Jun 27, 2025 Version 1

Quality control and preliminary data analysis of Sequencing Data: GalaxyTrakr ONT workflow V.1

DOI
https://dx.doi.org/10.17504/protocols.io.6qpvrq54plmk/v1

Narjol Gonzalez-Escalona[1], Maria Hoffmann[2]

[1]FDA; [2]US FDA

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

Narjol Gonzalez-Escalona
FDA

OPEN ACCESS

DOI: https://dx.doi.org/10.17504/protocols.io.6qpvrq54plmk/v1

Protocol Citation: Narjol Gonzalez-Escalona, Maria Hoffmann 2025. Quality control and preliminary data analysis of Sequencing Data: GalaxyTrakr ONT workflow. **protocols.io** https://dx.doi.org/10.17504/protocols.io.6qpvrq54plmk/v1

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

Created: May 27, 2025

Last Modified: June 27, 2025

Protocol Integer ID: 219010

Keywords: ONT, sequencing, long-read, Oxford Nanopore , R10.4.1, microbial genome assembly, AMR prediction, preliminary data analysis of sequencing data, draft de novo assembly, sequencing data, ont assembly workflow, galaxytrakr ont, preliminary sequence analysis for foodborne pathogen, ont workflow, galaxy trakr, galaxy trakr account, wgs sequence quality, account in galaxytrakr, galaxytrakr, custom galaxy instance, sequence type for each isolate, preliminary sequence analysis, workflow, most microbial pathogen, checking oxford nanopore technology, laboratory, de novo

Funders Acknowledgements:
FDA Foods Programme

Abstract

Step-by-step instructions for checking Oxford Nanopore Technologies (ONT) WGS sequence quality and preliminary sequence analysis for foodborne pathogens. The ONT workflow, implemented in a custom Galaxy instance, will produce draft *de novo* assemblies, along with reporting several characteristics such as: sequence type for each isolate, AMR profile, and serotype prediction in the case of Salmonella. This workflow will work on most microbial pathogens, so we advise laboratories to upload their entire ONT run through this workflow.

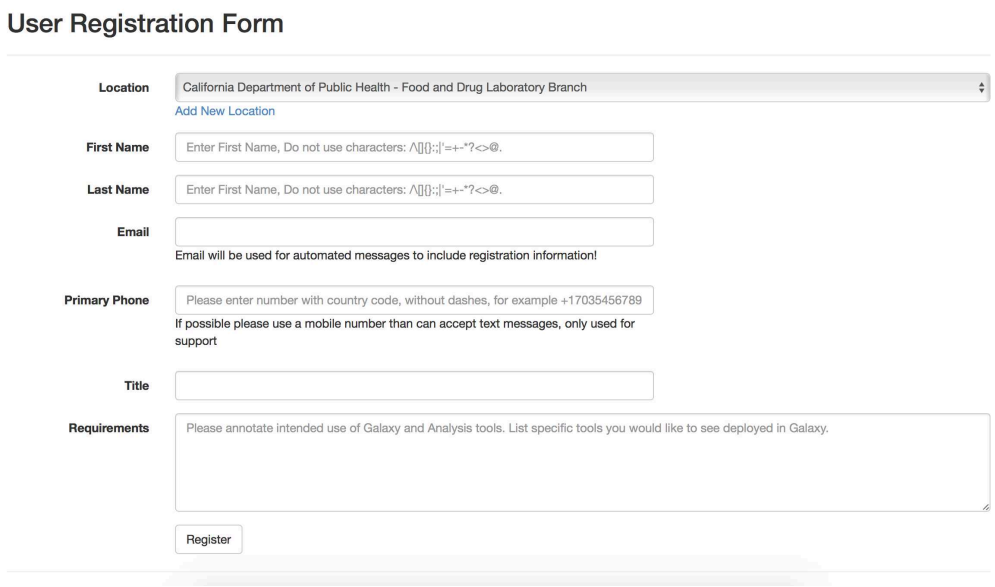
SCOPE: This protocol covers the following tasks:

1. set up an account in GalaxyTrakr (if not having one already)
2. Create a new history/workspace
3. Upload data
4. Execute the ONT assembly workflow
5. Interpret the results

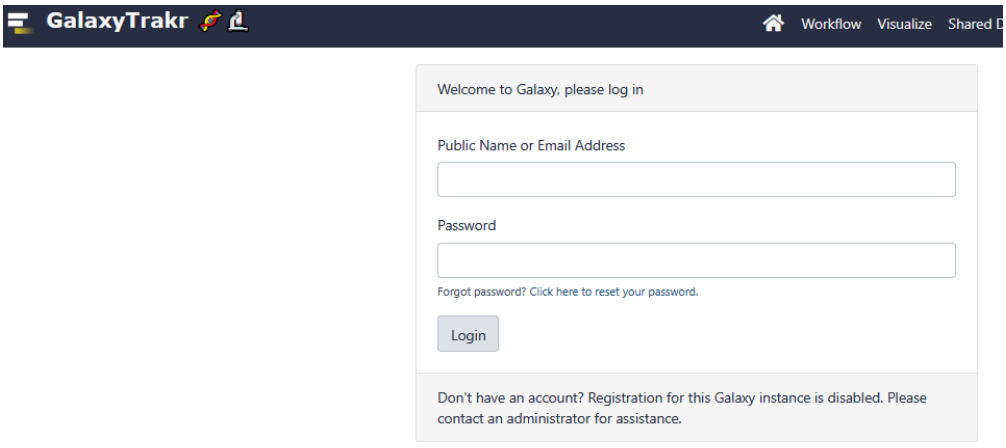
The first section "Account set up" it is applicable only to members that do not have a galaxy trakr account. If you do, please proceed to section 2.

Account set up

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



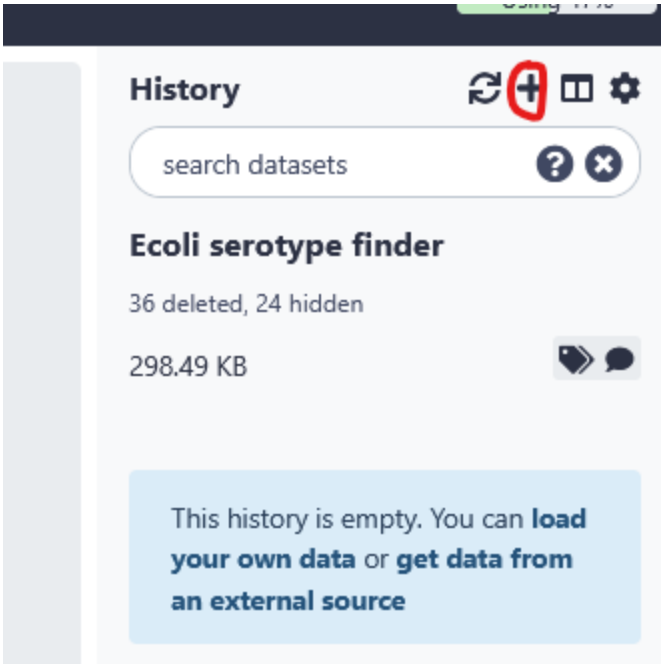
1.1 Log into your GalaxyTrakr account: **https://galaxytrakr.org**



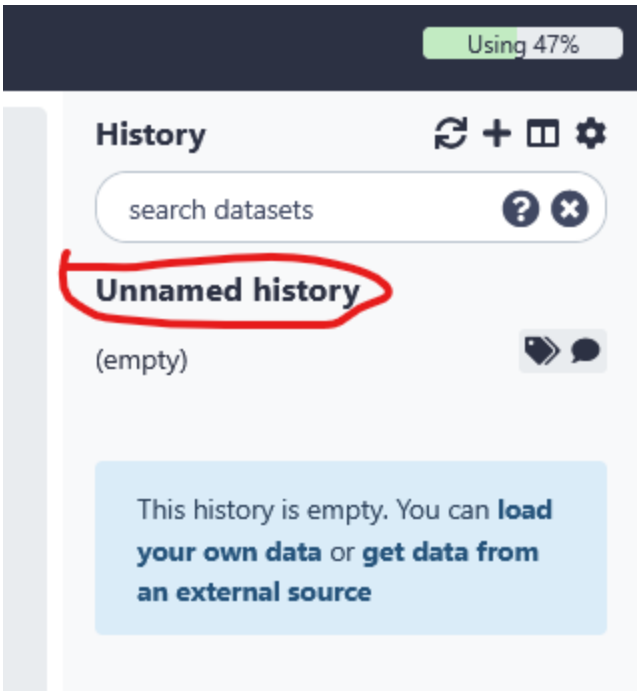
Create a new history

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

2.1 Click on the + icon in the upper right History panel



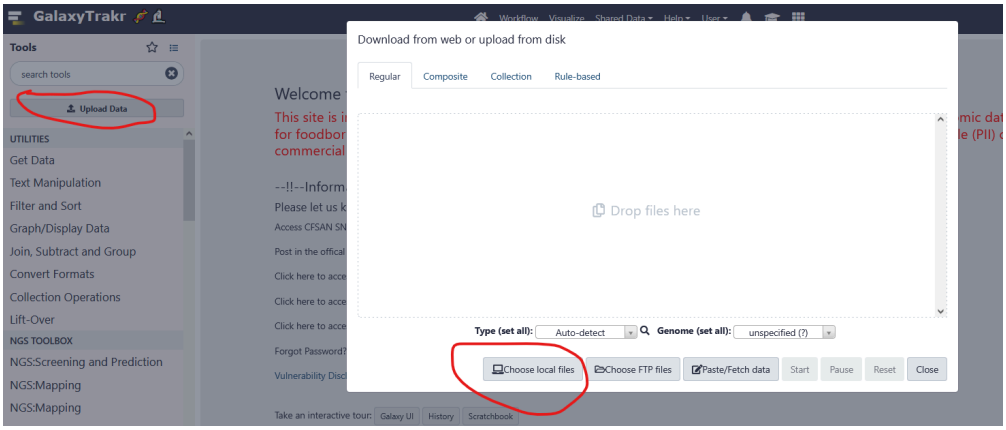
2.2 Name your new History by clicking on the “Unnamed history”, type in desired name and hit enter. We recommend including the flow cell ID and the date the run was started.



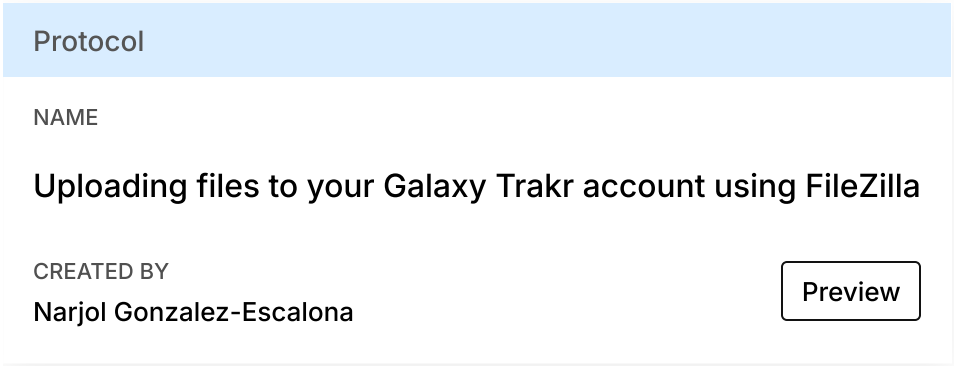
Upload data

3 This section will describe the process for uploading raw fastq files into your active History panel. After the files have been uploaded they will stay in your account until they are deleted.

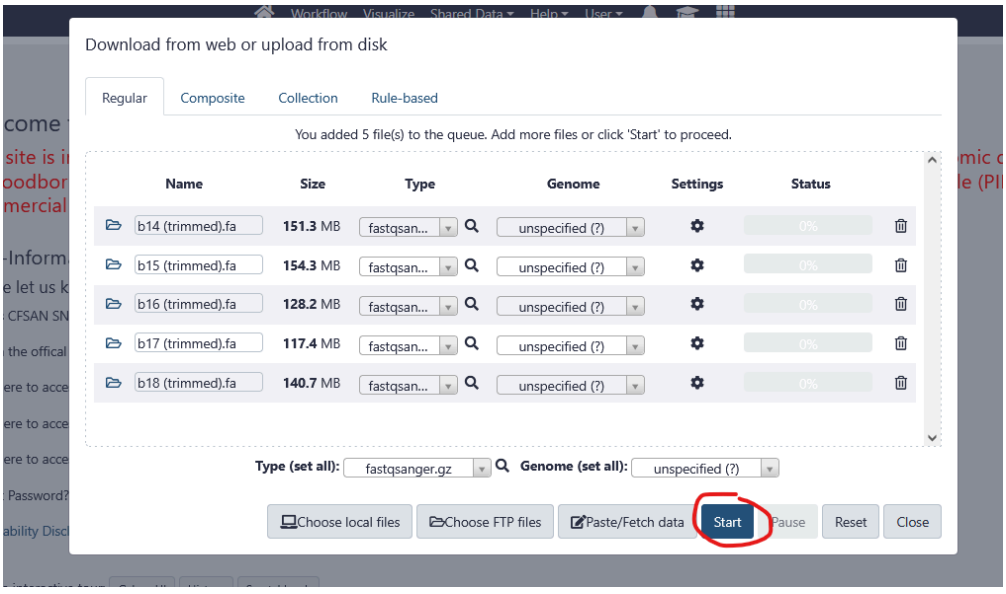
3.1 Click on the upload icon on the top of the left web page to start an upload process.



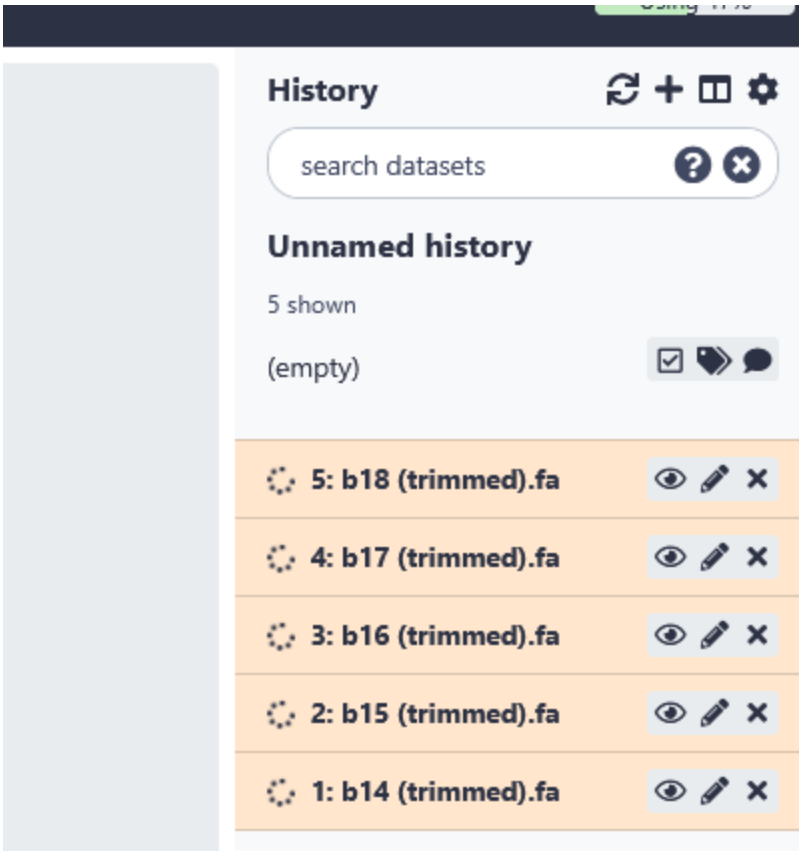
3.2 Alternative you can use FileZilla to upload your data to the Galaxy Trakr ftp and retrieve it from there as described below:

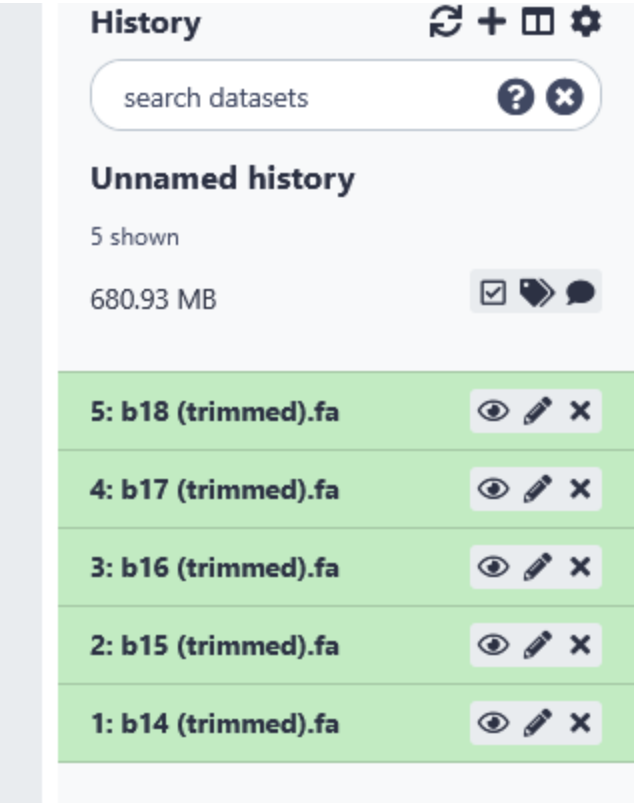


3.3 Select "Type (set all):fastqsanger.gz." Choose local file button and navigate to the desired fastq files, then click "start" to upload files. These files should be one single fastq file per sample/isolate).



As the file uploads complete, each row will turn green. Samples in yellow are still in process.

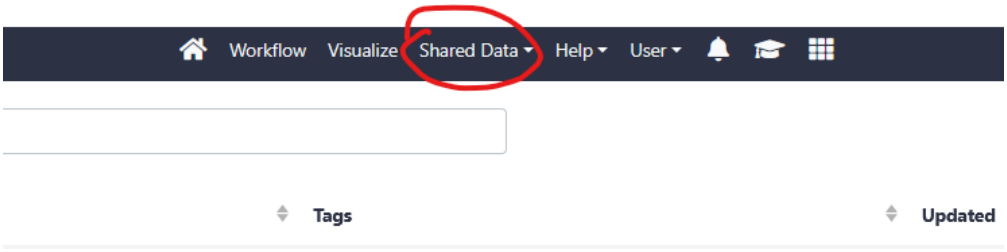




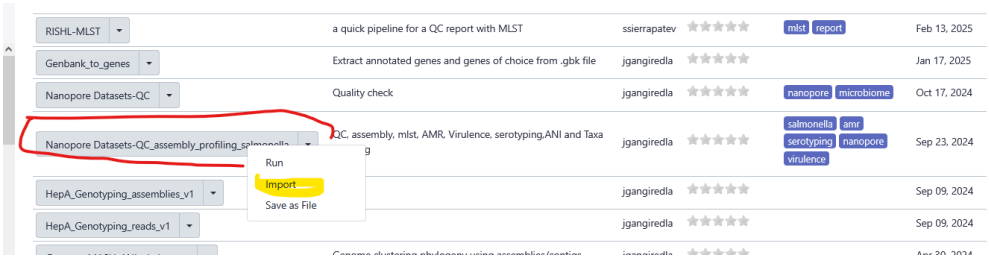
Execute the ONT assembly workflow

4 Add the ONT (**Nanopore Datasets-QC_assembly_profiling_salmonella)** workflow to your own "workflows" panel. You only have to do this step once for each new workflow you need.

4.1 Navigate to the “Shared Data" drop down menu, choose workflows and from the **Nanopore Datasets-QC_assembly_profiling_salmonella** drop down menu select import.

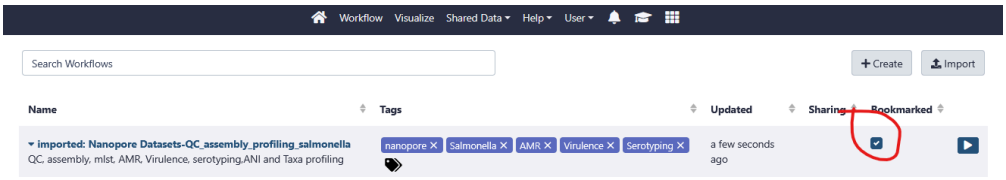


click on shared data and select "workflows"



click on the drop menu and select **import**

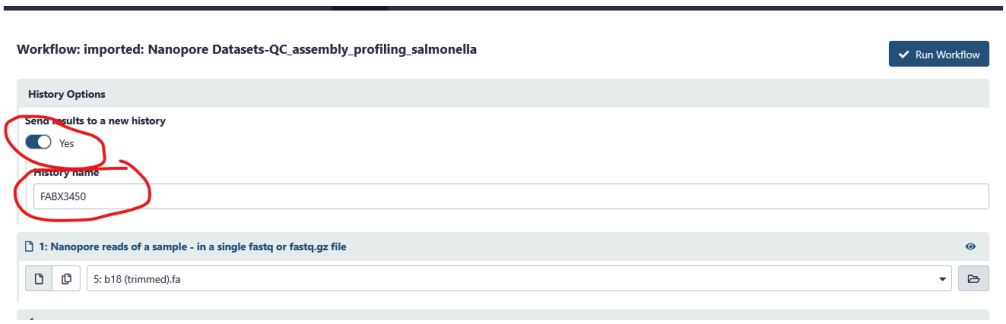
4.2 To see the new workflow in your “WORKFLOWS” tools panel on the left, open the Workflow tab and check “show in tools panel” for the workflow of interest.



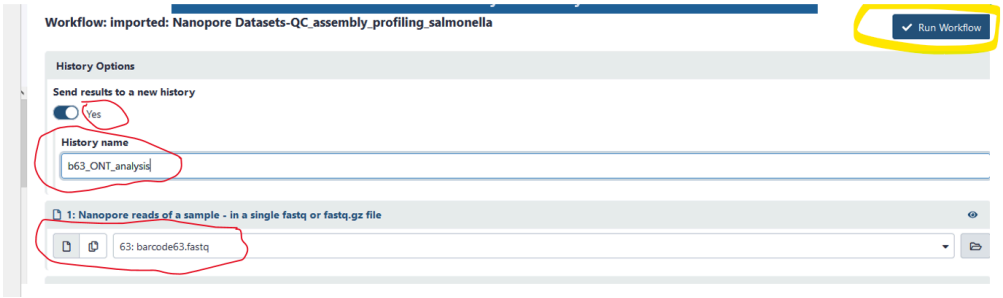
4.3 From the workflow menu select **Nanopore Datasets-QC_assembly_profiling_salmonella.** You can added to your favorites and this will make it easier to find it next time you need to use it.

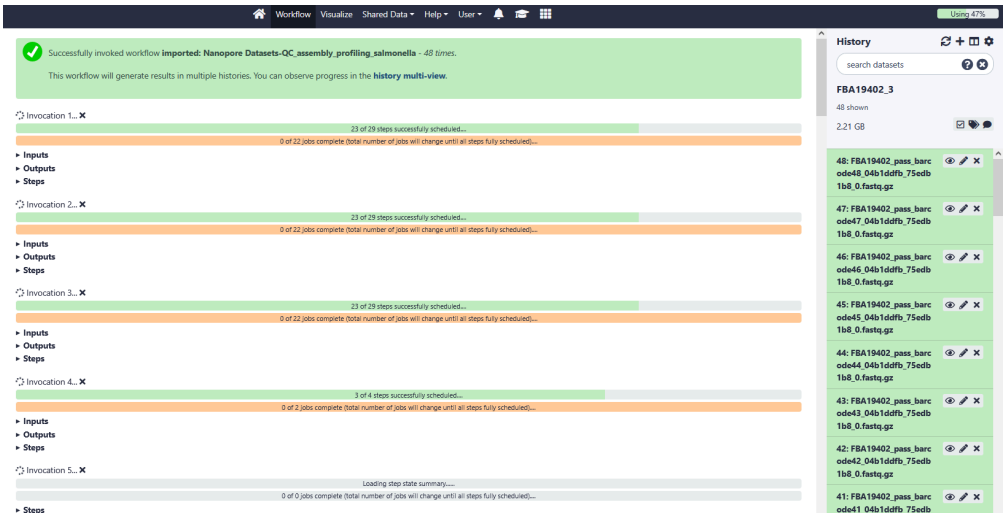
4.4 In the history options select send results to a new history "yes". Then you can change the history name . We suggest to call it as the flow cell ID. This will run each selected

sample/fastq in a separate history.



4.5 Select your fastq files that you uploaded previously. Using the multiple datasets button. Then click the Run workflow button. This can take some time depending on the number of samples you are analyzing. If you choose to you can log out of GalaxyTrakr and log back in at a later time to see if the job is completed. A new history will be created for each sample.

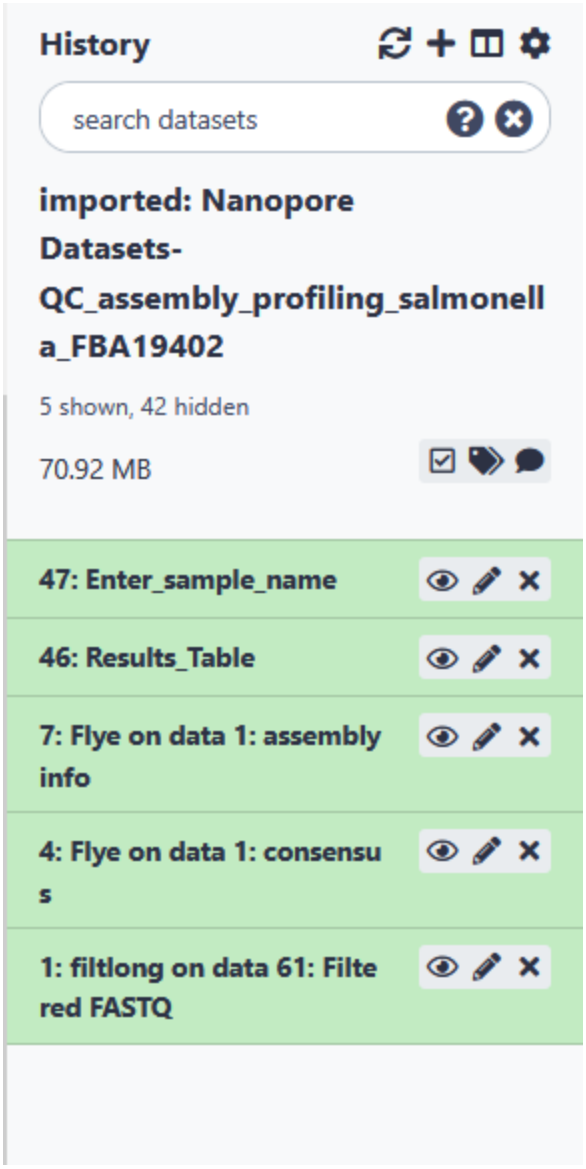




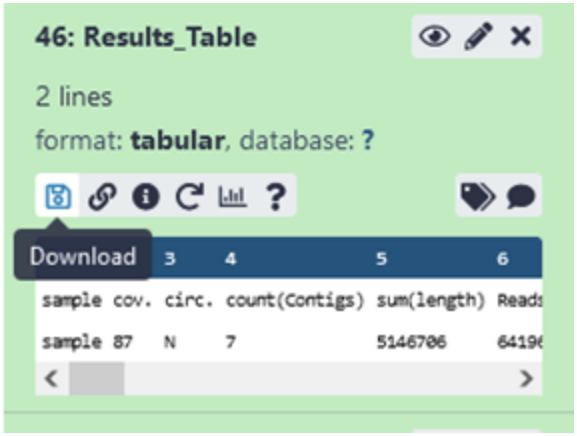
4.6 Upon completion of the pipeline all tiles in the new created history bar will be green.

Collate all results and interpretation

5 **Download and interpret the results:** After the end of the workflow you will have 5 files. Ignore the sample name file. Results_table will contain all your in silico analysis results (download it). THe Flye..consensus will have your assembly (download it). The Flye .. assembly info, contains all your assembly information (download it as well).



5.1 To download the files click on the title on the file (e.g. Results_Table) and click the floppy disc icon.



5.2 The Results_table output file includes the following metrics:

	A	B
	sample	*Change this to your sample number or ID*

	A	B
	Reads	Number of reads
	GC%	GC% content
	Avg_Len	Average length of the reads or reads N50
	Mean_Q	Mean Quality of the reads
	Q10%	percentage of reads > Q10
	Taxa	Taxa identified from the reads, closest strain match
	ANI	ANI number compared to the cloest match
	AMR	AMR gene content
	STRESS	Stress gene identified
	ST	Sequence Type
	Predicted identification	Salmonella enterica subspecies ID
	Predicted antigenic profile	In the case of Salmonella, the antigenic profile
	Predicted serotype	Salmonella Serotype
	cov.	Mean coverage for contigs in the flye assembly.
	circ.	If the contigs are circular or not
	count(Contigs)	Number of contigs in the flye assembly
	sum(length)	Genome size (bp)

ONT pipeline analysis workflow lists the summary metrics for sequence quality, number of contigs, and estimated genome size, along with other common metrics for reads (Mean Length, Quality, and Taxa ANI ID). Additionally, if the Multi-Locus Sequence Type (MLST) for the isolate is available from pubmlst, the workflow also reports Sequence Type (ST), AMR and Stress genes, and in the case of Salmonella will also report Predicted ID, Predicted antigenic profile, and predicted serotype. If no Salmonella, then those fields will left empty. Later iterations of this workflow will address other foodborne pathogens specific serotypes or ID as available.

**This output should be saved either to your LIMS or to a spreadsheet linked to the sequencing run and samples.

5.3 Due to the nature of Galaxy Trakr all individual Results_Tables for each sample should be collated into a single Table for easy interpretation. Example output for 8 samples run through the ONT assembly workflow:

	CFSAN#	Barcode	Reads	GC%	Avg_Len	Mean_Q	Q10%	Taxa	ANI	AMR	STRESS	VIRULENCE	ST	Predicted identification	Predicted antigenic profile	Predicted serotype	cov.	circ.	count(Contigs)	sum(length)
	CFSAN000318	48	397770	51	4233.6	35.99	100	Salmonella enterica subsp. enterica serovar Heid	99.998	aac(3)-IId,aadA1,blaTEM-1,fosA7,mdsA,mds	arsR,golS,golT	iroB,iroC,sinH,sodC1	15	Salmonella enterica subspecies enterica (subs	4:r:1,2	Heidelberg	32	Y	4	4957870

								elbergstr. 41578		B,tet(A)				peciesl)						
	CFSAN008585	47	616645	51	47227.2	35.71	100	Salmonellaenterica subsp.entericaserovarDerbystr.6266	99.837	aac(3)-VIa,aadA1,fosA7.3,mdsA,mdsB,sul1,tet(A)	golS,golT,pcoA,pcoB,pcoC,pcoD,pcoE,pcoR,pcoS,qacEdelta1,silA,silB,silC,silE,sil	iroB,iroC,sinH	40	Salmonellaenterica subspeciesenterica(subspeciesI)	4:f,g:-	Derby	47	Y	4	5086410

											F,silP,silR,silS									
	CFSAN051458	46	735568	50	58855.7	36.17	100	Escherichia coli O121 str. RM8352	99.97	acrF,blaEC,emrD,mdtM	arsC,arsR,asr,terD,terW,terZ,ymgB	eae,efa1,ehxA,espA,espB,espF,espI,espJ,espK,espP,espX1,fdeC,lpfA,nleA,n	655	-	-:-:-	-:-:-	79	N	7	5550579

	CFSAN068773	45	97883	56	49099	36.48	100	Cronobacter sakazakii ATCC BAA-894	99.81	blaCSA	fieF,pcoA,pcoB,pcoC,pcoD,pcoE,pcoR,pcoS,silA,silB,silC,sil	leB,nleC,stxA2c,stxB2a,tccP,tir,toxB	1	-	---:	-:-:-	99	Y	4	45817775

											E,silF,silP,silR,silS,terD,terW,terZ									
	CFSAN086180	44	673352	57	57551.5	36.68	100	Klebsiella sp. X1-16S-Nf21	99.8111	blaLEN-16,emrD,fosA,kdeA,oqxA,oqxB	fieF		-	-	-:-:-	-:-:-	69	Y	1	55364144
	CFSAN086182	43	498556	51	5250	36.36	100	Citrobacter sp. KTE32	98.6999	blaCMY	arsA,arsB,arsC,arsD,arsR		-	-	-:-:-	-:-:-	52	Y	1	49174486

	CFSAN086183	42	497555	55	99044.4	37.05	100	Enterobacter cancerogenus ATCC 35316	98.773	ampC,fosA	fieF	iroB,iroC,iroN	-	-	-:-:-	-:-:-	100	Y	2	49923559
	CFSAN122995	41	882297	50	57433.4	36.25	100	Salmonella enterica subsp. diarizonae serovar 50:k:	99.285			cdtB,iroB,iroC,iucA,iutA,ybtP,ybtQ	645	Salmonella enterica subspecies diarizonae (subs	6,14:z10:z	IIIb(6),14:z10:z	92	Y	1	53204223

								zstr.MZ00800						peciesIIIb)						