bybfpsjn

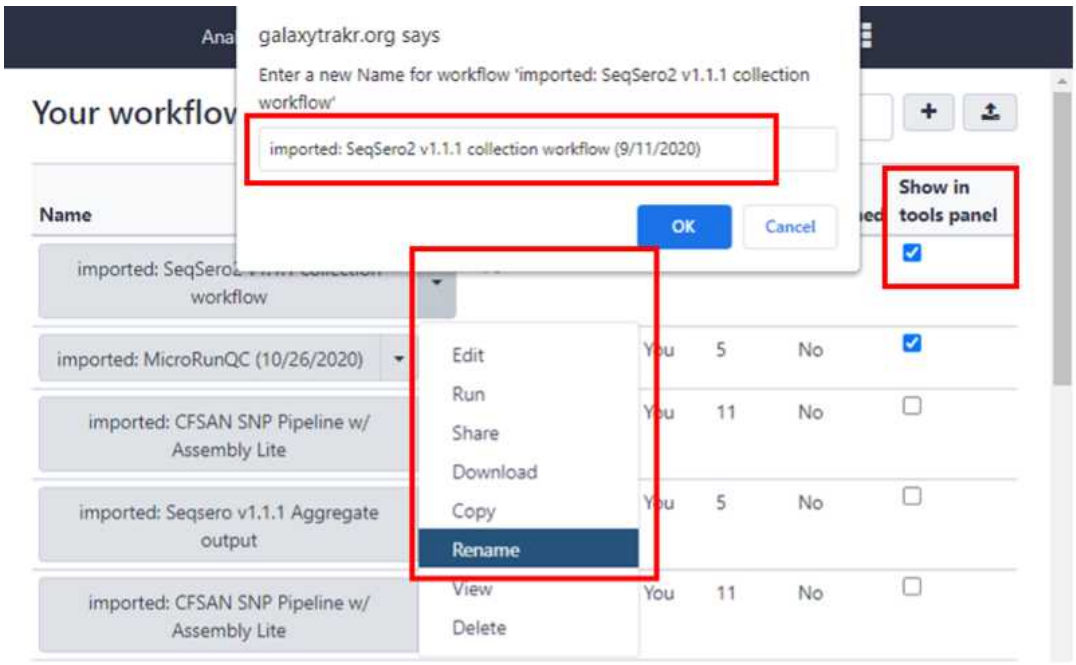


1.3 In **Your Workflows** you can use the dropdown arrow to rename of this workflow.

Adding a date to the name will help you in keeping track of newer versions of this workflow. Workflows do get updated periodically and you want to ensure you are working with the most recent version.

Check the box **"Show in tools panel"**.

This will move the Seqsero2 v1.1.1 collection workflow into your tools panel permanently and you will now have this workflow available to you every time you log into GalaxyTrakr.



Step 2 only needs to be done once for each workflow that is being imported into your Tools.

Import data for analysis

2 If your data is already in GalaxyTrakr, open the history containing that data to be analyzed or move the data to a new history for analysis and proceed to **Step #**. This option may be preferred if the data was already uploaded for other purposes such as MicroRunQC. It's ok if there are

Citation: Paul Morin, Ruth Timme, Michelle Moore, Shauna Madson, Evelyn Ladines, Julia Manetas, Karen Jinneman Salmonella serotype prediction using the GalaxyTrakr SeqSero2 workflow https://dx.doi.org/10.17504/protocols.io.bybfpsjn

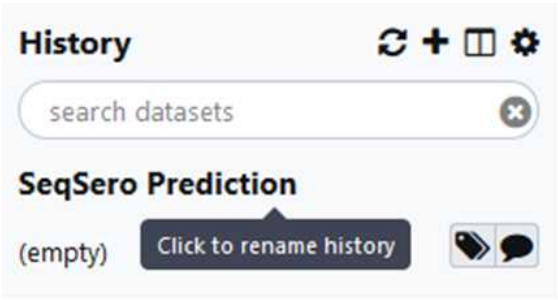
non-*Salmonella* isolates in your dataset.They will not return an antigenic formula or serovar name.

For uploading new data proceed to next step to create a new history and upload your data to be analyzed.

2.1 Create new History:

Click on the "+" button in the upper right corner.

Type in a custom name (i.e., "SeqSero Prediction")

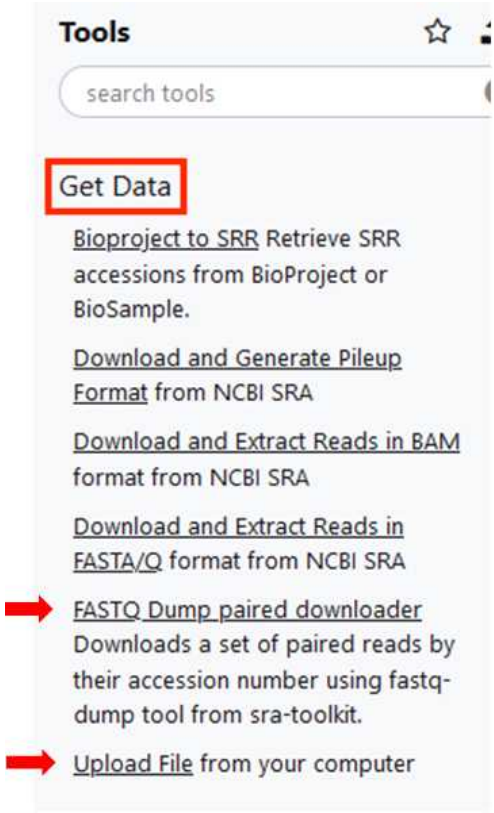


2.2 Import data:

Click on the Tool "**Get Data**", top of left panel.

Click on the sub-tool of choice to bring in data for analysis.

Citation: Paul Morin, Ruth Timme, Michelle Moore, Shauna Madson, Evelyn Ladines, Julia Manetas, Karen Jinneman Salmonella serotype prediction using the GalaxyTrakr SeqSero2 workflow https://dx.doi.org/10.17504/protocols.io.bybfpsjn

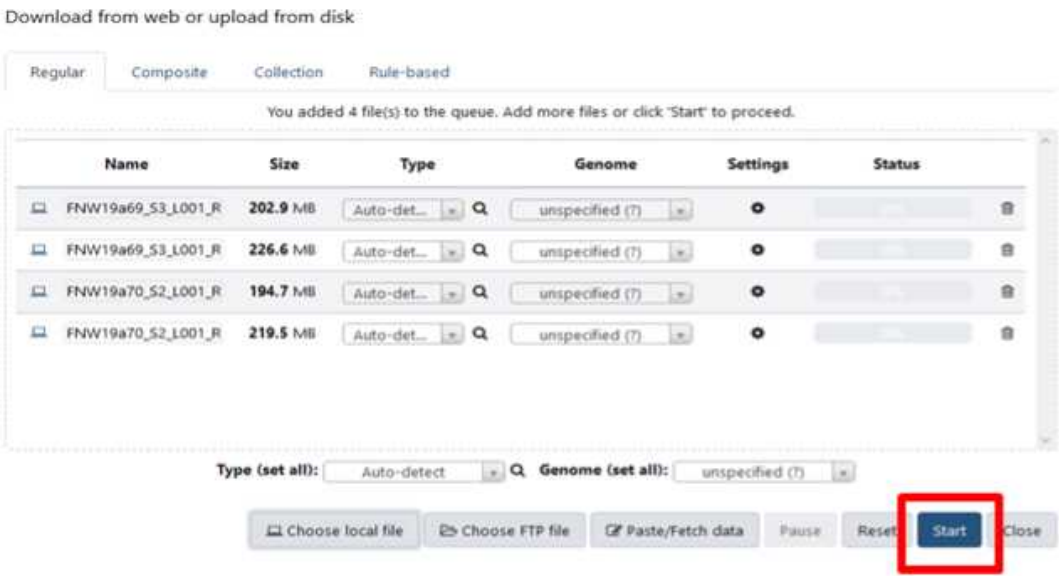


Next steps will show how to import data from NCBI OR upload from your computer.

2.3 "Upload File" for .gz files stored locally.

1. Click on "Choose files"
2. Find your WGS fastq.gz files and select those (2 data files: Read 1 and Read 2 per organism).
3. Click "Start" The amount of time to upload depends on how many files have been selected and the size of those files. The status bar will start to fill as upload progress is made and turn green when completed.

Citation: Paul Morin, Ruth Timme, Michelle Moore, Shauna Madson, Evelyn Ladines, Julia Manetas, Karen Jinneman Salmonella serotype prediction using the GalaxyTrakr SeqSero2 workflow https://dx.doi.org/10.17504/protocols.io.bybfpsjn



2.4 "FASTQ Dump Paired downloader" to import data from NCBI.

1. Enter the NCBI SRR for each sequence to be retrieved.
2. Click "Execute"

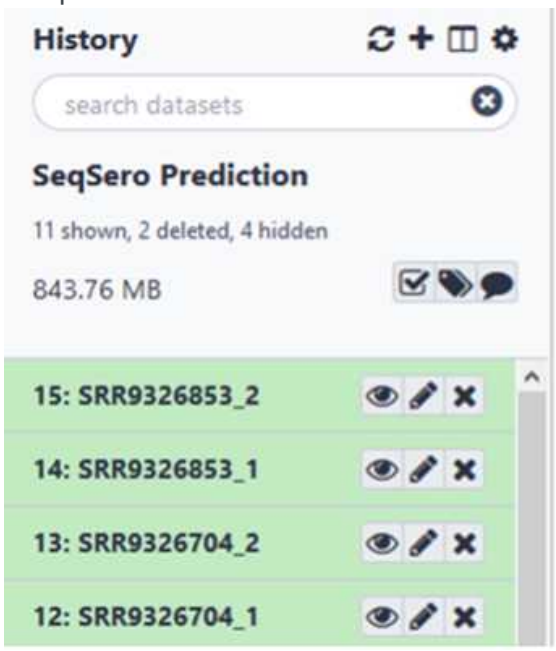
2.5 When the data has finished importing, you should see the successfully uploaded files listed in green in the right panel.

Files will be highlighted in RED if they were NOT successfully uploaded.

Example of .gz files uploaded:

Citation: Paul Morin, Ruth Timme, Michelle Moore, Shauna Madson, Evelyn Ladines, Julia Manetas, Karen Jinneman Salmonella serotype prediction using the GalaxyTrakr SeqSero2 workflow https://dx.doi.org/10.17504/protocols.io.bybfpsjn

Example of SRR data downloaded from NCBI:

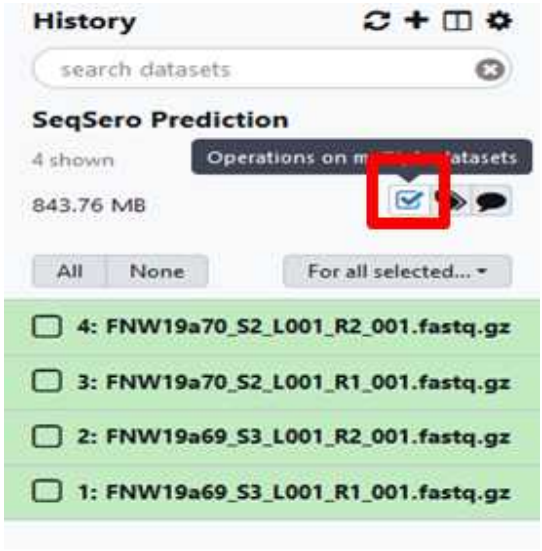


Build your dataset of paired-reads

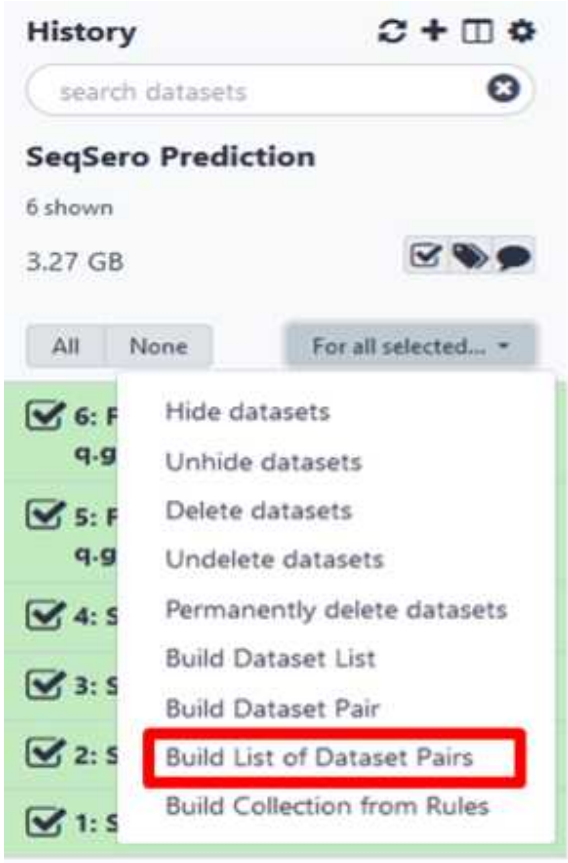
3 **Build "list of data set pairs"**

Citation: Paul Morin, Ruth Timme, Michelle Moore, Shauna Madson, Evelyn Ladines, Julia Manetas, Karen Jinneman Salmonella serotype prediction using the GalaxyTrakr SeqSero2 workflow https://dx.doi.org/10.17504/protocols.io.bybfpsjn

3.1 Click on the check mark in the history panel then select all files you want to include in the data set for SeqSero analysis.

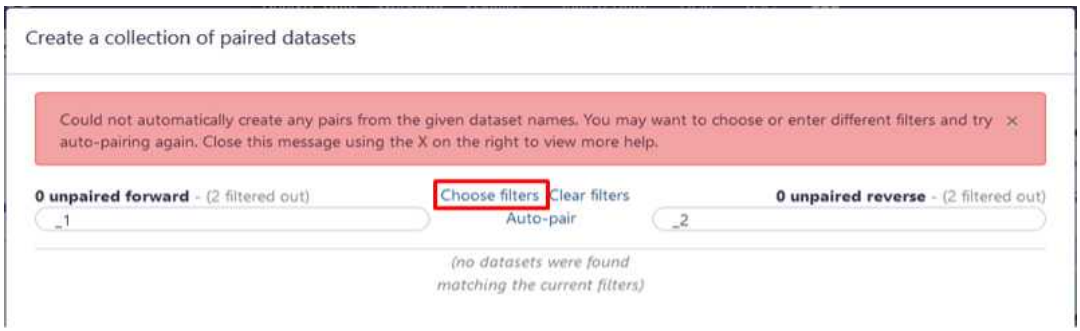


3.2 Open options under “For all selected” and then choose “Build List of Dataset Pairs”

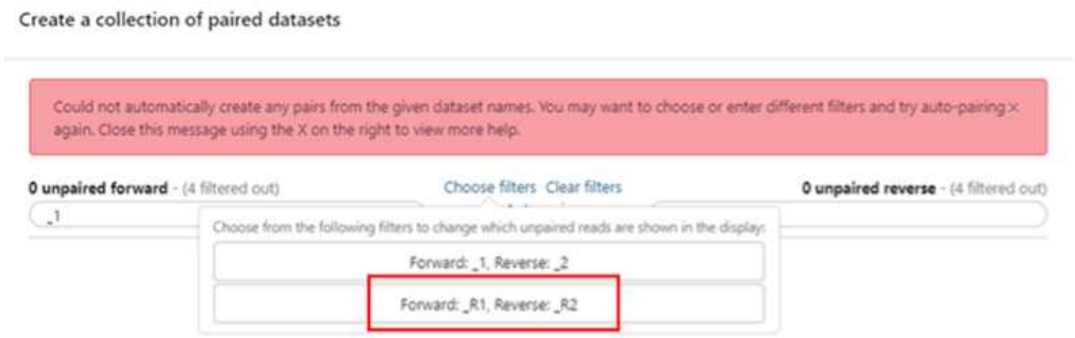


Citation: Paul Morin, Ruth Timme, Michelle Moore, Shauna Madson, Evelyn Ladines, Julia Manetas, Karen Jinneman Salmonella serotype prediction using the GalaxyTrakr SeqSero2 workflow https://dx.doi.org/10.17504/protocols.io.bybfpsjn

3.3 Click "**Choose Filters**"



3.4 Click the correct file extension "**Forward: _R1, Reverse: _R2"**



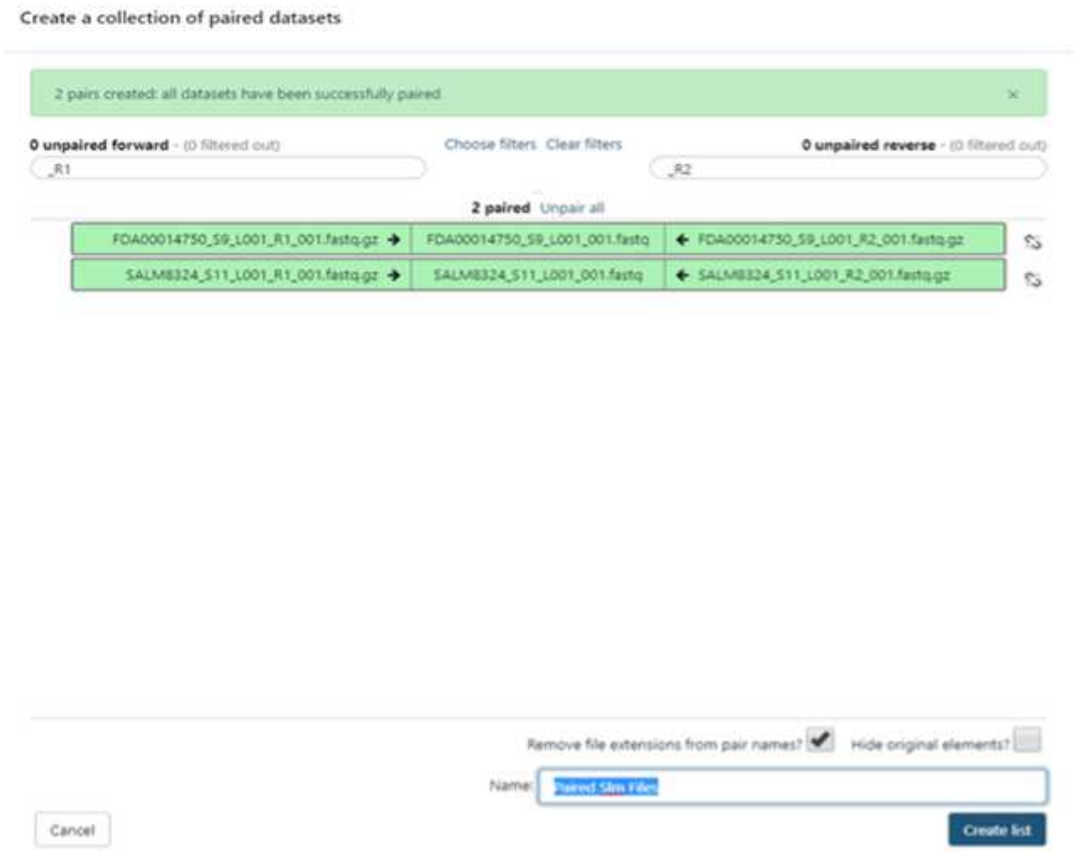
3.5 Click "**Auto-pair**"

The Read 1 and Read 2 fastq.gz files should automatically pair together. Note: For data downloaded from NCBI the two reads will already be paired and you will not need to select filter and auto-pair.

3.6 Type in a custom name for the dataset (i.e., "Paired Slm Files")

Citation: Paul Morin, Ruth Timme, Michelle Moore, Shauna Madson, Evelyn Ladines, Julia Manetas, Karen Jinneman Salmonella serotype prediction using the GalaxyTrakr SeqSero2 workflow https://dx.doi.org/10.17504/protocols.io.bybfpsjn

Click “**Create list”**



You should see your named list in the history panel

Analyze your data using the SeqSero2 workflow

4 In NGS TOOLBOX, left panel:

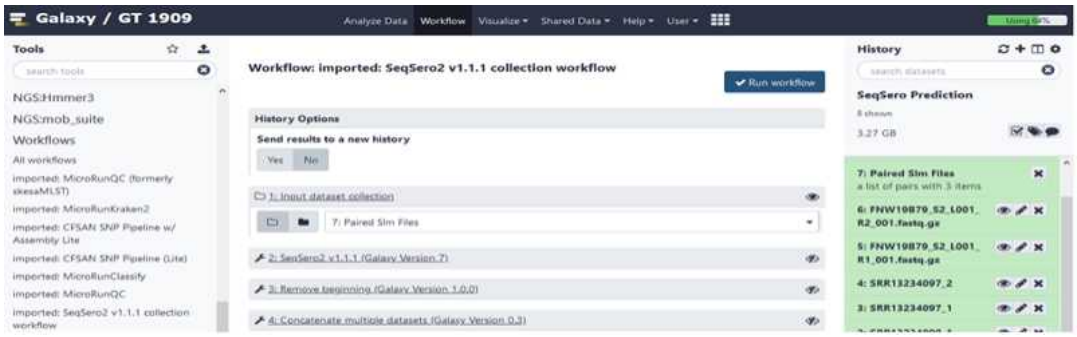
Click on the imported and saved version of the “**SeqSero2 v1.1.1 collection workflow**”.

4.1 In the Main window, the newly created list of paired files should automatically show up in the “Input dataset collection” window.

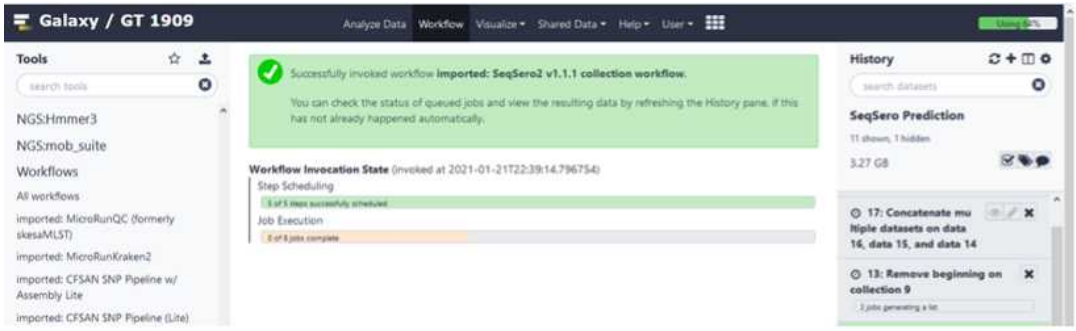
If it doesn’t, click and drag the file from your history panel into the “Input dataset collection” window.

Click “**Run Workflow”**

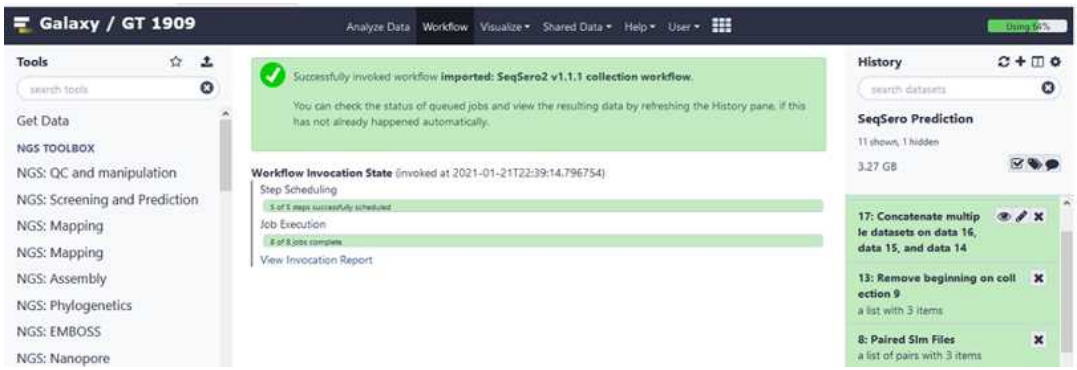
Citation: Paul Morin, Ruth Timme, Michelle Moore, Shauna Madson, Evelyn Ladines, Julia Manetas, Karen Jinneman Salmonella serotype prediction using the GalaxyTrakr SeqSero2 workflow https://dx.doi.org/10.17504/protocols.io.bybfpsjn



4.2 Your working panel should appear green with a white check mark on the upper left-hand corner.



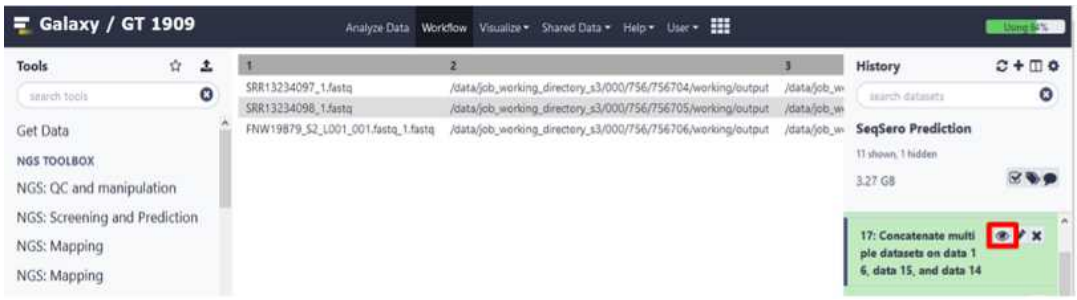
4.3 After the SeqSero analysis is complete, the “Concatenate multiple datasets on data” will appear green.



View and export results

5 Click on the “eyeball” in the “**Concatenate multiple datasets on data”** to view table of predicted serotypes for the collection.

Citation: Paul Morin, Ruth Timme, Michelle Moore, Shauna Madson, Evelyn Ladines, Julia Manetas, Karen Jinneman Salmonella serotype prediction using the GalaxyTrakr SeqSero2 workflow https://dx.doi.org/10.17504/protocols.io.bybfpsjn



Note: Scroll across the table to see additional information.

5.1 **Export SeqSero results:** cut/paste method

1	2	3	4	5	6	7	8	9
SRR13234097_1.1	/data/job_	/data/job_	60	r	e,n,x,z15	IIIb	60:r:e,n,x,	IIIb 60:r:e,n,x,z15
SRR13234098_1.1	/data/job_	/data/job_	9,46	z29	-	I	9,46:z29:-	Ouakam
FNW19B79_S2_L	/data/job_	/data/job_	4	k	e,n,z15	I	4:k:e,n,z1!	Texas

1. Click and drag to highlight text
2. Copy
3. Paste Special as “Text” or “Unicode Text” into Excel

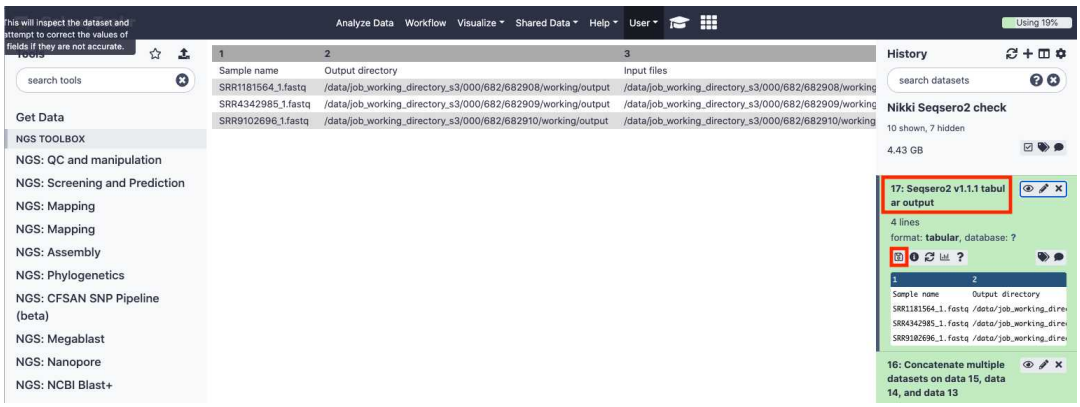
Alternatively, click on the table and “Ctrl-A” to select the entire Table, “Ctrl-C” to copy data and paste the copied data into Excel by “Ctrl-V”

5.2 **Export SeqSero2 results**: download tab-delimited text file

Click the dataset name.

The panel will expand, enabling more options.

Click the "**Save**" icon to download a tab-delimited file of results.

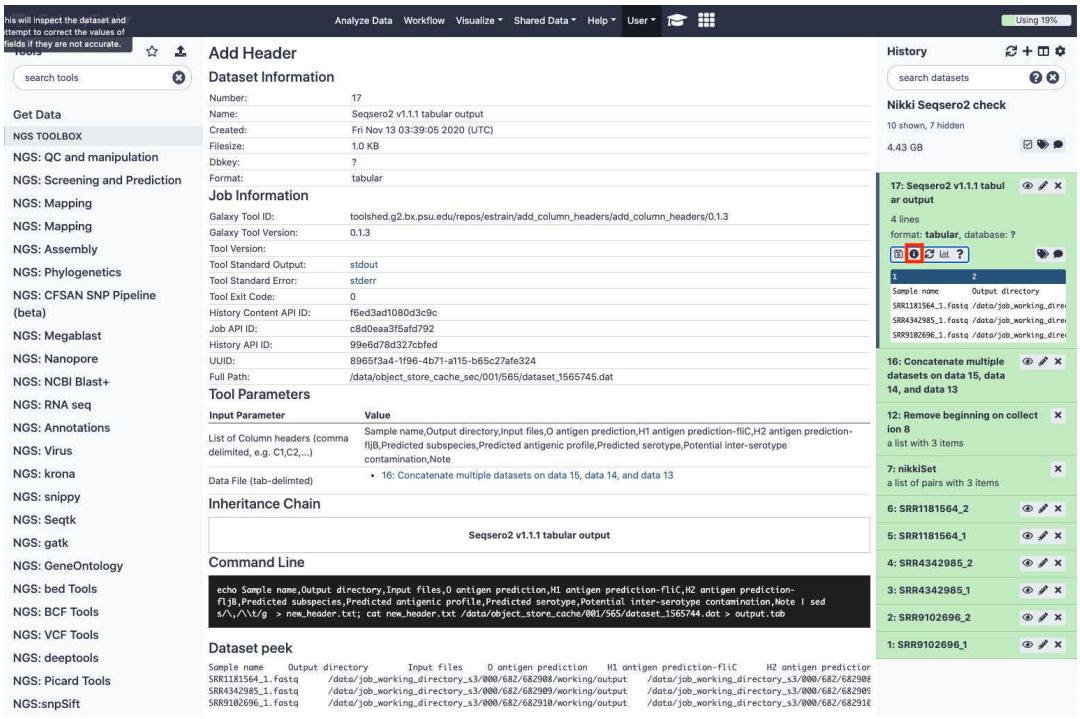


Citation: Paul Morin, Ruth Timme, Michelle Moore, Shauna Madson, Evelyn Ladines, Julia Manetas, Karen Jinneman Salmonella serotype prediction using the GalaxyTrakr SeqSero2 workflow https://dx.doi.org/10.17504/protocols.io.bybfpsjn

Example results file: SeqSeroExampleResults.tabular

5.3 Optional:

The small "Info" icon results in a detailed view of the dataset, analysis, parameters used, etc., which can be helpful for troubleshooting.



Citation: Paul Morin, Ruth Timme, Michelle Moore, Shauna Madson, Evelyn Ladines, Julia Manetas, Karen Jinneman Salmonella serotype prediction using the GalaxyTrakr SeqSero2 workflow https://dx.doi.org/10.17504/protocols.io.bybfpsjn