



Dec 27, 2021

# Wastewater QC workflow in GalaxyTrakr (SSQuAWK) V.1

Jasmine Amirzadegan<sup>1</sup>, Tunc Kayikcioglu<sup>1</sup>, hugh.rand<sup>1</sup>, Ruth Timme<sup>2</sup>, Maria Balkey<sup>1</sup>

<sup>1</sup>Center for Food Safety and Applied Nutrition, U.S. Food and Drug Administration, College Park, Maryland, USA;

<sup>2</sup>US Food and Drug Administration



protocol .

**GenomeTrakr**  
Tech. support email: [genomeTrakr@fda.hhs.gov](mailto:genomeTrakr@fda.hhs.gov)

Ruth Timme  
US Food and Drug Administration

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

## PURPOSE:

Step-by-step instructions for checking sequence quality for SARS-CoV-2 wastewater samples using **SSQuAWK: SARS - CoV - 2 Sequence Quality Assurance Workflow and Kontraption**. The SSQuAWK workflow, implemented in a custom Galaxy instance, will produce quality assessments for raw reads (Illumina MiSeq paired-end fastq files).

**SCOPE:** This protocol covers the following tasks:

1. Set up an account in GalaxyTrakr
2. Create a new history
3. Upload data
4. Execute the SSQuAWK workflow
5. Interpret the results

<https://galaxytrakr.org>

Jasmine Amirzadegan, Tunc Kayikcioglu, hugh.rand , Ruth Timme, Maria Balkey 2021. Wastewater QC workflow in GalaxyTrakr (SSQuAWK). **protocols.io**  
<https://protocols.io/view/wastewater-qc-workflow-in-galaxytrakr-ssquawk-b2h4qb8w>

WGS, Quality Control, GalaxyTrakr, GenomeTrakr, microbial pathogen surveillance

protocol ,

Dec 01, 2021

Dec 27, 2021

55580

:

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

Account set up

1. Create a GalaxyTrakr account here: <https://account.galaxytrakr.org/Account/Register>

**User Registration Form**

Location:

First Name:

Last Name:

Email:

Primary Phone:

Title:

Requirements:

## 1.1 Log into your GalaxyTrakr account: <https://galaxytrakr.org>

**Galaxy / GalaxyTrakr 1905** Analysis Data Workflow Visualize Shared Data Help Login

Welcome to Galaxy, please log in

Username or Email Address:

Password:

Forgot password? Click here to reset your password.

Don't have an account? Registration for this Galaxy instance is disabled. Please contact an administrator for assistance.

**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 re-import the skesamist workflow that was updated a few days ago. Previous versions are no longer working and are causing errors when running. Thank you.

Access CFSAH 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

Forgot Password? Email GalaxyTrakr Support Team

## Create a new history

### 2 Create a new history.

We recommend creating a new history for each new MiSeq Run and including the flow-cell ID and date in the history name.

Save your MicroRunQC output here and any other relevant analyses, like serotyping, or AMR detection.

After all the analysis output from this run is saved to your internal data network or computer, older history's should be purged/deleted so as not to occupy the limited storage space in your account. In some cases it may be useful to save, for a limited time, multiple histories or to run analyses concurrently in multiple histories. In these cases you need to pay attention to your % usage bar (shows % used of allocated storage space) in the upper right corner of the GalaxyTrakr page. If you need additional space you can contact [galaxytrakrsupport@fda.hhs.gov](mailto:galaxytrakrsupport@fda.hhs.gov) and request additional storage.

### 2.1 Create a new history with the "+" symbol in the upper right hand corner. Name your history and press "enter" on your keyboard to save the name.

**GalaxyTrakr** Analysis Data Workflow Visualize Shared Data Help User

Tools:

Get Data:

NGS: QC and manipulation

NGS: Screening and Prediction

NGS: Mapping

NGS: Assembly

NGS: Phylogenetics

NGS: CFSAH SNP Pipeline (Beta)

NGS: Mapblast

NGS: Nanopore

NGS: NCBI Blast+

NGS: RNA seq

NGS: Annotations

NGS: Virus

NGS: Ikonu

NGS: empty

NGS: Seqtk

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

History:

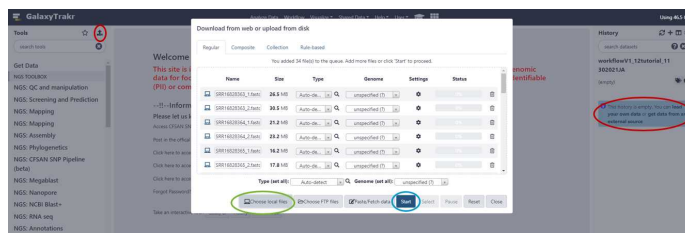
History:

History:

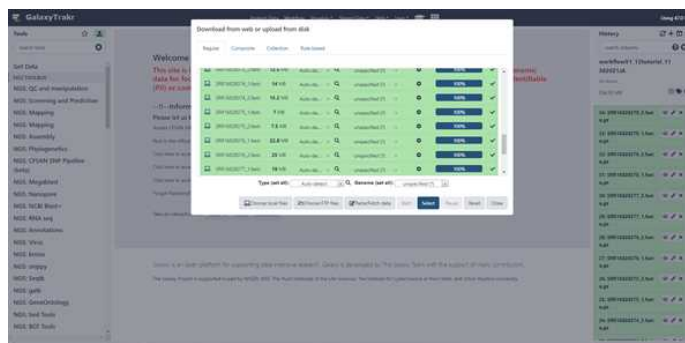
History:

History: <

After you've selected your files, press "Start" to initiate your data upload to GalaxyTrakr. The "Start" button is circled in blue.

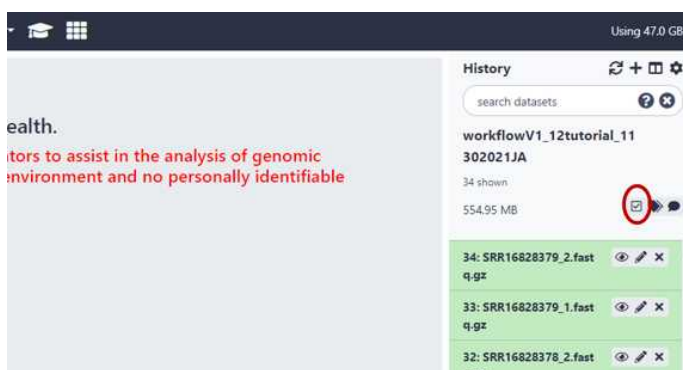


- 3.2 As the file uploads complete, each row will turn green. If samples are shown with yellow background, then are still uploading.



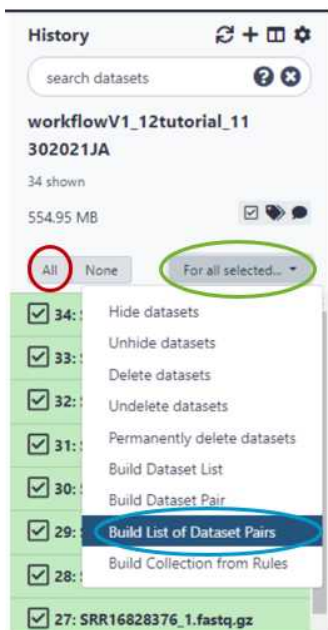
- 3.3 You have just upload a set of forward and reverse reads. For further analysis these files need to be paired properly so the platform knows which R1 and R2 files go with each sample. GalaxyTrakr does this by creating a **List of Dataset Pairs**.

Within your newly created History panel, click the "check box," then select all the files you just uploaded by clicking "All" or by individually selecting the ones you want to pair.



- 3.4 Check all the files belonging to a pair. In this example, all the files belong to a pair, so I will use the "All" button (circled in red).

Then, use the "For all selected..." dropdown (circled in green), and click on "Build List of Dataset Pairs" (circled in blue).



3.5 GalaxyTrkr will automatically pair the files, but it's good to double check.

Paired reads will pair in the middle column and turn green.

If everything looks good, then choose a name for your pairs (circled red) and "Create List" (also circled red).

Create a collection of paired datasets

17 pairs created: all datasets have been successfully paired

0 unpaired forward - (0 filtered out) Choose filters Clear filters 0 unpaired reverse - (0 filtered out)

17 paired Unpair all

SRR16828363_1.fastq.gz	→	SRR16828363.fastq.gz	←	SRR16828363_2.fastq.gz	⊗
SRR16828364_1.fastq.gz	→	SRR16828364.fastq.gz	←	SRR16828364_2.fastq.gz	⊗
SRR16828365_1.fastq.gz	→	SRR16828365.fastq.gz	←	SRR16828365_2.fastq.gz	⊗
SRR16828366_1.fastq.gz	→	SRR16828366.fastq.gz	←	SRR16828366_2.fastq.gz	⊗
SRR16828367_1.fastq.gz	→	SRR16828367.fastq.gz	←	SRR16828367_2.fastq.gz	⊗
SRR16828368_1.fastq.gz	→	SRR16828368.fastq.gz	←	SRR16828368_2.fastq.gz	⊗
SRR16828369_1.fastq.gz	→	SRR16828369.fastq.gz	←	SRR16828369_2.fastq.gz	⊗
SRR16828370_1.fastq.gz	→	SRR16828370.fastq.gz	←	SRR16828370_2.fastq.gz	⊗
SRR16828371_1.fastq.gz	→	SRR16828371.fastq.gz	←	SRR16828371_2.fastq.gz	⊗
SRR16828372_1.fastq.gz	→	SRR16828372.fastq.gz	←	SRR16828372_2.fastq.gz	⊗
SRR16828373_1.fastq.gz	→	SRR16828373.fastq.gz	←	SRR16828373_2.fastq.gz	⊗
SRR16828374_1.fastq.gz	→	SRR16828374.fastq.gz	←	SRR16828374_2.fastq.gz	⊗
SRR16828375_1.fastq.gz	→	SRR16828375.fastq.gz	←	SRR16828375_2.fastq.gz	⊗
SRR16828376_1.fastq.gz	→	SRR16828376.fastq.gz	←	SRR16828376_2.fastq.gz	⊗

Remove file extensions from pair names? ☐ Hide original elements? ☐

Name: Pairs\_11302021JA

Cancel Create list

Alternatively, instead of auto-pairing you can click "choose filters" and select the appropriate filter for the pairing:

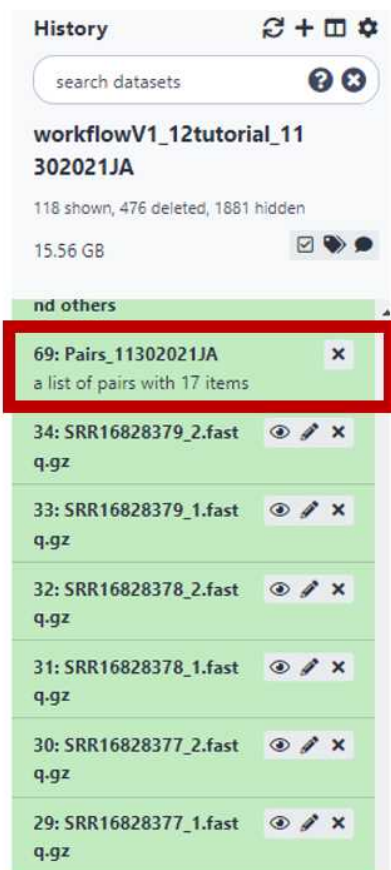
0 unpaired forward - (4 filtered out) Choose filters Clear filters 0 unpaired reverse - (4 filtered out)

Choose from the following filters to change which unpaired reads are shown in the display:

Forward: \_1, Reverse: \_2

Forward: \_R1, Reverse: \_R2

- 3.6 This paired dataset will now be available for analysis in your history panel. You can run multiple analyses on the same dataset in a history rather than upload the same sequence data to a new history to perform additional analyses. This will help you use your allocated storage space efficiently.

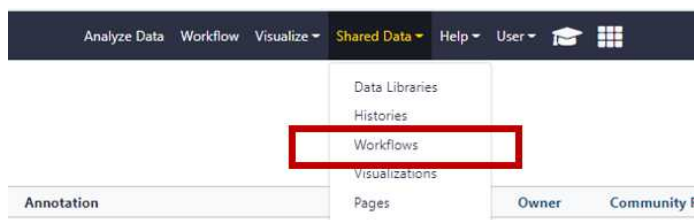


Run the <> workflow

- 4 Add the SSQuAWK\* workflow to your own "workflows" panel. You only have to do this step once for each new workflow you need.

\*SSQuAWK: SARS - CoV - 2 Sequence Quality Assurance Workflow and Kontraption

- 4.1 Navigate to the "Shared Data" drop down and choose workflows



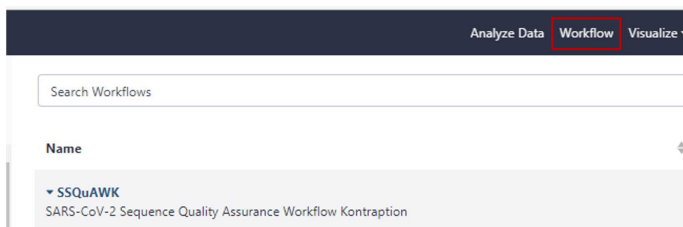
Then, from the SSQuAWK drop down menu, select import.

Published Workflows

search name, annotation, owner, or  
Advanced Search

Name	Annotation	Owner	Community Rating	Community Tags	Last Updated
SSQuAWK	SARS-CoV-2 Sequence Quality Assurance Workflow Kontrapion	jarmine_ami	★★★★★		3 seconds ago
NARMS: Unknown or Mixed Run-AMR Workflow V2.0	Not log specific. For mixed NISeg runs or unknown isolates	gmarin	★★★★★	amr, amr, NISeg	Oct 22, 2021
AMRfinderPlusGT Report WF		gmarin	★★★★★		Oct 22, 2021
NARMS: E. coli AMR Workflow V2.0	E. coli AMR, speciation, and QC	gmarin	★★★★★	amr, amr, Ecol	Oct 24, 2021

- 4.2 Navigate to the "Workflow" tab in the top ribbon (boxed in red). The workflow will be imported there.

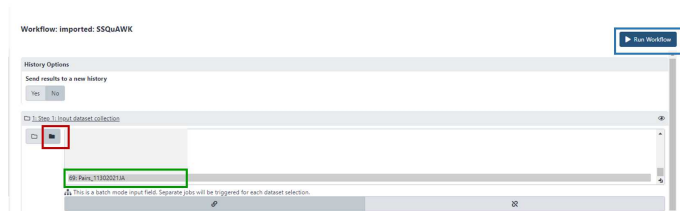


- 4.3 To use the workflow, press the 'play' button (boxed in red) on the right

Name	Tags	Updated	Sharing	Bookmarked	
Imported: SSQuAWK SARS-CoV-2 Sequence Quality Assurance Workflow Kontrapion		a few seconds ago			

- 4.4 Select the paired list you created earlier by selecting the folder icon (boxed in red), and then the list of pairs (boxed in green).


Click Run Workflow (boxed in blue).



Running the workflow can take some time depending on the number of samples you are analyzing. Once GalaxyTrakr adds the workflow invocation to the queue, you can choose to log out of GalaxyTrakr and log back in at a later time to see if the job is completed.



- 4.5 Upon completion of the pipeline, the NGSQC\_outfile will be green. Click on the "Eye" icon to view the output in the GalaxyTrakr window.

Analyze Data													Workflow	Visualize	Shared Data	Help	+		Using 100 T
1	2	3	4	5	6	7	8	9	10	11	12	13							
Sample	fastq1	fastq2	fastq3	fastq4	fastq5	fastq6	fastq7	fastq8	fastq9	fastq10	fastq11	fastq12	fastq13						
SRX1432351.fastq																			
SRX1432364.fastq	226058	755	3575	226058	755	3481	0.18	80.1	20.82	47620	2845	30751							
SRX1432355.fastq	175990	755	3430	175990	755	3579	0.50	874	30.04	52882	1579	27282							

History

search datasets

quickstart 2 Dec 2021

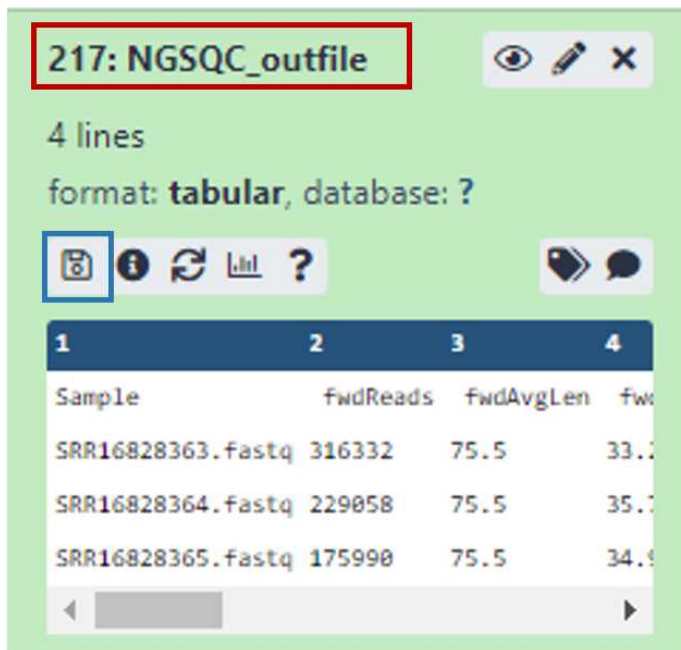
18 items, 257 views

637.87 MB

123 NEW! workflows

## 5 Download and interpret the results:

- 5.1 Click "NGSQC\_outfile" (boxed in red) and then the floppy disc save icon (boxed in blue). The tabular file can be opened in a text reader or converted to a format that can be opened on excel.



- 5.2 The SSQuAWK output file includes the following metrics:

A	B	C
Parameter	Input	Description
Sample	List of Pairs	Sample name from list of pairs
fwdReads	FASTQC	Number of forward reads contributing to the sample pair
fwdAvgLen	FASTQC	Average of all forward read lengths
fwdAvgQ	FASTQC	Average quality of all forward reads
revReads	FASTQC	Number of reverse reads contributing to the sample pair
revAvgLen	FASTQC	Average of all reverse read lengths
revAvgQ	FASTQC	Average quality of all reverse reads
percentHuman	Kraken2	Percentage of reads classified as <i>Homo sapiens</i>
readsHuman	Kraken2	Number of reads classified as <i>Homo sapiens</i>
percentSyntheticSeqs	Kraken2	Percentage of reads classified as non - biological sequences
readsSyntheticSeqs	Kraken2	Number of reads classified as non - biological sequences
percentCovid	Kraken2	Percentage of reads classified as SARS - CoV - 2
readsCovid	Kraken2	Number of reads classified as SARS - CoV - 2

- 5.3 Example output for 3 pairs run through the SSQuAWK workflow:

A	B	C	D	E	F	G	H	I	J	K	
Sample	fwdReads	fwdAvgLen	fwdAvgQ	revReads	revAvgLen	revAvgQ	percentHuman	readsHuman	percentSyntheticSeqs	readsSyntheticSeqs	perc
SRR16828363.fastq.qz	316332	75.5	33.21	316332	75.5	31.76	0.48	1517	70.88	224206	2
SRR16828364.fastq.qz	229058	75.5	35.71	229058	75.5	34.81	0.38	863	20.92	47920	3
SRR16828365.fastq.qz	175990	75.5	34.9	175990	75.5	33.79	0.5	874	30.04	52862	4