

Interim Summary Report

EURL-*Salmonella* Proficiency Test Cluster Analysis 2023

W.F. Jacobs-Reitsma, R.E. Diddens, A.H.A.M. van Hoek, and K.A. Mooijman
RIVM, Bilthoven, The Netherlands

26 June 2024
Z&O letter report 2024-0051

This investigation was co-funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the granting authority European Health and Digital Executive Agency (HaDEA). Neither the European Union nor the granting authority can be held responsible for them.



Co-funded by
the European Union

1. Introduction

This document provides an overview of the results as produced by the participants in the EURL-*Salmonella* Proficiency Test (PT) Typing 2023, concerning the optional part on Cluster Analysis (CA).

A total of 20 participants took part in the cluster analysis. All participants performed WGS analysis; PFGE (till 2020) as well as MLVA (till 2022) are no longer included as options in this part of the PT Typing. The full results will be reported in more detail in the final report on the EURL-*Salmonella* PT Typing 2023.

2. *Salmonella* strains for cluster analysis

A total of six *Salmonella* strains (23SCA01 – 23SCA06) in HI agar transport tubes were sent to the participants in the EURL-*Salmonella* PT Typing 2023, part CA. Background information on these 'wet' strains is given in Table 1A. In addition, raw sequence data (fastq.gz files, md5 checksums) of another six *Salmonella* strains (23SCA11 – 23SCA16) were made available to the participants via the RIVM secure ftp server for 'dry' evaluation. Background information on the 'dry' strains is given in Table 1B.

Table 1A. Background information on the 'wet' *Salmonella* strains used for cluster analysis in 2023

Strain code	Serovar	Abbreviation	ST	Origin
23SCA01 ^{a)}	Typhimurium	STM	19	Human
23SCA02 ^{b)}	Bovismorbificans	Ref2 SBm	142	Human
23SCA03	Bovismorbificans	SBm	142	Human
23SCA04	Bovismorbificans	SBm	142	Human
23SCA05 ^{b)} REF	Bovismorbificans	Ref SBm	142	Human
23SCA06	Bovismorbificans	SBm	142	Human

a) Biological duplicate variable strain PT 2019 SCA03

b) Technical duplicates

Table 1B. Background information on the 'dry' *Salmonella* strains used for cluster analysis in 2023

Strain code	Serovar	Abbreviation	ST	Origin
23SCA11	Bovismorbificans	SBm	142	Human
23SCA12 ^{c)}	non- <i>Salmonella</i>	Citro		ATCC 8090
23SCA13 ^{d)}	Enteritidis	SE	11	Human
23SCA14 ^{e)}	Mixed strains	Ref+Citro		
23SCA15	Bovismorbificans	SBm	142	Human
23SCA16 ^{f)}	Mixed strains	Ref+STM		

c) *Citrobacter freundii*

d) PT 2021 pre-tested strain

e) Wet-mixed culture of strains 23SCA05 (REF) and 23SCA12 (*Citrobacter freundii*)

f) Wet-mixed culture of strains 23SCA05 (REF) and 23SCA01 (*S. Typhimurium*)

In preparation of the PT 2023 on cluster analysis, 11 human surveillance *Salmonella* strains were re-cultured from storage (2020/2021) and submitted for WGS analysis both directly (July 2023) and after sub-culturing for ten times (August 2023). In addition, a variable strain from EURL-*Salmonella* PT CA 2019 (19SCA03), a *Citrobacter freundii* strain and some wet-mixed cultures were included in this pre-testing. Subsequently, six 'wet' strains and six 'dry' strains were selected for inclusion in the PT 2023 (also see Figure 1). One set of wet technical duplicates was included: strain 23SCA02 and strain 23SCA05 shipment tubes were both prepared from the same blood-agar plate containing strain 23SCA05.

Like in previous years, the PT Cluster Analysis 2023 was mimicking an outbreak situation, with a *Salmonella* Bovismorbificans as the reference strain (23SCA-REF). Raw WGS data of this strain (23SCA-REF_R1.fastq.gz and 23SCA-REF_R2.fastq.gz, as well as their md5 checksums)

were also made available through the RIVM secure ftp server. For this particular PT 2023, the cgMLST-based cluster definition was set at a maximum of five cgMLST allelic differences from the provided reference sequence. Participants were asked to analyse the six 'wet' *Salmonella* strains and the six 'dry' ones, and to report per strain whether a cluster with the reference strain was found or not. Analyses could be performed up to the choice of the participant by WGS, using their own routine method(s).

Evaluation of the participants' cluster analysis results was done by comparing the participants' results to the expected results in the outbreak investigation setting, as pre-defined by the EURL-*Salmonella* (Protocol EURL-*Salmonella* PT Typing 2023).

3. Evaluation of the cluster analysis results based on WGS data

Twenty participants (Table 2) submitted a total of 24 cluster analysis results based on WGS data; two participants submitted cgMLST-based as well as SNP-based data results, while one participant submitted two cgMLST-based and one SNP-based data analyses. Some details on the sequencing and analysis procedures performed by the participants are given in Annex 1. Pre-test results as well as the PT 2023 results from the EURL-*Salmonella* are shown in Figure 1. Sequencing was performed in-house, on an Illumina NextSeq platform. Raw data were processed via an in-house developed Juno-assembly pipeline (https://rivm-bioinformatics.github.io/ids_bacteriology_man/juno-assembly.html), which includes the SPAdes 3.15.3 assembler. Cluster analysis was done in Ridom SeqSphere+ (version: 9.0.8 (2023-06)), using the cgMLST Enterobase v2.0 scheme and visualised in a minimum spanning tree (MST, Figure 1).

Stable and consistent cgMLST analysis results were obtained for all *S. Bovismorbificans* strains, whereas the *S. Typhimurium* strain again showed variable results, like in PT 2019 (Figure 1: orange ELt0, ELt1, ELt2).

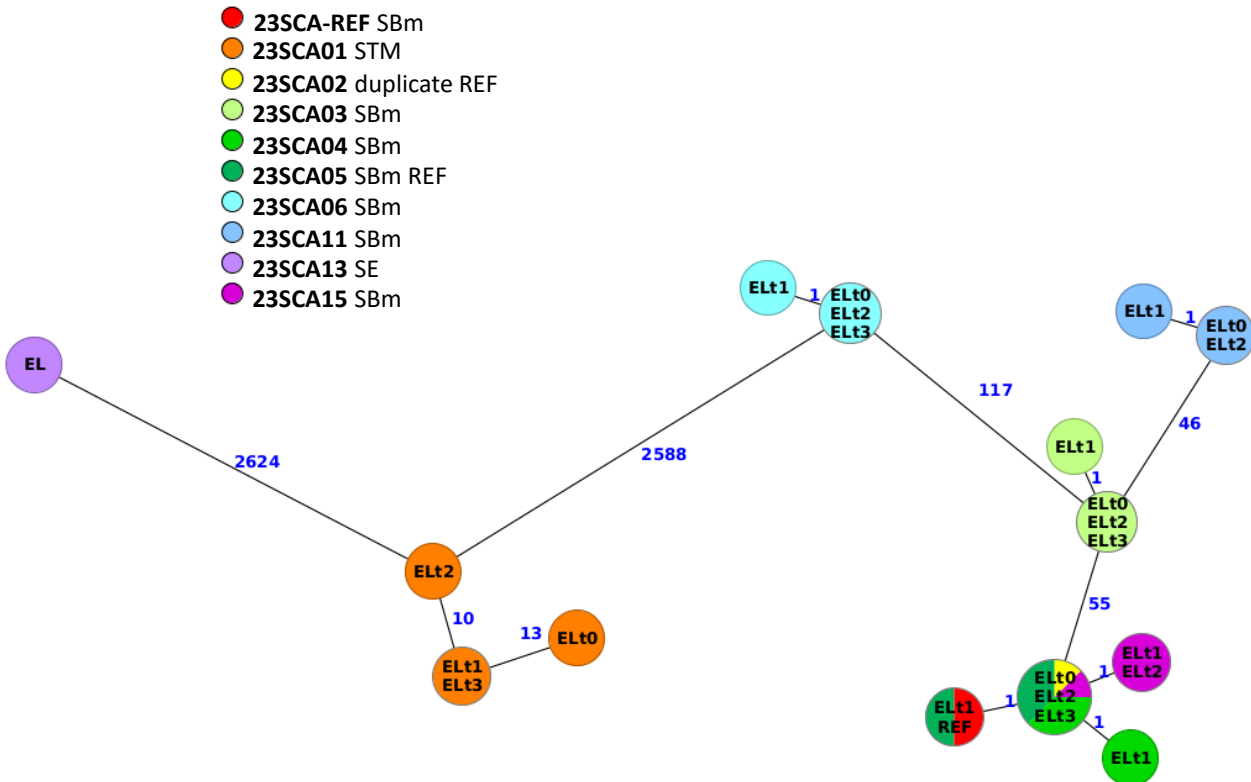


Figure 1. MST of the EURL-*Salmonella* (EL) pre-tests and PT 2023 results, (RidomSeqSphere+, cgMLST (3002), pairwise ignoring missing values and QC-failing strains excluded).

Participants were asked to report per strain:

- whether the data passed their Quality Control (QC) criteria or not,
- whether a cluster with the reference strain in the EURL-*Salmonella* PT Typing 2023 (23SCA-REF) was found or not.

As a general question, the participants were asked if and how they serotyped the strains. Seventeen participants indicated to have serotyped the strains. These serotyping results are given in Annex 5, for information purposes only.

Strains 23SCA12, 23SCA14, and 23SCA16 were expected not to pass the QC of the participants, because of their characteristics (Table 1B). The PT Typing 2023 Protocol indicated to exclude strains from the cluster analysis if the data did not pass the QC.

Participants' reasons for (not) excluding these strains are given in Annex 6.

Strain 23SCA12 (*Citrobacter freundii*) was reported not passing the QC criteria by 18 of the 20 participants, 2 participants did not exclude it from the distance matrix.

Strain 23SCA14 (mixture Ref+Citro) was reported not passing the QC criteria by 19 of the 20 participants, 3 participants did not exclude it from the distance matrix.

Strain 23SCA16 (mixture Ref+STM) was reported not passing the QC criteria by 12 of the 20 participants, 10 participants did not exclude it from the distance matrix.

Interpretation of what is part of QC criteria seems to differ per participant:

Two participants (Lab 11 and Lab 23) regarded strain 23SCA12 passing the QC criteria. Both noted this strain not to be *Salmonella* (Annex 4 and Annex 5). Consequently, Lab 11 indeed excluded this strain from the distance matrix, but Lab 23 did not.

The PT Protocol detailed the reference strain (23SCA-REF) to be serovar Bovismorbificans, and five participants (Labs 9, 11, 15, 20, and 22-SNPr) regarded this serovar as a criterium, excluding all other serovars (strains 23SCA01 and 23SCA13) from their distance matrix, see Annex 7. The EURL-*Salmonella* did not exclude other *Salmonella* serovars in this PT's cluster analysis, because not all participants may do (any kind of) serotyping before or during the WGS analysis since it is not obligatory in the CA part of the PT Typing 2023.

Annex 7 shows per submission the participants' distance matrix data, also regarding strains that were (not) excluded from the cluster analysis.

The cluster definition for this particular PT Typing 2023 situation was set at maximum 5 allelic differences from the reference sequence. Based on this cgMLST-based criterium, results were expected to indicate the 'wet' strains 23SCA05 (reference strain), 23SCA02 (technical duplicate of the reference strain), 23SCA04 (clustering with the reference strain), and the 'dry' strain 23SCA15 to form a cluster with the provided reference outbreak strain 23SCA-REF data (also see Figure 1).

Twenty of the 24 submissions (three participants with multiple submissions) reported those four strains to be part of the cluster (Table 2). However, seven of those submissions also included mixed serovars 'strain' 23SCA16 into the cluster. Consequently, 13 of the 24 submissions reported the cluster completely as expected (Table 2).

Technical duplicate strains 23SCA05 and 23SCA02 were reported within (the only) one cluster in all 24 submissions.

Optionally, any other cluster(s) apart from the reference could be submitted. Seven participants correctly reported no other clusters to be present.

Table 2. Expected cluster analysis results and those as reported by the 20 participants

Labcode-method	Strain code											
	23 SCA01	23 SCA02	23 SCA03	23 SCA04	23 SCA05	23 SCA06	23 SCA11	23 SCA12	23 SCA13	23 SCA14	23 SCA15	23 SCA16
	STM	Ref2 Bm	Bm	Bm	Ref Bm	Bm	Bm	Citro	SE	Ref+Citro	Bm	Ref+STM
Expected	No	Yes	No	Yes	Yes	No	No	n.a.	No	n.a.	Yes	n.a.
3-SNP	No	Yes	Yes	Yes	Yes	Yes	Yes	n.a.	n.a.	n.a.	Yes	No
4-cgMLST	No	Yes	No	Yes	Yes	No	No	n.a.	No	n.a.	Yes	Yes
5-cgMLST ^{a)}	No	Yes	No	Yes	Yes	No	No	n.a.	No	n.a.	Yes	Yes
8-cgMLST	No	Yes	No	Yes	Yes	No	No	n.a.	No	n.a.	Yes	n.a.
9-cgMLST ^{b)}	No	Yes	No	Yes	Yes	No	No	n.a.	No	n.a.	Yes	n.a.
11-cgMLST	No	Yes	No	Yes	Yes	No	No	No	No	n.a.	Yes	n.a.
14-cgMLST ^{c)}	No	Yes	No	Yes	Yes	No	No	n.a.	No	n.a.	Yes	n.a.
15-cgMLST	No	Yes	No	Yes	Yes	No	No	n.a.	No	n.a.	Yes	n.a.
16-cgMLST	No	Yes	No	Yes	Yes	No	No	n.a.	No	n.a.	Yes	n.a.
19-cgMLST ^{d)}	No	Yes	No	Yes	Yes	No	No	n.a.	No	n.a.	Yes	Yes
20-SNP ^{e)}	n.a.*	Yes	No	Yes	Yes	No	No	n.a.	n.a.*	n.a.	No	n.a.
21-cgMLST	No	Yes	No	Yes	Yes	No	No	n.a.	No	n.a.	Yes	n.a.
22-cgMLST1	No	Yes	No	Yes	Yes	No	No	n.a.	No	n.a.	Yes	n.a.
22-cgMLST2	No	Yes	No	Yes	Yes	No	No	n.a.	No	n.a.	Yes	n.a.
22-SNP ^{f)}	No	Yes	No	Yes	Yes	No	No	n.a.	No	n.a.	Yes	n.a.
23-SNP ^{g)}	No	Yes	No	Yes	Yes	No	No	No	No	n.a.	Yes	n.a.
25-cgMLST ^{h)}	No	Yes	No	Yes	Yes	No	No	n.a.	No	n.a.	Yes	Yes
26-cgMLST	No	Yes	No	Yes	Yes	No	No	n.a.	No	n.a.	Yes	Yes
27-cgMLST	No	Yes	No	Yes	Yes	No	No	n.a.	No	n.a.	Yes	Yes
28-cgMLST	No	Yes	No	Yes	Yes	No	No	n.a.	No	Yes	No	Yes
28-SNP	No	Yes	No	Yes	Yes	No	No	n.a.	No	Yes	No	Yes
30-cgMLST	No	Yes	No	Yes	Yes	No	No	n.a.	No	n.a.	Yes	Yes
85-cgMLST	No	Yes	No	Yes	Yes	No	No	n.a.	No	n.a.	Yes	n.a.
85-SNP ⁱ⁾	No	Yes	No	Yes	Yes	No	No	n.a.	No	n.a.	Yes	n.a.

n.a.: Not applicable (QC not passed) n.a.*: Other serovar than Bm

- Comment Lab 5: extra care should be taken with the conclusion of Strain 23SCA16 since the threshold of 10% missing loci was almost passed.
- Comment Lab 9: Samples not identified as *S. Bovismorbificans* were also excluded from the distance matrix.
- Comment Lab 14: 23SCA16: Exceeded our quality parameter for genome length by only 3,607bp.
- Comment Lab 19: 23SCA16 did not completely pass the quality check (contigs >1000: 154, N50: 99482), but was so close, that it was processed for further analyses.
- Comment Lab 20: Isolate 23SCA15 is close to the reference cluster, though may represent an isolate with shared lineage but not the same outbreak.
- Comment Lab 22: For SNP analysis, only isolates that were typed as *S. Bovismorbificans* ST142 and fulfilled all QC criteria were included in the analysis.
- Comment Lab 23: Strain 23SCA12 was included in the cluster analysis as it passed in the quality control. However this strains is not *Salmonella enterica* but was identified by kmerFinder as *Citrobacter freundii*.
- Comment Lab 25: For the time being, *S. Bovismorbificans* is not sequenced on a routine basis. Assuming that strain 23SCA15 is phenotypically fully susceptible, whereas the reference strain is multi-resistant, we probably (depending on the epidemiological information available) would not have it included in the sequencing / outbreak investigation.
- Comment Lab 85: For SNP analysis between 0 and 9 SNP of difference. Our threshold is at 21 SNP of difference for a cluster (Pightling et al., 2018). Strain 23SCA16 is identified also *Bovismorbificans* and <21 SNP of difference, but the analysis is unsure. In a case of outbreak we could make a focus on this strain also and ask to re-sequence it. As it don't pass our QC test, we haven't add it to the cluster analysis.

In blue: Deviation from the expected result.

4. Concluding remarks

Results from PT 2023 Cluster analysis (Table 2, Annex 7) indicate some points to consider:

- Check for contamination inter- and intraspecies;
- Use multiple typing tools, for example serotyping as well as MLST (7-housekeeping genes) analysis, and keep track of deviations;
- Implement WGS data quality checks in routine analysis. For example see ISO 23418, and the guidance documents produced by the Inter EURL Working Group on NGS.
- Exclude non-*Salmonella* isolates from a *Salmonella* outbreak analysis.

List of abbreviations

BPW	Buffered peptone water
CA	Cluster Analysis
cgMLST	core genome Multilocus Sequence Typing
EFSA	European Food Safety Authority
EL	EURL- <i>Salmonella</i> Laboratory
EU	European Union
EURL- <i>Salmonella</i>	European Union Reference Laboratory for <i>Salmonella</i>
ftp	file transfer protocol
HI agar	Hearth Infusion ager (in transport tubes)
MLVA	Multiple-Locus Variable number of tandem repeat Analysis
MST	Minimum Spanning Tree
n.a.	not applicable
NRLs- <i>Salmonella</i>	National Reference Laboratories for <i>Salmonella</i>
PFGE	Pulsed Field Gel Electrophoresis
PT	Proficiency Test
QC	Quality Control
REF	Reference
RIVM	National Institute for Public Health and the Environment
SBm	<i>Salmonella</i> Bovismorbificans
SE	<i>Salmonella</i> Enteritidis
SNPa	assembly-based Single-Nucleotide Polymorphism data
SNPr	reference-based Single-Nucleotide Polymorphism data
ST	Sequence Type
STM	<i>Salmonella</i> Typhimurium
WGS	Whole Genome Sequencing

References

- EN ISO 23418. Microbiology of the food chain – Whole genome sequencing for typing and genomic characterization of bacteria – General requirements and guidance. International Organization for Standardization, Geneva, Switzerland.
- Inter EURL Working Group on NGS. Guidance documents NGS.
<https://www.eurlsalmonella.eu/methods/ngs/guidance-documents> (accessed 26-6-204).

Acknowledgements

We would like to thank the IDS-bioinformatics team for their valuable help with the Juno-assembly pipeline.

Contact for this PT

Wilma Jacobs
E-Mail: wilma.jacobs@rivm.nl
Mobile: +31 6 3114 2419

EURL-*Salmonella* website: www.eurlsalmonella.eu

National Institute for Public Health and the Environment (RIVM)
Centre for Zoonosis and Environmental microbiology (Z&O/ internal mailbox 63)
Antonie van Leeuwenhoeklaan 9
P.O. Box 1, 3720 BA Bilthoven, The Netherlands

Annex 1. Sequencing and analyses details on the 20 participants and the EURL-*Salmonella*

Labcode	Wet lab ^{a)}	WGS platform	Data analysis ^{b)}	Tool for analysis	Method for cluster analysis
EURL-Salm	In-In-In	Illumina NextSeq	cgMLST-based	Ridom SeqSphere	Minimum Spanning Tree (MST)
3	In-In-In	Illumina MiSeq	SNPr-based	EnterobaseGalaxy	Neighbor joining (NJ)
4	In-In-In	Illumina MiSeq	cgMLST-based	In-house pipeline with Chewbbaca and scheme from Enterobase	Single linkage
5	In-In-In	Illumina MiSeq	cgMLST-based	in-house galaxy	MSTreeV2
8	In-In-In	Illumina MiSeq	cgMLST-based	linux command line	Neighbor joining (NJ)
9	In-In-In	Illumina MiSeq	cgMLST-based		MSTree V2 - Grapetree's modified version of MST
11	In-In-In	Illumina MiSeq	cgMLST-based	BioNumericsDTU-cgMLSTFinder 1.2	Complete linkage
14	In-In-In	Illumina MiSeq	cgMLST-based	BioNumerics	Minimum Spanning Tree (MST)
15	In-In-In	Illumina MiSeq	cgMLST-based	Ridom SeqSphere	Minimum Spanning Tree (MST)
16	In-In-In	Illumina NextSeq	cgMLST-based	Ridom SeqSphere	Minimum Spanning Tree (MST)
19	In-Out-Out	Illumina NovaSeq	cgMLST-based	cgMLSTFinder 1.2	Neighbor joining (NJ)
20	In-In-In	Illumina NextSeq	SNPr-based	TrimGalore, Snippy, snp-dist, Gubbins, RAxML	Maximum likelihood (ML)
21	In-In-In	Illumina MiSeq	cgMLST-based	chewBBACA and chewTree on Galaxy Aries, Enterobase as reference	Don't know
22-cgMLST1	In-In-In	Illumina NextSeq	cgMLST-based	inhouse chewieSnake pipeline (scheme: Enterobase)	single linkage hierarchical clustering
22-cgMLST2	In-In-In	Illumina NextSeq	cgMLST-based	inhouse chewieSnake pipeline (scheme: EFSA)	single linkage hierarchical clustering
22-SNPr	In-In-In	Illumina NextSeq	SNPr-based	SnippySnake	single linkage hierarchical clustering
23	In-In-In	Illumina MiSeq	SNPa-based	CSIPhylogeny 1.4	Maximum likelihood (ML)
25	In-In-In	Illumina MiSeq	cgMLST-based	Ridom SeqSphere	Minimum Spanning Tree (MST)
26	In-In-In	Illumina MiSeq	cgMLST-based	chewBBACA	Minimum Spanning Tree (MST)
27	In-In-In	Illumina MiniSeq	cgMLST-based	Ridom SeqSphere	MST to visualize but only Distance Matrix is used for interpreting
28-cgMLST	In-In-In	Illumina MiSeq	cgMLST-based	chewBBACA with scheme based on Enterobase cgMLST scheme. ReporTree for clustering.	Minimum Spanning Tree (MST)
28-SNPr	In-In-In	Illumina MiSeq	SNPr-based	Python-based pipeline using bowtie2 and bcftools.	Minimum Spanning Tree (MST)
30	In-In-In	Illumina MiSeq	cgMLST-based	Ridom SeqSphere	Minimum Spanning Tree (MST)
85-cgMLST	In-Out-Out	Illumina NovaSeq	cgMLST-based	Ridom SeqSphere	Minimum Spanning Tree (MST)
85-SNPr	In-Out-Out	Illumina NovaSeq	SNPr-based	Snippy	Maximum likelihood (ML)

a) Wet lab preparations: DNA extraction, Library preparation, Sequencing. In: In-house, Out: Outsourced.

b) SNPa: assembly-based, SNPr: reference-based.

Annex 2. Md5 checksums of the 14 fastq.gz files that had to be downloaded from the RIVM secure ftp server for further analysis (content of file 'ALLmd5-PT2023')

03b4fcc4d51aa1388842c1ac0cc1eae6	23SCA11_R1.fastq.gz
50749a789ccfe296bea0a603526ac2cd	23SCA11_R2.fastq.gz
2ea88c574109869383bbfcf6e9d73ad8	23SCA12_R1.fastq.gz
d01555ede19b9a4382da5c2a677a5805	23SCA12_R2.fastq.gz
0051903560ee287a65d052ba068674a2	23SCA13_R1.fastq.gz
2558324ace8d372a76914a4755b1e7e2	23SCA13_R2.fastq.gz
ee91a6519d73c6db84600df1272fec71	23SCA14_R1.fastq.gz
52210c19b18e7447a1d07efe6b758194	23SCA14_R2.fastq.gz
421e5785522c276a55e0fed03ff20b3c	23SCA15_R1.fastq.gz
920c36c8b66a49ad26d34ae85e49fe09	23SCA15_R2.fastq.gz
fe8e9eb9fc9acc5a7600faeddba6d2e5	23SCA16_R1.fastq.gz
620a7d30a1510e14549897c0c9ada08a	23SCA16_R2.fastq.gz
ee02957b374c359e7b0cd48cbaad668f	23SCA-REF_R1.fastq.gz
b1ed161c4ccdd6a64e6b572e821b841e	23SCA-REF_R2.fastq.gz

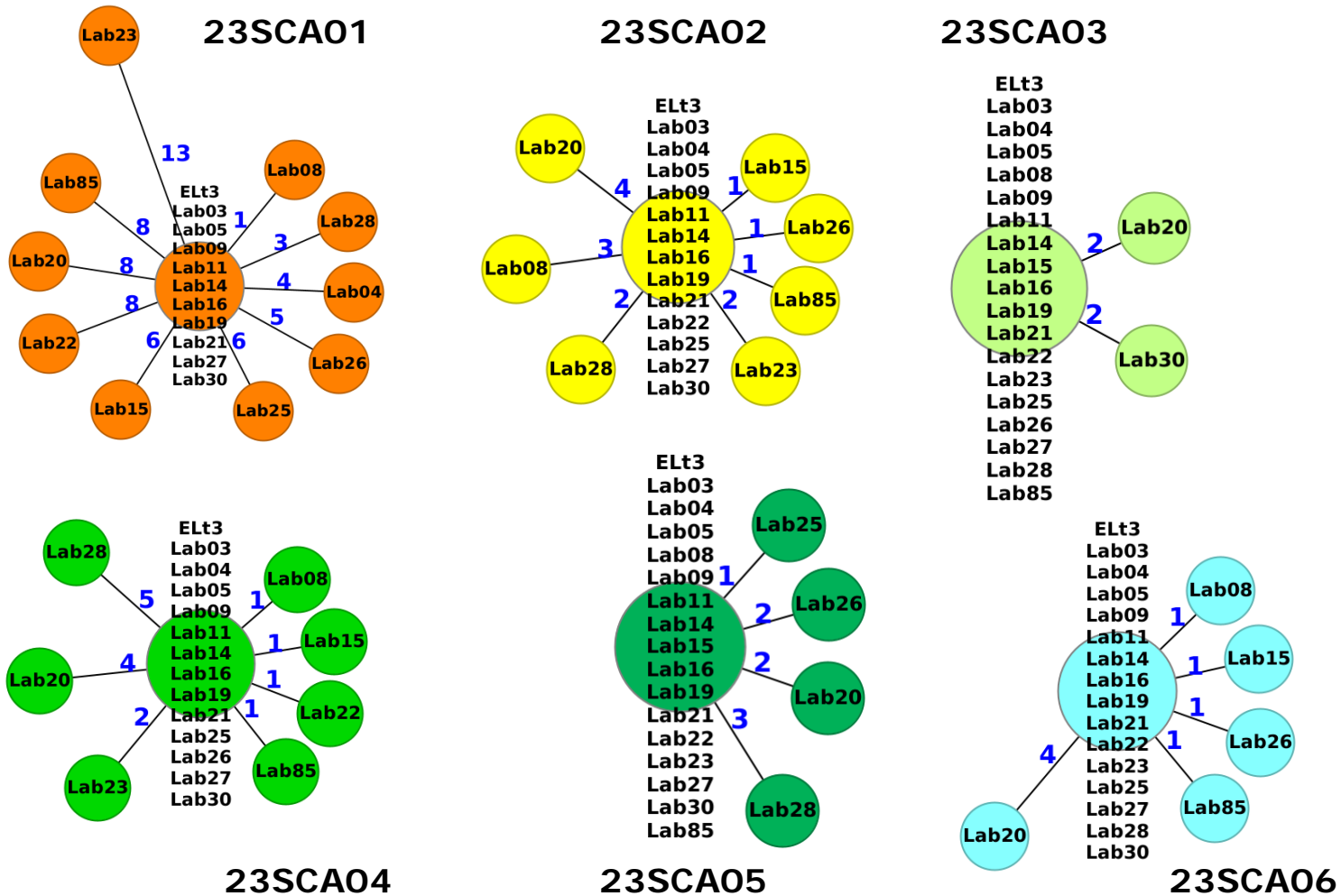
Annex 3. Integrity checks of downloaded and uploaded files using md5sums

Labcode	'Dry' strains 23SCA11 – 23SCA16			'Wet' strains 23SCA01 – 23SCA06	
	Did you check the md5sum values of the downloaded files	Lab's md5sum data submitted	Check by the EURL- <i>Salmonella</i>	Lab's md5sum data submitted	Md5sum determination and check by the EURL- <i>Salmonella</i> after downloading
3	No	No	na	No	na
4	Yes	Yes	OK	Yes	OK
5	Yes	No	na	No	na
8	No	Yes	OK	Yes	OK
9	Yes	Yes	OK	Yes	OK
11	No	No	na	No	na
14	Yes	Yes	OK	Yes	OK
15	Yes	Yes	OK	Yes	OK, after re-sending 23SCA04 data
16	Yes	Yes	OK	No	na
19	Yes	No	na	No	na
20	Yes	Yes	OK	Yes	OK
21	Yes	Yes	OK	Yes	OK
22	Yes	Yes	OK	Yes	OK
23	Yes	Yes	OK	Yes	OK, after several re-tries
25	Yes	Yes	OK	Yes	OK
26	Yes	Yes	OK	Yes	OK
27	No	No	na	No	na
28	Yes	Yes	OK	Yes	OK
30	Yes	Yes	OK	Yes	OK
85	Yes	Yes	OK*	Yes	OK*, except for 23SCA02

na: not applicable, OK: matching data.

*: OK for the uncompressed fastq files, instead of the requested compressed fastq.gz files. Failure in uncompressing strain 23SCA02 data.

Annex 4. MSTs of each 'wet' strain, using all participants' processed raw data (Ridom SeqSphere+, cgMLST (3002), pairwise ignoring missing values)



Annex 5. Serotyping results (optional)

No data submitted by Participants 5, 27, and 30.

In blue: Deviation from the expected result.

Lab code	Serotyping method(s) used	Strain code					
		23SCA01	23SCA02	23SCA03	23SCA04	23SCA05	23SCA06
EL	In-house pipeline	Typhimurium	Bovismorbificans	Bovismorbificans	Bovismorbificans	Bovismorbificans	Bovismorbificans
3	WGS	Typhimurium	Bovismorbificans	Bovismorbificans	Bovismorbificans	Bovismorbificans	Bovismorbificans
4	SISTR	Typhimurium	Bovismorbificans	Bovismorbificans	Bovismorbificans	Bovismorbificans	Bovismorbificans
8	WGS serotypefinder	Typhimurium	Bovismorbificans	Bovismorbificans	Bovismorbificans	Bovismorbificans	Bovismorbificans
9	Classical serology	Typhimurium	not serotyped	not serotyped	not serotyped	not serotyped	not serotyped
11	WGS	Typhimurium	Bovismorbificans	Bovismorbificans	Bovismorbificans	Bovismorbificans	Bovismorbificans
14	Bionumerics SeqSero	Typhimurium	Bovismorbificans	Bovismorbificans	Bovismorbificans	Bovismorbificans	Bovismorbificans
15	SeqSero+SISTR	Typhimurium	Bovismorbificans	Bovismorbificans	Bovismorbificans	Bovismorbificans	Bovismorbificans
16	In-house pipeline, based on SeqSero2	Typhimurium	Bovismorbificans	Bovismorbificans	Bovismorbificans	Bovismorbificans	Bovismorbificans
19	SeqSero2 v1.1.0	Typhimurium	Bovismorbificans	Bovismorbificans	Bovismorbificans	Bovismorbificans	Bovismorbificans
20	Classical serology	Typhimurium	Bovismorbificans	Bovismorbificans	Bovismorbificans	Bovismorbificans	Bovismorbificans
21	SeqSero_v2	Typhimurium	Hindmarsh or Bovismorbificans	Hindmarsh or Bovismorbificans	Hindmarsh or Bovismorbificans	Hindmarsh or Bovismorbificans	Hindmarsh or Bovismorbificans
22	Classical serologySISTR v1.1.1	Typhimurium	Bovismorbificans	Bovismorbificans	Bovismorbificans	Bovismorbificans	Bovismorbificans
23	Classical serologySeqSero 1.2	Typhimurium 4:i:1,2	Bovismorbificans 8:r:1,5	Bovismorbificans 8:r:1,5	Bovismorbificans 8:r:1,5	Bovismorbificans 8:r:1,5	Bovismorbificans 8:r:1,5
25	Classical serologySeqSero, SISTR	Typhimurium	Bovismorbificans	Bovismorbificans	Bovismorbificans	Bovismorbificans	Bovismorbificans
26	NGS, sistr	Typhimurium	Bovismorbificans	Bovismorbificans	Bovismorbificans	Bovismorbificans	Bovismorbificans
28	Sistr	Typhimurium	Bovismorbificans	Bovismorbificans	Bovismorbificans	Bovismorbificans	Bovismorbificans
85	SeqSero2 + Sistr	Typhimurium	Bovismorbificans	Bovismorbificans	Bovismorbificans	Bovismorbificans	Bovismorbificans

Lab code	Serotyping method(s) used	Strain code					
		23SCA11	23SCA12	23SCA13	23SCA14	23SCA15	23SCA16
EL	In-house pipeline	Bovismorbificans	failed QC	Enteritidis	failed QC	Bovismorbificans	failed QC
3	Enterobase	Bovismorbificans	-	Enteritidis	Bovismorbificans	Bovismorbificans	Bovismorbificans
4	SISTR	Bovismorbificans	Not typable/not Salmonella	Enteritidis	Bovismorbificans/ contaminated	Bovismorbificans	Bovismorbificans
8	WGS serotypefinder	Bovismorbificans	N/A	Enteritidis	N/A	Bovismorbificans	N/A
9	SeqSero2	Bovismorbificans	Not Salmonella	Enteritidis	Bovismorbificans	Bovismorbificans	Lindenburg
11	SeqSero, BioNumerics	Bovismorbificans	not Salmonella (Citrobacter freundii)	Enteritidis	Bovismorbificans-QC failed	Bovismorbificans	Bovismorbificans-QC failed
14	Bionumerics SeqSero	Bovismorbificans	Not applicable	Enteritidis	Bovismorbificans (contaminated)	Bovismorbificans	Bovismorbificans (contaminated)
15	SeqSero+SISTR	Bovismorbificans	/(Citrobacter)	Enteritidis	/(Salmonella and Citrobacter)	Bovismorbificans	Bovismorbificans + other Salmonella serovar
16	In-house pipeline, based on SeqSero2	Bovismorbificans	insufficient WGS quality	Enteritidis	insufficient WGS quality	Bovismorbificans	multiple serotypes
19	SeqSero2 v1.1.0	Bovismorbificans	not Salmonella	Enteritidis	Bovismorbificans	Bovismorbificans	Bovismorbificans
20	MOST, SeqSero and sistr	Bovismorbificans	No Type	Enteritidis	Bovismorbificans	Bovismorbificans	Bovismorbificans and Bsilla
21	SeqSero_v2	Hindmarsh or Bovismorbificans	N/A - failed QC	Enteritidis	N/A - failed QC	Hindmarsh or Bovismorbificans	N/A - failed QC
22	SISTR v1.1.1	Bovismorbificans	no Salmonella (Citrobacter)	Enteritidis	Bovismorbificans (with QC Warning)	Bovismorbificans	Bovismorbificans
23	SeqSero 1.2	Bovismorbificans 8:r:1,5	Not applicable	Enteritidis 9:g,m:-	Bovismorbificans 8:r:1,5	Bovismorbificans 8:r:1,5	Bovismorbificans 8:r:1,5
25	SeqSero, SISTR	Bovismorbificans		Enteritidis		Bovismorbificans	Bovismorbificans
26	NGS, sistr	Bovismorbificans		Enteritidis		Bovismorbificans	Bovismorbificans
28	Sistr	Bovismorbificans		Enteritidis	Bovismorbificans	Bovismorbificans	Bovismorbificans
85	SeqSero2 + Sistr	Bovismorbificans	Non Salmonella (Citrobacter)	Enteritidis	Contaminated	Bovismorbificans	Contaminated

Annex 6. Reasons for (not) passing participant's QC and (not) excluding strains 23SCA12, 23SCA14, and 23SCA16 from cluster analysis

In blue: Deviation from the expected result.

Lab code	Strain 23SCA12 passed QC	Reason(s) not passing QC	Strain 23SCA12 excluded from cluster analysis ^{a)}
Expected	No	Not <i>Salmonella</i> but <i>Citrobacter freundii</i>	Yes
3-SNPr	No	Per sequence GC content	No
4-cgMLST	No	This is <i>Citrobacter freundii</i> , not passing the contamination/species check	Yes
5-cgMLST	No	Contamination: <i>Citrobacter freundii</i>	Yes
8-cgMLST	No	contamination with <i>Citrobacter freundii</i>	Yes
9-cgMLST	No	<i>Citrobacter</i> genome	Yes
11-cgMLST	Yes		Yes
14-cgMLST	No	Failed Core Percent quality parameter. Kmer Finder confirmed isolate was not <i>Salmonella</i> .	Yes
15-cgMLST	No	Non <i>Salmonella</i> spp. : <i>Citrobacter</i>	Yes
16-cgMLST	No	contamination too high, contaminated with <i>Citrobacter</i>	Yes
19-cgMLST	No	<i>Citrobacter</i> sp.	Yes
20-SNPr	No	Contaminated: KmerID largely <i>Citrobacter</i>	Yes
21-cgMLST	No	Identified as <i>Citrobacter</i> by kmerfinder	Yes
22-cgMLST1	No	Read Fraction Majority Genus: no <i>Salmonella</i> (AQUAMIS), fraction of loci found in cgMLST below 95% (86%) - ChewieSnake	Yes
22-cgMLST2	No	Read Fraction Majority Genus: no <i>Salmonella</i> (AQUAMIS), fraction of loci found in cgMLST below 95% (86%) - ChewieSnake	Yes
22-SNPr	No	Read Fraction Majority Genus: no <i>Salmonella</i> (AQUAMIS), fraction of loci found in cgMLST below 95% (86%) - ChewieSnake	Yes
23-SNPa	Yes*		No
25-cgMLST	No	<i>Citrobacter freundii</i>	Yes
26-cgMLST	No	The 89% of the reads are classified as <i>Citrobacter</i> sp.	Yes
27-cgMLST	No	ID species = 100% <i>Citrobacter freundii</i>	Yes
28-cgMLST	No	Not <i>Salmonella enterica</i>	Yes
28-SNPr	No	Not <i>Salmonella enteritidis</i>	Yes
30-cgMLST	No	Different species: <i>Citrobacter freundii</i>	Yes
85-cgMLST	No	Not a <i>Salmonella</i> , detected <i>Citrobacter freundii</i>	Yes
85-SNPr	No	Not a <i>Salmonella</i> , detected <i>Citrobacter freundii</i>	Yes

^{a)} Based on the submitted Distance matrix.

*Comment Lab 23: Strain 23SCA12 was included in the cluster analysis as it passed in the quality control. However this strains is not *Salmonella enterica* but was identified by kmerFinder as *Citrobacter freundii*.

Lab code	Strain 23SCA14 passed QC	Reason(s) not passing QC	Strain 23SCA14 excluded from cluster analysis ^{a)}
Expected	No	Wet mixture of 23SCA05 REF S. Bovismorbificans and 23SCA12 <i>Citrobacter freundii</i>	Yes
3-SNPr	No	Per sequence GC content	No
4-cgMLST	No	not passing contamination/species check - contaminated, also very large assembly (total length)	Yes
5-cgMLST	No	number of contigs, total length, contamination	Yes
8-cgMLST	No	contamination with <i>Citrobacter freundii</i>	Yes
9-cgMLST	No	<i>Citrobacter</i> contamination	Yes
11-cgMLST	No	Total length too big, low GC%, too many contigs, low N50	Yes
14-cgMLST	No	Failed on Genome Length. Kmer Finder confirmed contamination.	Yes
15-cgMLST	No	Total length of assembly incorrect >5 Mb + number of contigs>150 + contamination : <i>Citrobacter</i>	Yes
16-cgMLST	No	Contamination too high, contaminated with <i>Citrobacter</i> , Total length assembly too high	Yes
19-cgMLST	No	assembly length (9722959), number of contigs >1000: 314	No
20-SNPr	No	Contaminated: Assembly > 5.8Mbp and large portion is <i>Citrobacter</i>	Yes
21-cgMLST	No	Genome size 9,7MBp and Identified as mix of <i>Citrobacter</i> and <i>Salmonella</i> by kmerfinder	Yes
22-cgMLST1	No	total length >9000000kb, Inter species contamination (3750 SNVs), read and contig fraction majority genus below cut-off	Yes
22-cgMLST2	No	total length >9000000kb, Inter species contamination (3750 SNVs), read and contig fraction majority genus below cut-off	Yes
22-SNPr	No	total length >9000000kb, Inter species contamination (3750 SNVs), read and contig fraction majority genus below cut-off	Yes
23-SNPa	No	contaminated; <i>Citrobacter freundii</i> and <i>Salmonella enterica</i>	Yes
25-cgMLST	No	contamination > 10% - <i>Citrobacter freundii</i>	Yes
26-cgMLST	No	Assembly exceeds 5.5Mb and is contaminated with <i>Citrobacter</i> sp. (12% of the reads)	Yes
27-cgMLST	No	Mixed culture was detected with <i>Citrobacter freudii</i> >2%	Yes
28-cgMLST	Yes		No
28-SNPr	Yes		No
30-cgMLST	No	Double the assembly length, failed cgMLST	Yes
85-cgMLST1	No	Contamination by 13% of <i>Citrobacter freundii</i>	Yes
85-SNPr	No	Contamination by 13% of <i>Citrobacter freundii</i>	Yes

^{a)} Based on the submitted Distance matrix.

Lab code	Strain 23SCA16 passed QC	Reason(s) not passing QC	Strain 23SCA16 excluded from cluster analysis ^{a)}
Expected	No	Wet mixture of 23SCA05 REF <i>S. Bovismorbificans</i> and 23SCA01 <i>S. Typhimurium</i>	Yes
3-SNPr	Yes		No
4-cgMLST	Yes		No
5-cgMLST	Yes*		No
8-cgMLST	No	Contamination with <i>Klebsiella pneumoniae</i>	Yes
9-cgMLST	No	14% contamination	Yes
11-cgMLST	No	too large total length, low GC%, a lot of contigs, low N50	Yes
14-cgMLST	No	Failed on Genome Length. Kmer Finder confirmed contamination.	Yes
15-cgMLST	No	Number of contig >150 + suspected contamination with other serotypes of <i>Salmonella</i>	Yes
16-cgMLST	No	Indication of multiple serotypes	No
19-cgMLST	Yes*		No
20-SNPr	No	Contaminated: multiple serotypes (<i>Bovismorbificans</i> and <i>Bsilla</i>)	Yes
21-cgMLST	No	Genome size 5,4Mbp	Yes
22-cgMLST1	No	total length above threshold, Intra species contamination (340 SNVs)	Yes
22-cgMLST2	No	total length above threshold, Intra species contamination (340 SNVs)	Yes
22-SNPr	No	total length above threshold, Intra species contamination (340 SNVs)	Yes
23-SNPa	No	N. contigs>500; Total length higher than expected	Yes
25-cgMLST	Yes		No
26-cgMLST	Yes		No
27-cgMLST	No	number of contigs = 443 (warning) but all the other criteria are OK	No
28-cgMLST	Yes		No
28-SNPr	Yes		No
30-cgMLST	Yes		No
85-cgMLST	No	Contaminated with other <i>Salmonella</i> , Out of EFSA-recommended thresholds: 5,5Mb (max 5,3Mb); 190 contigs (max 150)	Yes
85-SNPr	No	Contaminated with other <i>Salmonella</i> , Out of EFSA-recommended thresholds: 5,5Mb (max 5,3Mb); 190 contigs (max 150)	Yes

^{a)} Based on the submitted distance matrix.

*Comment Lab 5: extra care should be taken with the conclusion of Strain 23SCA16 since the threshold of 10% missing loci was almost passed.

*Comment Lab 19: 23SCA16 did not completely pass the quality check (contigs >1000: 154, N50: 99482), but was so close, that it was processed for further analyses.

Annex 7. Per submission, the participants' distance matrix data

In blue: Deviation from the expected result.

Labcode-method	Strain code													
	23 SCA-REF	23 SCA01	23 SCA02	23 SCA03	23 SCA04	23 SCA05	23 SCA06	23 SCA11	23 SCA12	23 SCA13	23 SCA14	23 SCA15	23 SCA16	
	Ref Bm	STM	Ref2 Bm	Bm	Bm	Ref Bm	Bm	Bm	Citro	SE	Ref+Citro	Bm	Ref+STM	
EL-cgMLST	0	2595	1	56	1	1	128	60		2635		2		
4-cgMLST	0	2383	0	35	0	0	101	46		2459		0	0	
5-cgMLST ^{a)}	0	2472	0	57	0	0	135	58		2504		3	4	
8-cgMLST	0	2576	5	54	1	0	129	57		2624		2		
9-cgMLST ^{b)}	0		1	65	2	1	151	69				4		
11-cgMLST	0		1	59	3	1	135	61				3		
14-cgMLST ^{c)}	0	200	0	54	0	0	134	60		200		1		
15-cgMLST	0		0	55	0	0	128	58				1		
16-cgMLST	0	2595	1	56	1	1	128	60		2635		2	4	
19-cgMLST ^{d)}	0	2425	0	52	0	0	124	57		2442	14	0	1	
21-cgMLST	0	2559	0	53	0	0	127	57		2562		3		
22-cgMLST1	0	2569	1	53	1	1	128	57		2605		2		
22-cgMLST2	0	2875	1	61	2	1	147	66		2922		3		
25-cgMLST ^{h)}	0	2599	0	55	0	1	125	58		2641		1	0	
26-cgMLST	0	2822	0	19	0	0	96	32		2862		0	0	
27-cgMLST	0	2595	0	55	0	0	127	58		2622		1	0	
28-cgMLST	0	2548	3	58	2	3	132	62		2592	1	6	4	
30-cgMLST	0	2595	1	56	1	1	128	59		2635		2	1	
85-cgMLST	0	2594	1	55	1	0	127	58		2630		1		
23-SNP ^{a)}	0	11170	3	46	2	2	101	52	12277	11403		6		
3-SNP ^r	0	8	0	2	0	0	7	4	2971	447	0	0	0	
20-SNP ^{r^{c)}}	0		0	111	2	0	241	103				6		
22-SNP ^{r^{f)}}	0		0	141	2	1	550	245				7		
28-SNP ^r	0	4487	0	55	1	0	137	46		21385	0	6	0	
85-SNP ^{rⁱ⁾}	0	39206	1	269	4	0	671	645		39141		9		

The 4 clustering strains indicated in green. The 3 strains expected to be excluded indicated in orange.

Empty cell: strain not included in the lab's distance matrix.

Grey cell: strain excluded from the lab's distance matrix because of not being serovar Bovismorbificans.

Numbers in **bold** were reported to be clustering with the REF strain.

- a) Comment Lab 5: extra care should be taken with the conclusion of Strain 23SCA16 since the threshold of 10% missing loci was almost passed.
- b) Comment Lab 9: Samples not identified as *S. Bovismorbificans* were also excluded from the distance matrix.
- c) Comment Lab 14: 23SCA16: Exceeded our quality parameter for genome length by only 3,607bp.
- d) Comment Lab 19: 23SCA16 did not completely pass the quality check (contigs >1000: 154, N50: 99482), but was so close, that it was processed for further analyses.
- e) Comment Lab 20: Isolate 23SCA15 is close to the reference cluster, though may represent an isolate with shared lineage but not the same outbreak.
- f) Comment Lab 22: For SNP analysis, only isolates that were typed as *S. Bovismorbificans* ST142 and fulfilled all QC criteria were included in the analysis.
- g) Comment Lab 23: Strain 23SCA12 was included in the cluster analysis as it passed in the quality control. However this strains is not *Salmonella enterica* but was identified by kmerFinder as *Citrobacter freundii*.
- h) Comment Lab 25: For the time being, *S. Bovismorbificans* is not sequenced on a routine basis. Assuming that strain 23SCA15 is phenotypically fully susceptible, whereas the reference strain is multi-resistant, we probably (depending on the epidemiological information available) would not have it included in the sequencing / outbreak investigation.
- i) Comment Lab 85: For SNP analysis between 0 and 9 SNP of difference. Our threshold is at 21 SNP of difference for a cluster (Pightling et al., 2018). Strain 23SCA16 is identified also *Bovismorbificans* and <21 SNP of difference, but the analysis is unsure. In a case of outbreak we could make a focus on this strain also and ask to re-sequence it. As it don't pass our QC test, we haven't add it to the cluster analysis.