

Aug 14, 2025

Version 3

# Quality control and preliminary data analysis of Sequencing Data: GalaxyTrakr ONT workflow V.3

DOI

<https://dx.doi.org/10.17504/protocols.io.6qpvrq54plmk/v3>

Narjol Gonzalez-Escalona<sup>1</sup>, Maria Hoffmann<sup>1</sup>, Jayanthi Gangiredla<sup>1</sup>

<sup>1</sup>US FDA

GenomeTrakr

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



Narjol Gonzalez-Escalona

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.6qpvrq54plmk/v3>

**Protocol Citation:** Narjol Gonzalez-Escalona, Maria Hoffmann, Jayanthi Gangiredla 2025. Quality control and preliminary data analysis of Sequencing Data: GalaxyTrakr ONT workflow. **protocols.io**

<https://dx.doi.org/10.17504/protocols.io.6qpvrq54plmk/v3> Version created by [Narjol Gonzalez-Escalona](#)



**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:** August 08, 2025

**Last Modified:** August 14, 2025

**Protocol Integer ID:** 224356

**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, galaxytrkr ont, preliminary sequence analysis for foodborne pathogen, ont workflow, galaxy trkr, galaxy trkr account, wgs sequence quality, account in galaxytrkr, galaxytrkr, custom galaxy instance, sequence type for each isolate, preliminary sequence analysis, workflow, most microbial pathogen, checking oxford nanopore technology, laboratory

**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 trkr account. If you do, please proceed to section 2.



## Account set up

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

### User Registration Form

Location

California Department of Public Health - Food and Drug Laboratory Branch

[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

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

GalaxyTrakr

Welcome to Galaxy, please log in

Public Name or Email Address

Password

Forgot password? [Click here to reset your password.](#)

Login

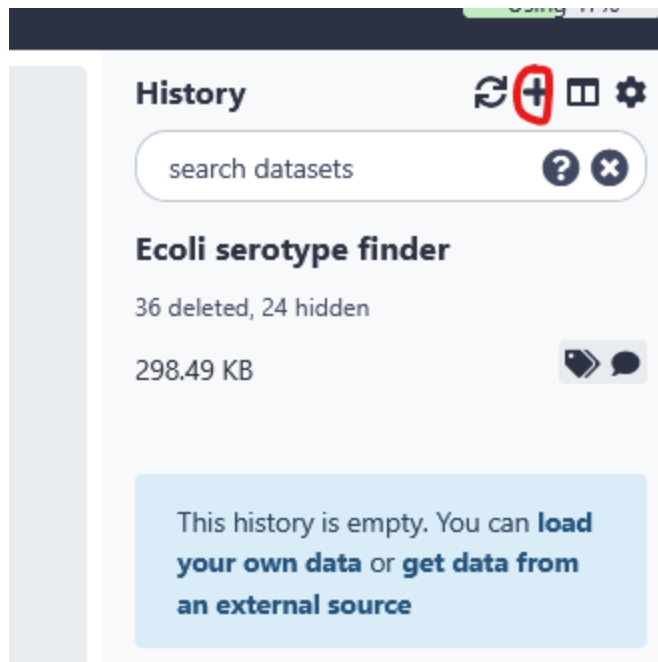
Don't have an account? Registration for this Galaxy instance is disabled. Please contact an administrator for assistance.



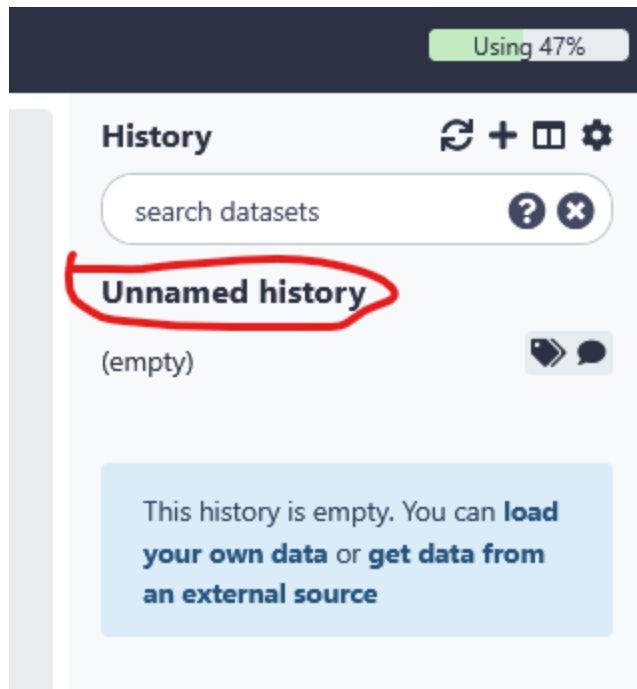
## 1.2

### 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

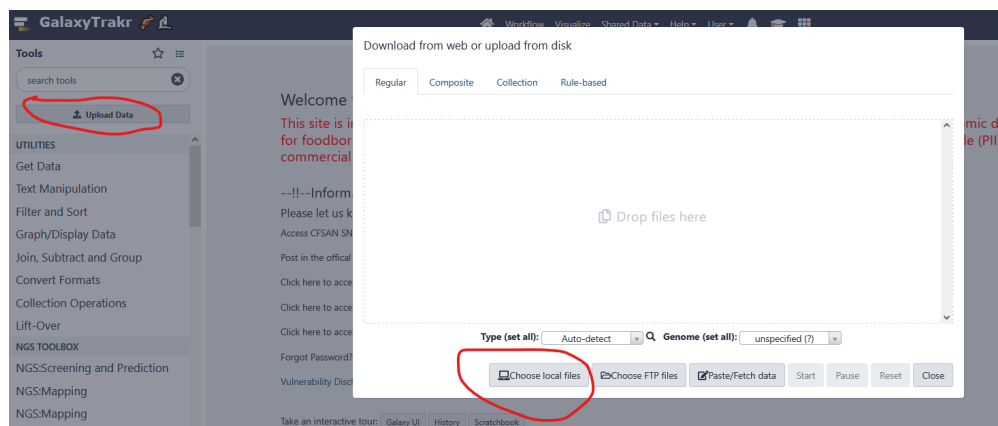


- 2.2 Name your new History by clicking on the "Unnamed history", type in desired name and hit enter. We recommend including the flow cell ID and the date the run was started.



## 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:

Protocol

NAME

Uploading files to your Galaxy Trakr account using FileZilla

CREATED BY

Narjol Gonzalez-Escalona

Preview

- 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).

Download from web or upload from disk

Regular Composite Collection Rule-based

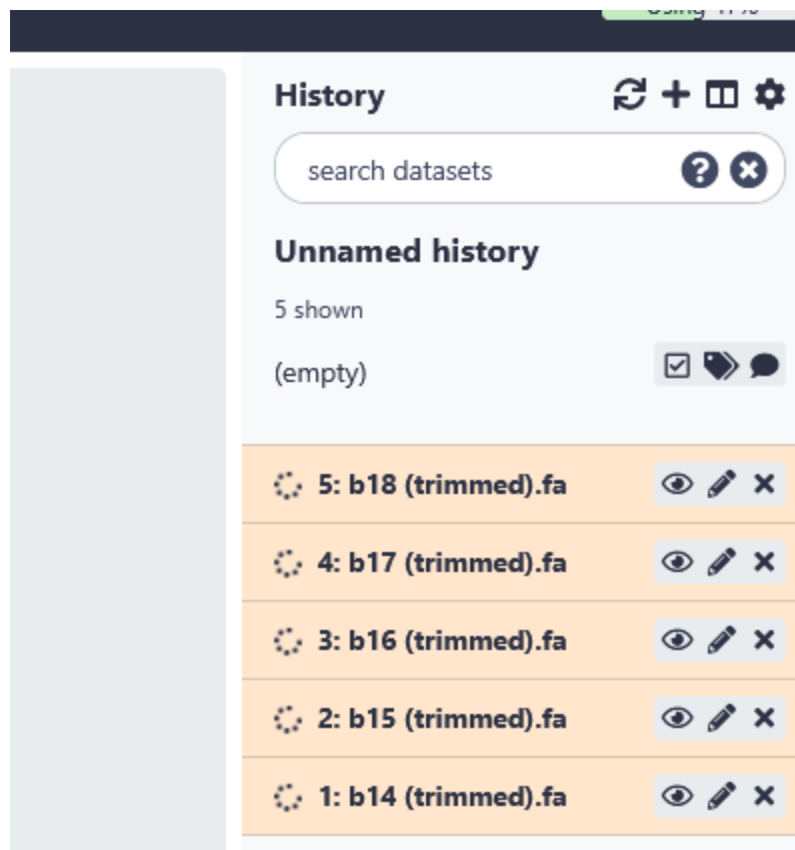
You added 5 file(s) to the queue. Add more files or click 'Start' to proceed.

| Name             | Size     | Type        | Genome          | Settings | Status |
|------------------|----------|-------------|-----------------|----------|--------|
| b14 (trimmed).fa | 151.3 MB | fastqsan... | unspecified (?) |          | 0%     |
| b15 (trimmed).fa | 154.3 MB | fastqsan... | unspecified (?) |          | 0%     |
| b16 (trimmed).fa | 128.2 MB | fastqsan... | unspecified (?) |          | 0%     |
| b17 (trimmed).fa | 117.4 MB | fastqsan... | unspecified (?) |          | 0%     |
| b18 (trimmed).fa | 140.7 MB | fastqsan... | unspecified (?) |          | 0%     |

Type (set all): fastqsanger.gz Genome (set all): unspecified (?)

Choose local files Choose FTP files Paste/Fetch data Start Pause Reset Close

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



The screenshot shows the 'History' panel in the protocols.io interface. It features a search bar labeled 'search datasets' and a section titled 'Unnamed history' with '5 shown'. Below this, there is a list of five datasets, each in a yellow state, indicating they are still in process. Each dataset entry includes a circular progress indicator, a label, and three action icons (eye, pencil, and X).

| Dataset Label       | Status              | Actions            |
|---------------------|---------------------|--------------------|
| 5: b18 (trimmed).fa | Yellow (In Process) | View, Edit, Delete |
| 4: b17 (trimmed).fa | Yellow (In Process) | View, Edit, Delete |
| 3: b16 (trimmed).fa | Yellow (In Process) | View, Edit, Delete |
| 2: b15 (trimmed).fa | Yellow (In Process) | View, Edit, Delete |
| 1: b14 (trimmed).fa | Yellow (In Process) | View, Edit, Delete |

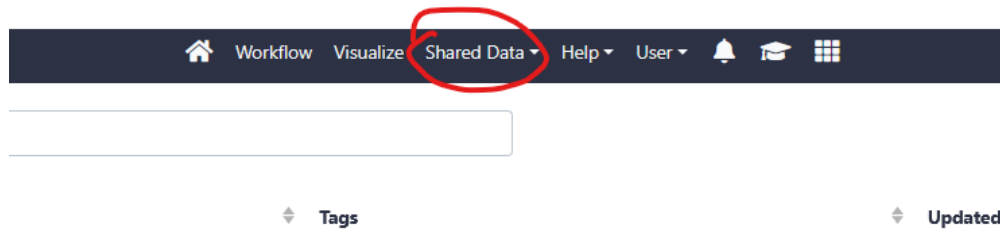


The screenshot shows the 'History' panel in the protocols.io interface. At the top, there are icons for refresh, add, view, and settings. Below is a search bar labeled 'search datasets' with a help icon and a close icon. The main section is titled 'Unnamed history' and indicates '5 shown' and '680.93 MB'. Below this is a list of five datasets, each with a green background and icons for view, edit, and delete. The datasets are:

- 5: b18 (trimmed).fa
- 4: b17 (trimmed).fa
- 3: b16 (trimmed).fa
- 2: b15 (trimmed).fa
- 1: b14 (trimmed).fa

## Execute the ONT assembly workflow

- 4 Add the ONT (**Nanopore Datasets-QC\_assembly\_genome\_profiling\_v1**) 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\_genome\_profiling\_v1** drop down menu select import.



click on shared data and select "workflows"

then search for "Nanopore"

#### Published Workflows

Nanopore    
Advanced Search

| Name                                                | Annotation                                                                                                | Owner     | Community Rating | Community Tags                               | Last Updated |
|-----------------------------------------------------|-----------------------------------------------------------------------------------------------------------|-----------|------------------|----------------------------------------------|--------------|
| Nanopore Datasets- QC_assembly_genome_profiling_v1  | QC-Assembly, MLST, AMR, Virulence, serotyping, ANI and Taxa profiling- works on raw sequences as an input | jgangedla | ★★★★★            | amr serotyping nanopore assembly qc          | Jul 16, 2025 |
| ONT-Assembly polishing with long reads              | Racon Assembly polishing with long reads, x4                                                              | jgangedla | ★★★★★            | nanopore assembly polishing                  | Jul 01, 2025 |
| Nanopore Datasets-QC                                | Quality check                                                                                             | jgangedla | ★★★★★            | nanopore microbiome                          | Oct 17, 2024 |
| Nanopore Datasets- QC_assembly_profiling_salmonella | QC. assembly, mlst, AMR, Virulence, serotyping, ANI and Taxa profiling                                    | jgangedla | ★★★★★            | salmonella amr serotyping nanopore virulence | Sep 23, 2024 |

#### Published Workflows

Nanopore    
Advanced Search

| Name                                                | Annotation                                                                                                | Owner     | Community Rating | Community Tags                               | Last Updated |
|-----------------------------------------------------|-----------------------------------------------------------------------------------------------------------|-----------|------------------|----------------------------------------------|--------------|
| Nanopore Datasets- QC_assembly_genome_profiling_v1  | QC-Assembly, MLST, AMR, Virulence, serotyping, ANI and Taxa profiling- works on raw sequences as an input | jgangedla | ★★★★★            | amr serotyping nanopore assembly qc          | Jul 16, 2025 |
| ONT-Assembly polishing with long reads              | Racon Assembly polishing with long reads, x4                                                              | jgangedla | ★★★★★            | nanopore assembly polishing                  | Jul 01, 2025 |
| Nanopore Datasets-QC                                | Quality check                                                                                             | jgangedla | ★★★★★            | nanopore microbiome                          | Oct 17, 2024 |
| Nanopore Datasets- QC_assembly_profiling_salmonella | QC. assembly, mlst, AMR, Virulence, serotyping, ANI and Taxa profiling                                    | jgangedla | ★★★★★            | salmonella amr serotyping nanopore virulence | Sep 23, 2024 |

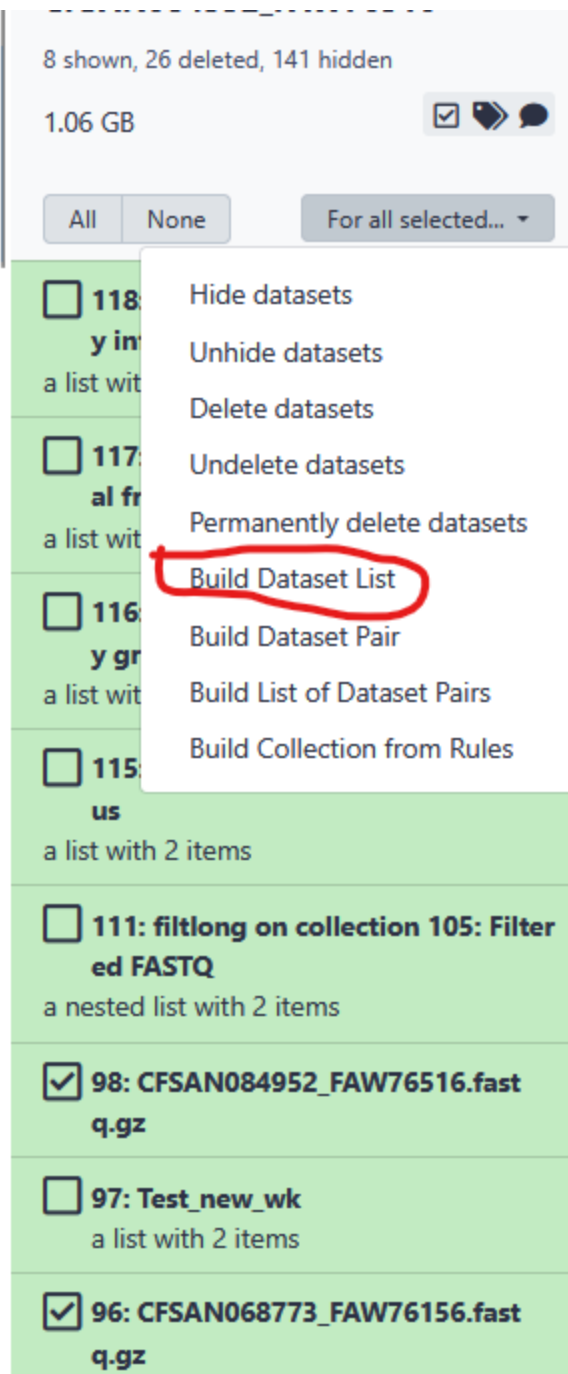
Then select import and you will have it in your workflows.

- 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.



| Name                                                                                                                                                                              | Tags   | Updated           | Sharing | Bookmarked                          |
|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|--------|-------------------|---------|-------------------------------------|
| ▼ imported: <b>Nanopore Datasets-QC_assembly_genome_profiling_v1</b><br>QC-Assembly, MLST, AMR, Virulence, serotyping, ANI and Taxa profiling- works on raw sequences as an input | 6 Tags | a few seconds ago |         | <input checked="" type="checkbox"/> |

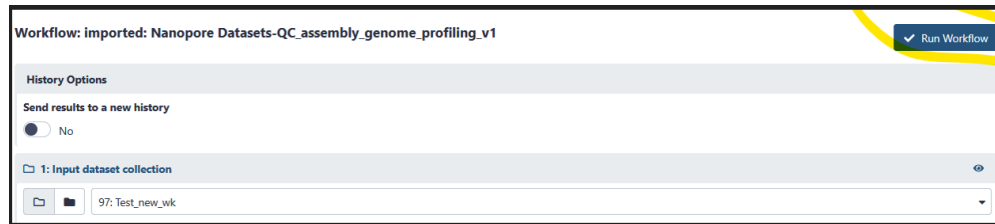
- 4.3 From the workflow menu select **Nanopore Datasets-QC\_assembly\_genome\_profiling\_v1**. You can added to your favorites and this will make it easier to find it next time you need to use it.
- 4.4 Create a new history and add your fastq files. Then select your fastq files and create a list of Datasets collection (Build Dataset List) with a name of your choosing.



- 4.5 Run the **Nanopore Datasets-QC\_assembly\_genome\_profiling\_v1** workflow using that dataset collection as input. 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.



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

## Results and interpretation

- 5 **Results:** After the end of the workflow you will have 6 extra files. The combined profiles result will be on the **Profiles-table** (download it). This Table will contain all your in silico analysis results for all your samples. **Flye on collection - consensus** will have all of the assemblies (download it). The **Flye on collection: assembly info**, contains all your assembly information (download it as well).



**History**

search datasets

**FAZ97012\_2**

7 shown, hide deleted, 319 hidden

20.46 GB

**326: Profiles-table**

**107: Flye on collection 86: assembly info**   
a list with 16 items

**106: Flye on collection 86: graphical fragment assembly**   
a list with 16 items

**105: Flye on collection 86: assembly graph**   
a list with 16 items

**104: Flye on collection 86: consensus**   
a list with 16 items

**86: filtlong on collection 52: Filtered FASTQ**   
a nested list with 16 items

**17: FAZ97012**   
a list with 16 items

- 5.1 To download the files click on the title on the file (e.g. Results\_Table) and click the floppy disc icon.

**FAZ97012\_2**  
7 shown, hide deleted, 319 hidden  
20.46 GB

326: Profiles-table

17 lines

format: **tabular**, database: ?

1

2

3

4

5

6

7

Sample

Reads

GC%

Avg\_Len

Mean\_Q

Q10%

Taxa

b64

180281

51

2226.7

32.49

99.71

Salmon

b63

230615

51

2673.8

33.88

99.92

Salmon

b62

138398

50

3813.6

33.89

99.78

Escher:

b61

265439

56

3245.8

32.26

99.77

Cronob:

5.2 The Results\_table output file includes the following metrics:

| A              | B                                                    |
|----------------|------------------------------------------------------|
| <b>sample</b>  | <i>Change this to your sample number or ID</i>       |
| <b>Reads</b>   | Number of reads                                      |
| <b>GC%</b>     | GC% content                                          |
| <b>Avg_Len</b> | Average length of the reads or reads N50             |
| <b>Mean_Q</b>  | Mean Quality of the reads                            |
| <b>Q10%</b>    | percentage of reads > Q10                            |
| <b>Taxa</b>    | Taxa identified from the reads, closest strain match |
| <b>ANI</b>     | ANI number compared to the closest match             |
| <b>AMR</b>     | AMR gene content                                     |
| <b>STRESS</b>  | Stress gene identified                               |

| A                                  | B                                                |
|------------------------------------|--------------------------------------------------|
| <b>ST</b>                          | Sequence Type                                    |
| <b>Predicted identification</b>    | Salmonella enterica subspecies ID                |
| <b>Predicted antigenic profile</b> | In the case of Salmonella, the antigenic profile |
| <b>Predicted serotype</b>          | Salmonella Serotype                              |
| <b>cov.</b>                        | Mean coverage for contigs in the flye assembly.  |
| <b>circ.</b>                       | If the contigs are circular or not               |
| <b>count(Contigs)</b>              | Number of contigs in the flye assembly           |
| <b>sum(length)</b>                 | 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 Passing Metrics for the assemblies:

minimum depth/**coverage**:  $\geq 30X$

**ANI** score:  $\geq 95\%$

average quality score: **Mean Q**  $\geq 30$

Note: The coverage in the Table is based on the longest contig. If you want to know the coverage and close and circularity of each contig, please check the assembly\_info file.

### 5.4 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) |
|--------------|---------|-------|-----|---------|--------|------|----------------------------------------------------------------|--------|--------------------------------------------------|------------------|----------------------|----|--------------------------------------------------------|-----------------------------|--------------------|------|-------|----------------|-------------|
| CFSAN0000318 | 48      | 39770 | 51  | 4233.6  | 35.99  | 100  | Salmonella enterica subsp. enterica serovar Heidelberg str. 41 | 99.998 | aac(3)-IIId,aadA1,blaTEM-1,fosA7,mdsA,mdsB,te(A) | arsR,ggolS,ggolT | iroB,iroC,iroH,sodC1 | 15 | Salmonella enterica subspecies enterica (subspecies I) | 4:rr:1,2                    | Heidelberg         | 32   | Y     | 4              | 4957870     |

|             |    |       |    |        |       |     |                                                            |        |                                                |                                                                                                 |               |    |                                                        |         |       |    |   |   |         |
|-------------|----|-------|----|--------|-------|-----|------------------------------------------------------------|--------|------------------------------------------------|-------------------------------------------------------------------------------------------------|---------------|----|--------------------------------------------------------|---------|-------|----|---|---|---------|
|             |    |       |    |        |       |     |                                                            | 578    |                                                |                                                                                                 |               |    |                                                        |         |       |    |   |   |         |
| CFSAN008585 | 47 | 61645 | 51 | 4727.2 | 35.71 | 100 | Salmonella enterica subsp. enterica serovar Derby str. 626 | 99.837 | aac(3)-VIa,aadA1,fosA7.3,mdsA,mdsB,sul1,tet(A) | golS,golT,pcpA,pcpB,pcpC,pcpD,pcpE,pcpR,pcpS,qacEdelta1,silA,silB,silC,silE,silF,silP,silR,silS | irB,iroC,sinH | 40 | Salmonella enterica subspecies enterica (subspecies I) | 4:f,g:- | Derby | 47 | Y | 4 | 5086410 |



|                                                                                        |  |
|----------------------------------------------------------------------------------------|--|
|                                                                                        |  |
| CFSAN051458                                                                            |  |
| 46                                                                                     |  |
| 73568                                                                                  |  |
| 50                                                                                     |  |
| 5885.7                                                                                 |  |
| 36.17                                                                                  |  |
| 100                                                                                    |  |
| EscherichiacoliO121str.RM8352                                                          |  |
| 99.97                                                                                  |  |
| acrF,blaEC,emrD,mdtM                                                                   |  |
| arsC,arsR,astert,erW,tterZ,ymgB                                                        |  |
| ea,e,efa1,ehxA,espA,espB,espF,espl,espj,espk,espp,esppX1,fd,eC,lpfA,nleA,nleB,nleC'stx |  |
| 655                                                                                    |  |
| -                                                                                      |  |
| -:-:-                                                                                  |  |
| -:-:-                                                                                  |  |
| 79                                                                                     |  |
| N                                                                                      |  |
| 7                                                                                      |  |
| 5550579                                                                                |  |

|                       |     |           |     |             |           |       |                                                               |           |             |                                                                                                                                                       |                                          |   |   |           |          |     |   |   |               |
|-----------------------|-----|-----------|-----|-------------|-----------|-------|---------------------------------------------------------------|-----------|-------------|-------------------------------------------------------------------------------------------------------------------------------------------------------|------------------------------------------|---|---|-----------|----------|-----|---|---|---------------|
|                       |     |           |     |             |           |       |                                                               |           |             |                                                                                                                                                       | A 2 c, st x B 2 a, tc c P, ti r, t o x B |   |   |           |          |     |   |   |               |
| C F S A N 0 6 8 7 7 3 | 4 5 | 9 7 8 8 3 | 5 6 | 4 9 0 9 . 9 | 3 6 . 4 8 | 1 0 0 | C r o n o b a c t e r s a k a z a k i i A T C C B A A - 8 9 4 | 9 9 . 8 1 | b l a C S A | f i e F , p c o A , p c o B , p c o C , p c o D , p c o E , p c o R , p c o S , s i l A , s i l B , s i l C , s i l E , s i l F , s i l P , s i l R . |                                          | 1 | - | - : - : - | - -: - : | 9 9 | Y | 4 | 4 5 8 1 7 7 5 |

|  |             |    |       |    |        |       |     |                             |        |                                                             |                              |                           |   |   |       |       |     |   |   |         |
|--|-------------|----|-------|----|--------|-------|-----|-----------------------------|--------|-------------------------------------------------------------|------------------------------|---------------------------|---|---|-------|-------|-----|---|---|---------|
|  |             |    |       |    |        |       |     |                             |        | sil<br>S,<br>te<br>r<br>D,<br>te<br>r<br>W,<br>t<br>er<br>Z |                              |                           |   |   |       |       |     |   |   |         |
|  | CFSAN086180 | 44 | 67352 | 57 | 5751.5 | 36.68 | 100 | Klebsiella sp. X-16S - Nf21 | 99.811 | blaLEN-16, emrD, fosA, kdsA, oqxA, oqxB                     | fifE                         |                           | - | - | -:-:- | -:-:- | 69  | Y | 1 | 5536414 |
|  | CFSAN086182 | 43 | 49856 | 51 | 5250   | 36.36 | 100 | Citrobacter sp. KTE32       | 98.699 | blaCMY                                                      | arsA, arsB, arsC, arsD, arsR |                           | - | - | -:-:- | -:-:- | 52  | Y | 1 | 4917486 |
|  | CFSAN08     | 42 | 49755 | 55 | 9904.4 | 37.05 | 100 | Enterobact                  | 98.773 | ampC, fosA                                                  | fifE                         | ir<br>oB,<br>i<br>r<br>oC | - | - | -:-:- | -:-:- | 100 | Y | 2 | 4992359 |

|                         |                                                                |
|-------------------------|----------------------------------------------------------------|
|                         |                                                                |
| 6183                    | CFSAN122995                                                    |
|                         | 41                                                             |
|                         | 88297                                                          |
|                         | 50                                                             |
|                         | 5743.4                                                         |
|                         | 36.25                                                          |
|                         | 100                                                            |
| cancerogenus ATCC 35316 | Salmonella enterica subsp. diarizonae serovar 50:k:z str. MZO0 |
|                         | 99.285                                                         |
|                         |                                                                |
|                         |                                                                |
| ,iron                   | cdtB,iroB,iroC,iucA,iutA,ybtP,ybtQ                             |
|                         | 645                                                            |
|                         | Salmonella enterica subspecies diarizonae (subspecies IIIb)    |
|                         | 6,14 : z10 : z                                                 |
|                         | IIlb(6),14 : z10 : z                                           |
|                         | 92                                                             |
|                         | Y                                                              |
|                         | 1                                                              |
|                         | 5320423                                                        |

[illegible]

### 5.5 PT exercise circular closed reference genomes characteristics:

|  | Taxa       | CFSA N No.   | contig No. expected | Chromosome |        | Plasmids |       |       |       | Taxa                                                                  |
|--|------------|--------------|---------------------|------------|--------|----------|-------|-------|-------|-----------------------------------------------------------------------|
|  | GTV SS-021 | CFSA N068773 | 4                   | 4,408,038  | 13,196 | 31,208   | 3,983 |       |       | Cronobacter sakazakii ATCC BAA-894                                    |
|  | GTV SS-016 | CFSA N030807 | 6                   | 4,813,409  | 8,401  | 7,098    | 4,238 | 5,396 | 5,111 | Shigella sonnei 08-7761                                               |
|  | GTV SS-025 | CFSA N084952 | 1                   | 6,326,461  |        |          |       |       |       | Pseudomonas sp. HMWF034                                               |
|  | GTV SS-020 | CFSA N051458 | 2                   | 5,421,422  | 81,969 |          |       |       |       | Escherichia coli O121 str. RM8352                                     |
|  | GTV SS-019 | CFSA N044836 | 2                   | 2,864,614  | 57,565 |          |       |       |       | Listeria innocua ATCC 33091                                           |
|  | GTV SS-014 | CFSA N023469 | 1                   | 2,939,749  |        |          |       |       |       | Listeria monocytogenes CFSAN002349                                    |
|  | GTV SS-011 | CFSA N023464 | 1                   | 2,939,778  |        |          |       |       |       | Listeria monocytogenes CFSAN002349                                    |
|  | GTV SS-001 | CFSA N000189 | 2                   | 4,724,863  | 81,814 |          |       |       |       | Salmonella enterica subsp. enterica serovar Bareilly str. CFSAN000189 |

