

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-Escalona¹, Maria Hoffmann²

¹FDA; ²US FDA

GenomeTrakr

Tech. support email: 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/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>



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: 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, 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, 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 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>

Workflow Visualize Shared D

Welcome to Galaxy, please log in

Public Name or Email Address

Password

Forgot password? Click here to reset your password.

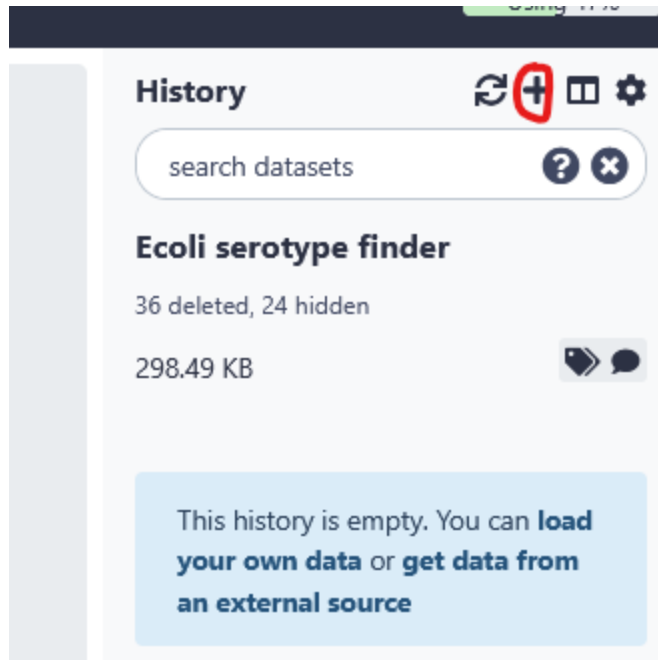
Login

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

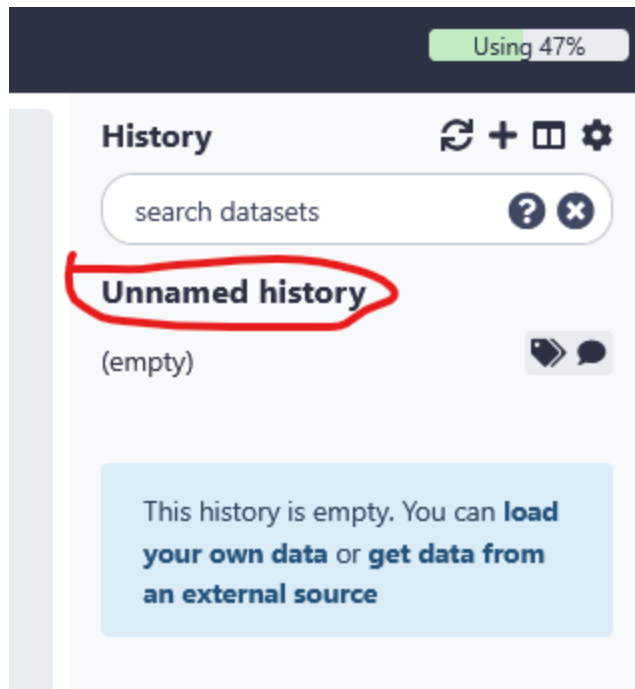
Create a new history

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

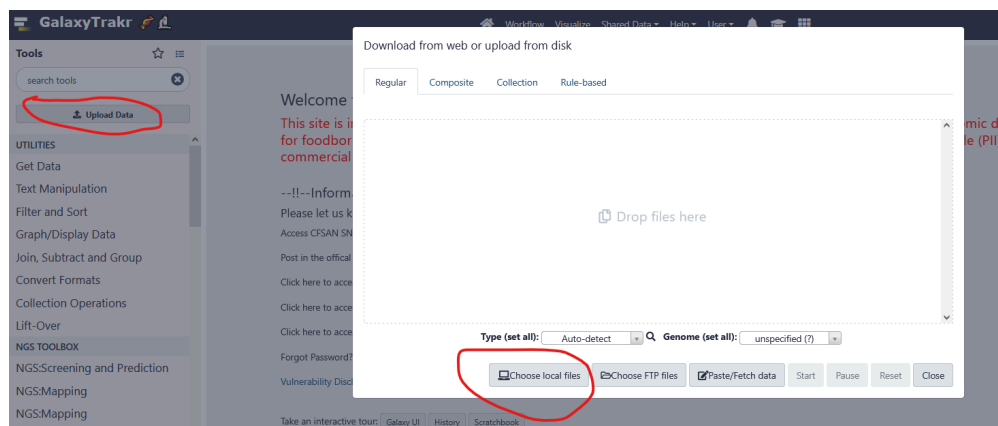


2.2 Name your new History by clicking on the "Unnamed history", type in desired name and hit enter. We recommend including the flow cell ID and the date the run was started.



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

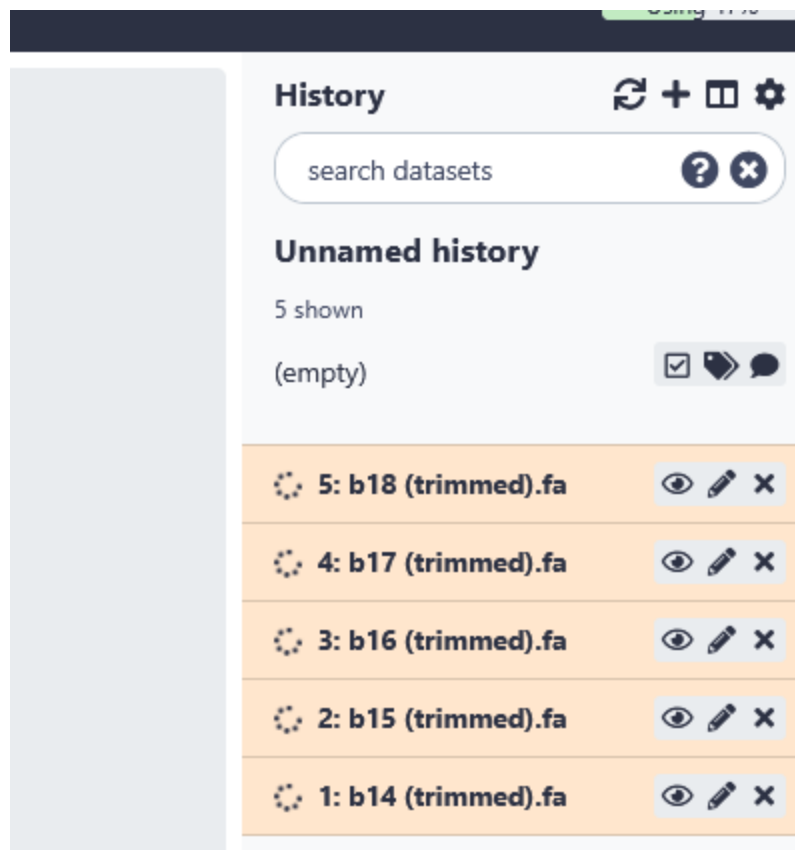
You added 5 file(s) to the queue. Add more files or click 'Start' to proceed.

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



History

? x

Unnamed history

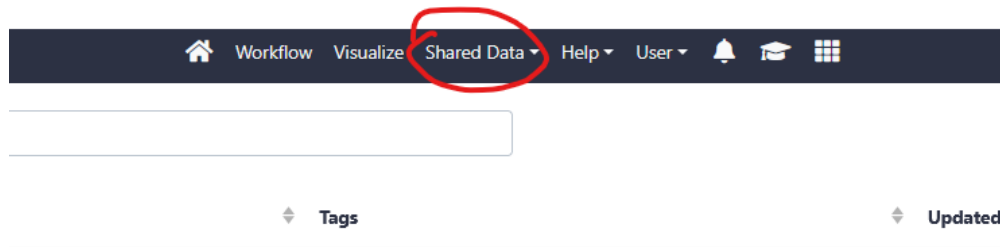
5 shown

680.93 MB ☒

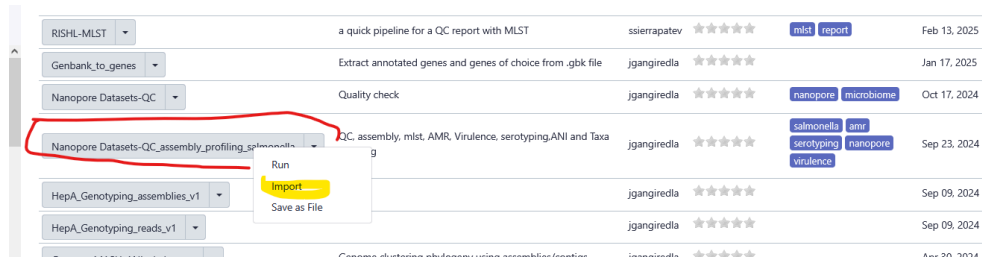
5: b18 (trimmed).fa	
4: b17 (trimmed).fa	
3: b16 (trimmed).fa	
2: b15 (trimmed).fa	
1: b14 (trimmed).fa	

Execute the ONT assembly workflow

- 4 Add the ONT (**Nanopore Datasets-QC_assembly_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.

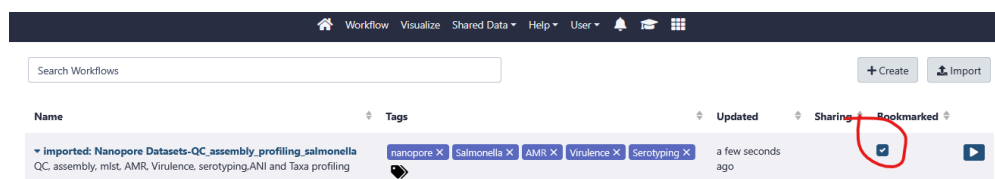


click on shared data and select "workflows"



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.

Workflow: imported: Nanopore Datasets-QC_assembly_profiling_salmonella ✓ Run Workflow

History Options

Send results to a new history ☒ Yes

History name FABX3450

1: Nanopore reads of a sample - in a single fastq or fastq.gz file

S: b18 (trimmed).fa

- 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.

Workflow: imported: Nanopore Datasets-QC_assembly_profiling_salmonella ✓ Run Workflow

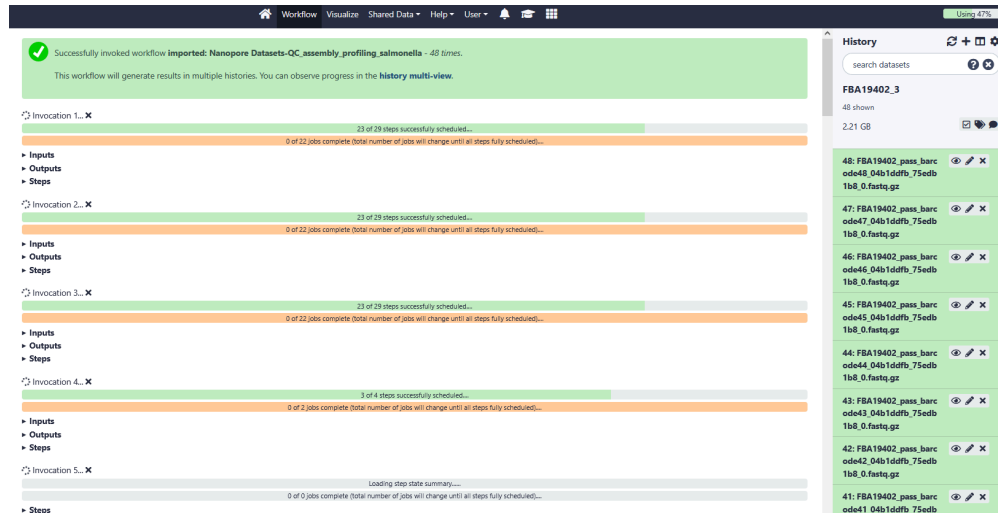
History Options

Send results to a new history ☒ Yes

History name b63_ONT_analysis

1: Nanopore reads of a sample - in a single fastq or fastq.gz file

63: barcode63.fastq



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).



History

search datasets

?

×

imported: Nanopore
Datasets-
QC_assembly_profiling_salmonella_FBA19402

5 shown, 42 hidden

70.92 MB

☒

47: Enter_sample_name

×

46: Results_Table

×

7: Flye on data 1: assembly info

×

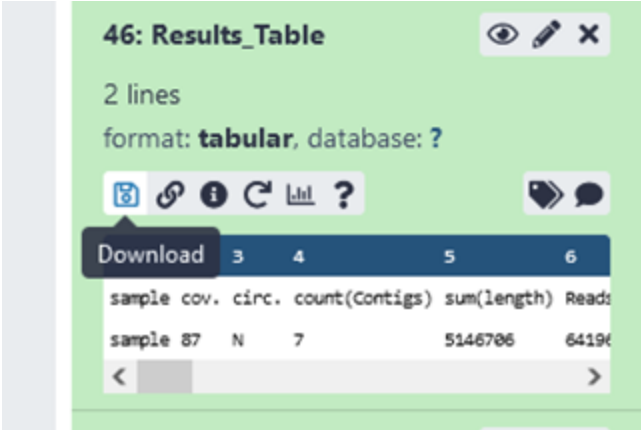
4: Flye on data 1: consensus

×

1: filtlong on data 61: Filtered FASTQ

×

- 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 closest 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	VI R U L E N C E	ST	Predicted identification	Predicted antigenic profile	Predicted serotype	cov.	circ.	count(Contigs)	sum(length)
CFSAN000318	48	39770	51	4233.6	35.99	100	Salmonella enterica subsp. enterica serovar Heidelberg	99.998	aac(3)-IIId,aadA1,blaTEM-1,fosA7,mdsA,mds	arsR,ggolS,ggolT	iroB,iroC,sinH,sodC1	15	Salmonella enterica subspecies enterica (sub	4:rr:1,2	Heidelberg	32	Y	4	4957870

								el b e r g s t r. 41 5 7 8		B, t e t (A)					p e c i e s l)								
	C F S A N 0 0 8 5 8 5	4 7	6 1 6 4 5	5 1	4 7 2 7 . 2	3 5. 7 1	1 0 0	S a l m o n e l a e n t e r i c a s u b s p. e n t e r i c a s e r o v a r D e r b y s t r. 6 2 6	9 9 . 8 3 7	a a c (3) - V I a a a d A 1, f o s A 7. 3, m d s A, m d s B, s u l t e t (A)	g o l S, g o l T, 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, q a c E d e l t a 1, s i l A, s i l B, s i l C, s i l E, s i l	i r o B, i r o C, s i n H	4 0	S a l m o n e l a e n t e r i c a s u b s p e c i e s e n t e r i c a (s u b s p e c i e s l)	4 : f , g : :-	D e r b y	4 7	Y	4	5 0 8 6 4 1 0			

											F, sil P, sil R, sil S								
C F S A N O 5 1 4 5 8	4 6	7 3 5 6 8	5 0	5 8 8 5 7	3 6. 1 7	1 0 0	E s c h e r i c h i a c o l i O 12 1 s t r. R M 8 3 5 2	9 9 . 9 7	a c r F, b l a E C, e m r D, m d t M	a r s C, a r s R, a s r t e r D, t e r W t e r Z, y m g B	e a e, e f a 1, e h x A e s p A e s p B e s p F, e s p J, e s p K, e s p P, e s p X 1, f d e C l p f A n i e A n	6 5 5	-	- : : : : -	- : : : : -	7 9	N	7	5 5 5 0 5 7 9

[illegible]

										E, sil F, sil P, sil R, sil S, te r D, te r W , t er Z										
	C F S A N O 8 6 1 8 0	4 4	6 7 3 5 2	5 7	5 7 5 1 .5	3 6. 6 8	1 0 0	Kl e b si el la s p. X 1- 16 S - N f2 1	9 9 .8 1 1	bl a L E N -1 6, e m r D, fo s A, k d e A, o q x A, o q x B	fi e F		-	-	- : : : : -	- : : : -	6 9	Y	1	5 5 3 6 4 1 4
	C F S A N O 8 6 1 8 2	4 3	4 9 8 5 6	5 1	5 2 5 0	3 6. 3 6	1 0 0	Ci tr o b a ct er s p. K T E 3 2	9 8 .6 9 9	bl a C M Y	ar s A, ar s B, ar s C, ar s D, ar s R		-	-	- : : : : -	- : : : -	5 2	Y	1	4 9 1 7 4 8 6

	CFSAN086183	42	49755	55	9904.4	37.05	100	Enterobacter cancerogenus ATCC 35316	98.773	ampC, fosA	fi e F	ir o B , i r o C , i r o N	-	-	- : - : -	- -: - :	100	Y	2	4992359
	CFSAN122995	41	88297	50	5743.4	36.25	100	Salmonella enterica subsp. diarizonae serovar 50:k:	99.285			c d t B , i r o B , i r o C , i u c A , i u t A , y b t P , y b t Q	645	Salm on el la en ter ic a s u b s p e ci es di ar iz on ae (s u b s	6,14 : z 10 : z	I l l b (6), 1 4 : z 1 0 : z	92	Y	1	5320423

							z st r. M Z 0 0 8 0						p e c i e s III b)							