Interim Summary Report
EURL-*Salmonella* Proficiency Test Cluster Analysis 2024

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

19 May 2025
Z&O letter report 2025-0043

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.

EURL Salmonella

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 2024, concerning the optional part on Cluster Analysis (CA).
A total of 24 participants took part in the WGS-based cluster analysis. The full results will be reported in more detail in the final report on the EURL-*Salmonella* PT Typing 2024.

2. *Salmonella* strains for cluster analysis

A total of six *Salmonella* strains (24SCA01 – 24SCA06) in HI agar transport tubes were sent to the participants in the EURL-*Salmonella* PT Typing 2024, 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 (24SCA11 – 24SCA16) 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 2024*

Strain code	Serovar	Abbreviation	ST	Origin
24SCA01	Infantis	Inf	32	Human
24SCA02	Infantis	Inf=REF	32	Broilers
24SCA03[a)]	Infantis	Inf	32	Broilers
24SCA04	Infantis	Inf	32	Human
24SCA05[a)]	Infantis	Inf	32	Broilers
24SCA06	Infantis	Inf	32	Broilers

a) Technical duplicates

Table 1B. Background information on the 'dry' Salmonella *strains used for cluster analysis in 2024*

Strain code	Serovar	Abbreviation	ST	Origin
24SCA11	Infantis	Inf	32	Broilers
24SCA12[b)]	Mixed strains	Inf+STM		
24SCA13	Infantis	Inf	32	Human
24SCA14	Infantis	Inf=REF	32	Broilers
24SCA15	Infantis	Inf	32	Broilers
24SCA16[c)]	Mixed strains	Inf+Java		

b) Wet-mixed culture of strains 24SCA02 (Inf=REF) and 19SCA03 (*S.* Typhimurium)
c) Wet-mixed culture of strains 24SCA03 (Inf) and 2013S16 (*S.* Paratyphi B, var. Java)

In preparation of the PT 2024 on cluster analysis, five broiler strains from a research project (Mughini-Gras et al., 2021) and three human surveillance strains were re-cultured from storage (2018/2019 and 2023 respectively). All *Salmonella* strains were submitted for WGS analysis both directly (July 2024) and after sub-culturing for ten times on blood-agar and BPW (August 2024). In addition, several wet-mixed cultures were included in this pre-testing. Subsequently, six 'wet' strains and six 'dry' strains were selected for inclusion in the PT 2024 (also see Figure 1). One set of wet technical duplicates was included: the shipment tubes of strain 24SCA03 and strain 24SCA05 were both prepared from the same blood-agar plate containing strain 24SCA03.
Like in previous years, the PT Cluster Analysis 2024 was mimicking an outbreak situation, with a *Salmonella* Infantis as the reference strain (24SCA-REF). Raw WGS data of this strain (24SCA-REF_R1.fastq.gz and 24SCA-REF_R2.fastq.gz, as well as their md5 checksums) were also made available through the RIVM secure ftp server.
For this particular PT 2024, 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 2024).

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

Twenty-four participants (Table 2) submitted a total of 27 cluster analysis results based on WGS data; two participants submitted cgMLST-based as well as SNP-based data results, and one participant submitted cgMLST-based and wgMLST-based data analyses. Some details on the sequencing and analysis procedures performed by the participants are given in Annex 1. The pre-test 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://github.com/RIVM-bioinformatics/juno-assembly), 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*. Infantis strains.

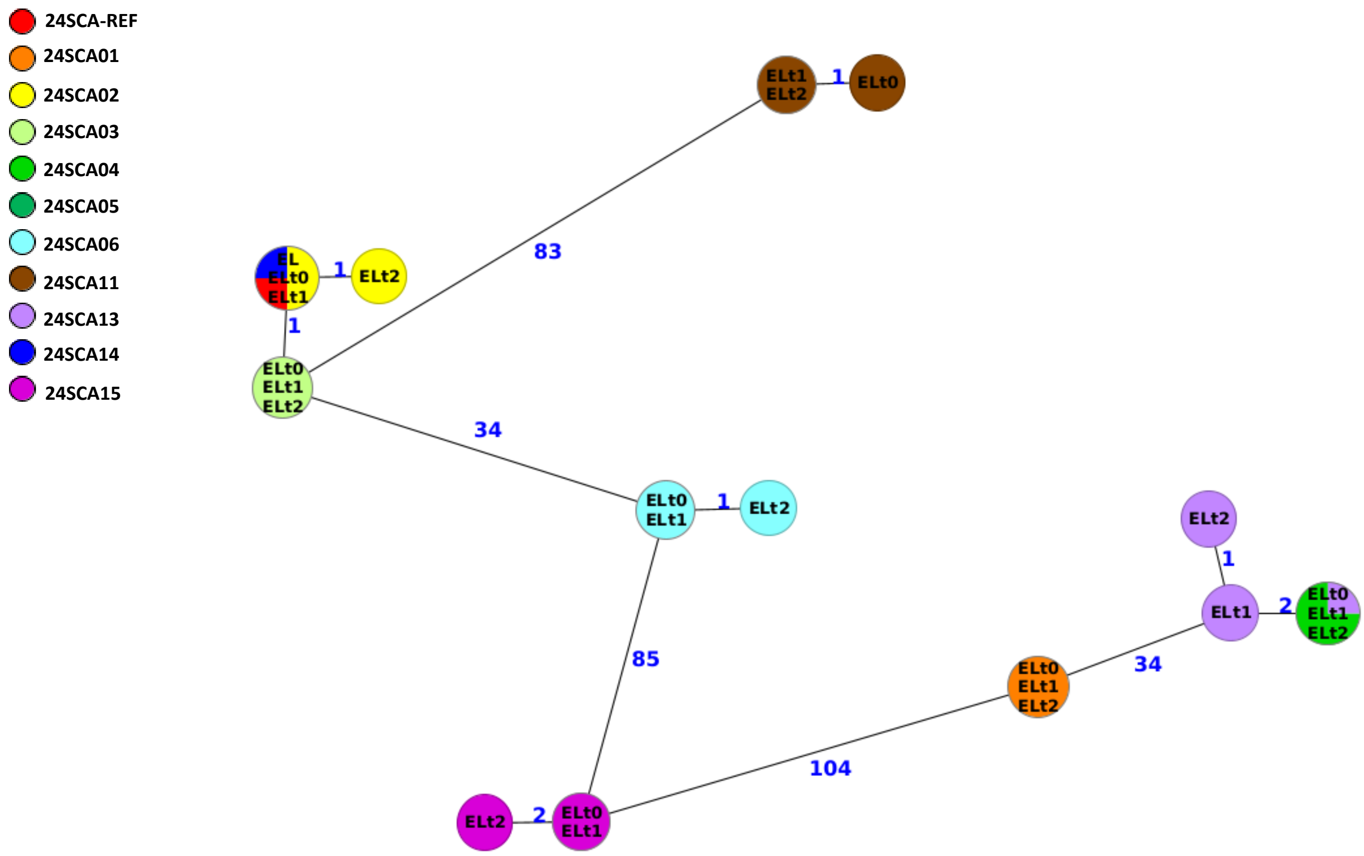


*Figure 1. MST of the EURL-*Salmonella *(EL) pre-tests (Ridom SeqSphere$^+$, cgMLST (3002), pairwise ignoring missing values and QC-failing strains excluded)*

EL: One data set, no testing in time;
ELt0: Original WGS data from the stored poultry research and human surveillance Salmonella *strains;*
ELt1: WGS data from initial pre-testing for the PT (July 2024);
ELt2: WGS data after ten times sub-culturing (blood-agar/BPW) for the PT (August 2024).

Fourteen compressed paired-end fastq files (strains 24SCA11 - 24SCA16 plus 24SCA-REF) had to be downloaded for 'dry' lab analysis from the RIVM secure ftp server. The md5 checksums for these files were available on the server as well (Annex 2). Participants were asked whether they checked the md5sum values after downloading, and 23 participants indicated that they had done this. Nineteen of those 23 participants also submitted the 'dry' strain md5sum values, which were then checked by the EURL-*Salmonella* again (Annex 3).
Twenty-two participants submitted the md5sum values of their 'wet' strain analysis before uploading. After downloading by the EURL-*Salmonella*, the md5sums were determined and checked for consistency with the participants' data (Annex 3).

All participants' raw data (compressed fastq files) for the six 'wet' strains (24SCA01 – 24SCA06) were successfully processed through the EURL-*Salmonella* in-house Juno-assembly pipeline as mentioned before. The *de novo* assembled genomes (fasta files) were analysed in Ridom SeqSphere⁺, using the cgMLST Enterobase v2.0 and visualised in a MST, this also included the 'dry' strain data of 24SCA-REF, 24SCA11, 24SCA13, 24SCA14 and 24SCA15 (Figure 2). Data per 'wet' strain are given in Annex 4.

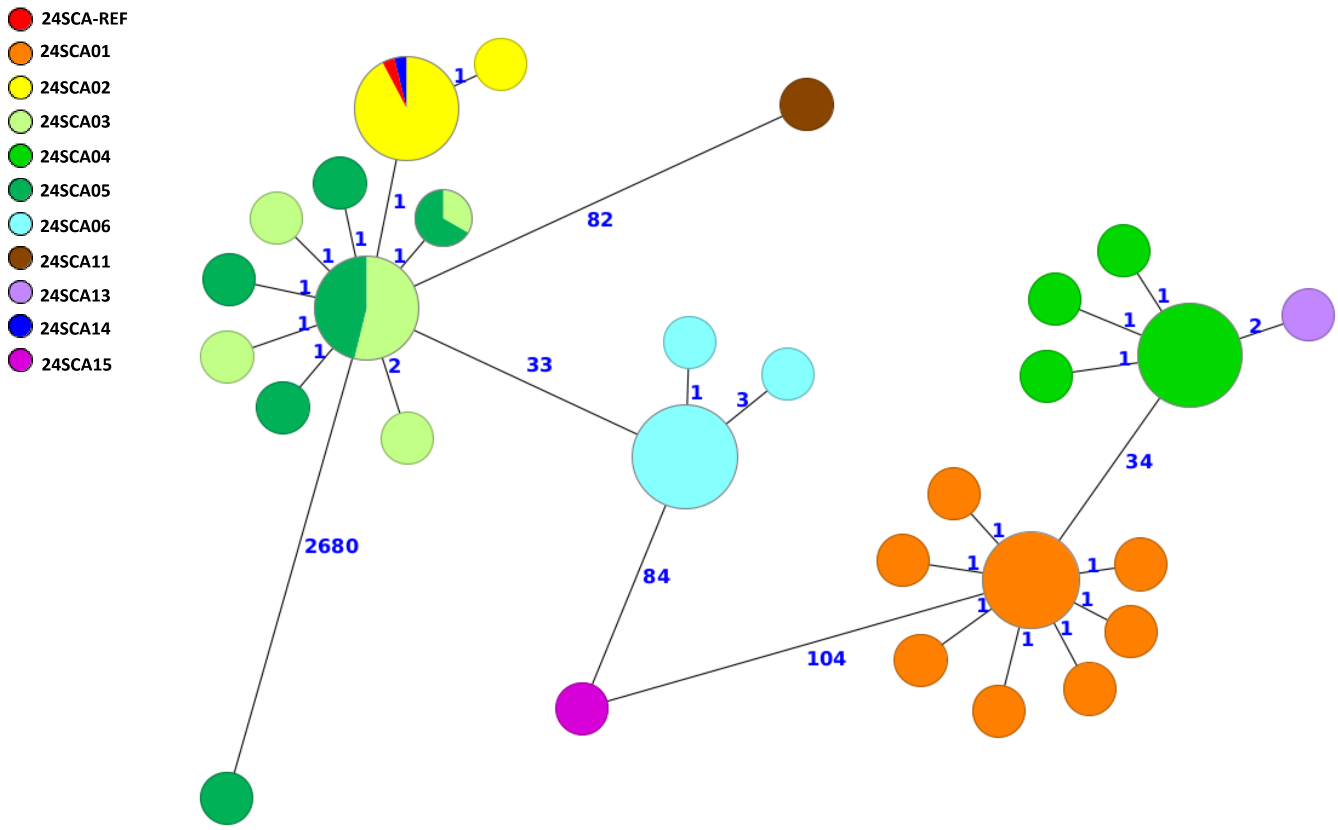


Figure 2. MST of the strains from the participants' processed raw data plus the 'dry' strain data (24SCA-REF, 24SCA11, 24SCA13, 24SCA14, 24SCA15) (Ridom SeqSphere⁺, cgMLST (3002), pairwise ignoring missing values)

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 2024 (24SCA-REF) was found or not.

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

Strains 24SCA12 and 24SCA16 were expected to not pass the QC of the participants, because of their characteristics (Table 1B). The PT Typing 2024 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 24SCA12 (wet mixture of serovars Infantis and Typhimurium) was reported not passing the QC criteria by 23 of the 24 participants, and these 23 participants excluded this strain from their distance matrix.
Strain 24SCA16 (wet mixture of serovars Infantis and Paratyphi B, var. Java) was reported not passing the QC criteria by all 24 participants and was excluded from all distance matrices.
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 2024 situation was set at maximum five allelic differences from the reference sequence. Based on this cgMLST-based criterium, results were expected to indicate the 'wet' strains 24SCA02 (reference strain), technical duplicate strains 24SCA03 and 24SCA05 (clustering with the reference strain), and the 'dry' strain 24SCA14 (reference strain) to form a cluster with the provided reference outbreak strain 24SCA-REF data (also see Figure 1).
Twenty-three of the 27 submissions (three participants with multiple submissions) reported those four strains to be part of the cluster. Lab 24 also included mixed serovars 'strain' 24SCA12 into this cluster. Consequently, 22 of the 27 submissions reported the cluster completely as expected (Table 2).
Technical duplicate strains 24SCA03 and 24SCA05 were reported within one cluster in 24 of the 27 submissions.
Optionally, any other cluster(s) apart from the reference could be submitted. Seventeen of 21 submissions correctly reported strains 24SCA04 and 24SCA13 to form a second cluster.

In addition, participants were asked if their NGS analysis included identification of AMR markers. Twelve participants submitted data on this topic, to be discussed in more detail in the final report on the EURL-*Salmonella* PT Typing 2024.

Table 2. Expected cluster analysis results and those as reported by the 24 participants

Labcode-method	Strain code											
	24 SCA01	24 SCA02	24 SCA03	24 SCA04	24 SCA05	24 SCA06	24 SCA11	24 SCA12	24 SCA13	24 SCA14	24 SCA15	24 SCA16
	Inf	Inf=REF	Inf	Inf	Inf	Inf	Inf	Inf+STM	Inf	Inf=REF	Inf	Inf+Java
Expected	**No**	**Yes**	**Yes**	**No**	**Yes**	**No**	**No**	**n.a.**	**No**	**Yes**	**No**	**n.a.**
1-cgMLST	No	Yes	Yes	No	Yes	No	No	n.a.	No	Yes	No	n.a.
2-cgMLST	No	Yes	Yes	No	Yes	No	No	n.a.	No	Yes	No	n.a.
3-cgMLST	No	No	Yes	Yes	No	Yes	No	n.a.	No	Yes	No	n.a.
4-cgMLST	No	Yes	Yes	No	Yes	No	No	n.a.	No	Yes	No	n.a.
5-cgMLST	No	Yes	Yes	No	Yes	No	No	n.a.	No	Yes	No	n.a.
7-cgMLST	No	Yes	Yes	No	Yes	No	No	n.a.	No	Yes	No	n.a.
11-cgMLST	No	Yes	Yes	No	Yes	No	No	n.a.	No	Yes	No	n.a.
12-cgMLST[a)]	No	Yes	Yes	No	Yes	No	No	n.a.	No	Yes	No	n.a.
13-cgMLST[b)]	No	Yes	Yes	No	Yes	No	No	n.a.	No	n.a.	No	n.a.
14-cgMLST	No	Yes	Yes	No	Yes	No	No	n.a.	No	Yes	No	n.a.
16-cgMLST[c)]	No	Yes	Yes	No	Yes	No	No	n.a.	No	Yes	n.a.	n.a.
17-cgMLST	No	Yes	Yes	No	Yes	No	No	n.a.	No	Yes	No	n.a.
17-wgMLST	No	Yes	Yes	No	Yes	No	No	n.a.	No	Yes	No	n.a.
19-SNPa[d)]	No	No	No	No	Yes	No	No	n.a.	No	No	No	n.a.
20-cgMLST	No	Yes	Yes	No	Yes	No	No	n.a.	No	Yes	No	n.a.
21-SNPr	No	Yes	Yes	No	Yes	No	No	n.a.	No	Yes	No	n.a.
23-cgMLST	No	Yes	Yes	No	Yes	No	No	n.a.	No	Yes	No	n.a.
23-SNPr	No	Yes	Yes	No	Yes	No	No	n.a.	No	Yes	No	n.a.
24-cgMLST	No	Yes	Yes	No	Yes	No	No	Yes	No	Yes	No	n.a.
27-SNPr	No	Yes	Yes	No	Yes	No	No	n.a.	No	Yes	No	n.a.
28-cgMLST	No	Yes	Yes	No	Yes	No	No	n.a.	No	Yes	No	n.a.
29-cgMLST	No	Yes	Yes	No	Yes	No	No	n.a.	No	Yes	No	n.a.
31-SNPr	No	Yes	Yes	No	Yes	No	No	n.a.	No	Yes	No	n.a.
32-cgMLST	No	Yes	Yes	No	Yes	No	No	n.a.	No	Yes	No	n.a.
63-SNPr[e)]	No	Yes	Yes	No	n.a.	No	No	n.a.	No	Yes	No	n.a.
80-cgMLST	No	Yes	Yes	No	Yes	No	No	n.a.	No	Yes	No	n.a.
80-SNPa	No	Yes	Yes	No	Yes	No	No	n.a.	No	Yes	No	n.a.

Yes/No: Wether or not a cluster with the reference strain in the EURL-*Salmonella* PT Typing 2024 (24SCA-REF) was found
n.a.: Not applicable (QC not passed)

a) Comment Lab 12: We would have repeated NGS of strain 24SCA12 as it has an AD=3 to the reference strain, but failed QC.
b) Comment Lab 13: 24SCA12 included *S*. Infantis which clustered with the REF strain but was contaminated with other *Salmonella*. 24SCA14 included *S*. Infantis which clustered with REF strain but was contaminated slightly with *E. coli*. 24SCA16 included *S*. Paratyphi B and some other *Salmonella*. All of them were excluded from the final cluster analysing.
c) Comment Lab 16: Strain 24SCA15: Intraspecies contamination with 37 SNVs, excluded from analysis/distance matrix.
d) Comment Lab 19: Strain 24SCA02, Strain 24SCA03 and 24SCA0314 shared 11 SNPs.
e) Comment Lab 63: Data set passed QC criteria for *Salmonella*, but was excluded from reference-based SNP-analysis as it is a more distant sequence type.

In blue: Deviation from the expected result.

List of abbreviations

AMR	Antimicrobial Resistance
BPW	Buffered peptone water
CA	Cluster Analysis
cgMLST	core genome Multilocus Sequence Typing
EFSA	European Food Safety Authority
EL	EURL-*Salmonella* Laboratory
EU	European Union
EURL-*Salmonella*	European Union Reference Laboratory for *Salmonella*
ftp	file transfer protocol
HI agar	Hearth Infusion ager (in transport tubes)
Inf	*Salmonella* Infantis
Java	*S*. Paratyphi B, var. Java
MST	Minimum Spanning Tree
n.a.	not applicable
NGS	Next Generation Sequencing
NRLs-*Salmonella*	National Reference Laboratories for *Salmonella*
PT	Proficiency Test
QC	Quality Control
REF	Reference
RIVM	National Institute for Public Health and the Environment
SNPa	assembly-based Single-Nucleotide Polymorphism data
SNPr	reference-based Single-Nucleotide Polymorphism data
ST	Sequence Type
STM	*Salmonella* Typhimurium
wgMLST	whole genome Multilocus Sequence Typing
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 5-5-2025).

Mughini-Gras L, AHAM van Hoek, T Cuperus, et al., Prevalence, risk factors and genetic traits of *Salmonella* Infantis in Dutch broiler flocks. Vet Microbiol. 2021. doi: 10.1016/j.vetmic.2021.109120.

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 24 participants and the EURL-*Salmonella*

Labcode	Wet lab[a)]	WGS platform[b)]	Data analysis[c)]	Tool for analysis	Method for cluster analysis[d)]
EURL-Salm	In-In-In	NextSeq	cgMLST	Ridom SeqSphere	MST
1	In-In-In	MiSeq	cgMLST	In-house pipeline with Chewbbaca, scheme from Enterobase	Single linkage
2	In-In-In	NextSeq	cgMLST	inhouse ChewieSnake Pipeline (scheme: Enterobase 3000 loci)	single linkage hierarchical clustering
3	In-In-In	NextSeq	cgMLST	Ridom SeqSphere	MST
4	In-In-In	MiSeq	cgMLST	in-house galaxy	MSTreeV2
5	In-In-In	MiSeq	cgMLST	Ridom SeqSphere	MST
7	In-In-In	NextSeq	cgMLST	Ridom SeqSphere	UPGMA Tree
11	In-In-In	MiniSeq	cgMLST	Ridom SeqSphere	Distance matrix only
12	In-In-In	MiSeq	cgMLST	Ridom SeqSphere	MST
13	In-In-In	MiSeq	cgMLST	Ridom SeqSphere	MST
14	In-In-In	MiSeq	cgMLST	chewBBACA + Grapetree	MSTree V2 - Grapetree's modified version of MST
16	In-In-In	MiSeq	cgMLST	chewBBACA	MST
17-cgMLST	In-In-In	MiSeq	cgMLST	Ridom SeqSphere	NJ
17-wgMLST	In-In-In	MiSeq	wgMLST	chewBBACA	UPGMA
19	In-In-In	MiSeq	SNPa	CSIPhylogeny 1.4	ML
20	In-In-In	NextSeq	cgMLST	Ridom SeqSphere	MST
21	In-In-Out	NextSeq	SNPr	MINTyper 1.0	NJ
23-cgMLST	In-In-In	MiSeq	cgMLST	chewBBACA and chewTree on Galaxy ARIES	chewTree
23-SNPr	In-In-In	MiSeq	SNPr	CSIPhylogeny on https://cge.food.dtu.dk/	ML
24	In-In-In	NextSeq	cgMLST	Ridom SeqSphere	MST
27	In-In-In	NextSeq	SNPr		ML
28	In-In-In	MiSeq	cgMLST	linux command line	NJ
29	In-In-In	MiSeq	cgMLST	chewBBACA and ReporTree	Hierarchical clustering, average linkage
31	In-In-In	MiSeq	SNPr	Galaxy Samtools	ML
32	In-In-In	MiniSeq	cgMLST	Ridom SeqSphere	MST
63	In-Out-Out	NovaSeq	SNPr	bwa, gatk, custom scripts, raxml	ML
80-cgMLST	In-Out-Out	NextSeq	cgMLST	Ridom SeqSphere	MST
80-SNPa	In-Out-Out	NextSeq	SNPa	SNIPPY	ML

a) Wet lab preparations: DNA extraction, Library preparation, Sequencing. In: In-house, Out: Outsourced.
b) All Illumina platforms.
c) SNPa: assembly-based, SNPr: reference-based.
d) ML:Maximum Likelihood, MST: Minimum Spanning Tree, NJ: Neighbor Joining.

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 ‘ALL-md5’)

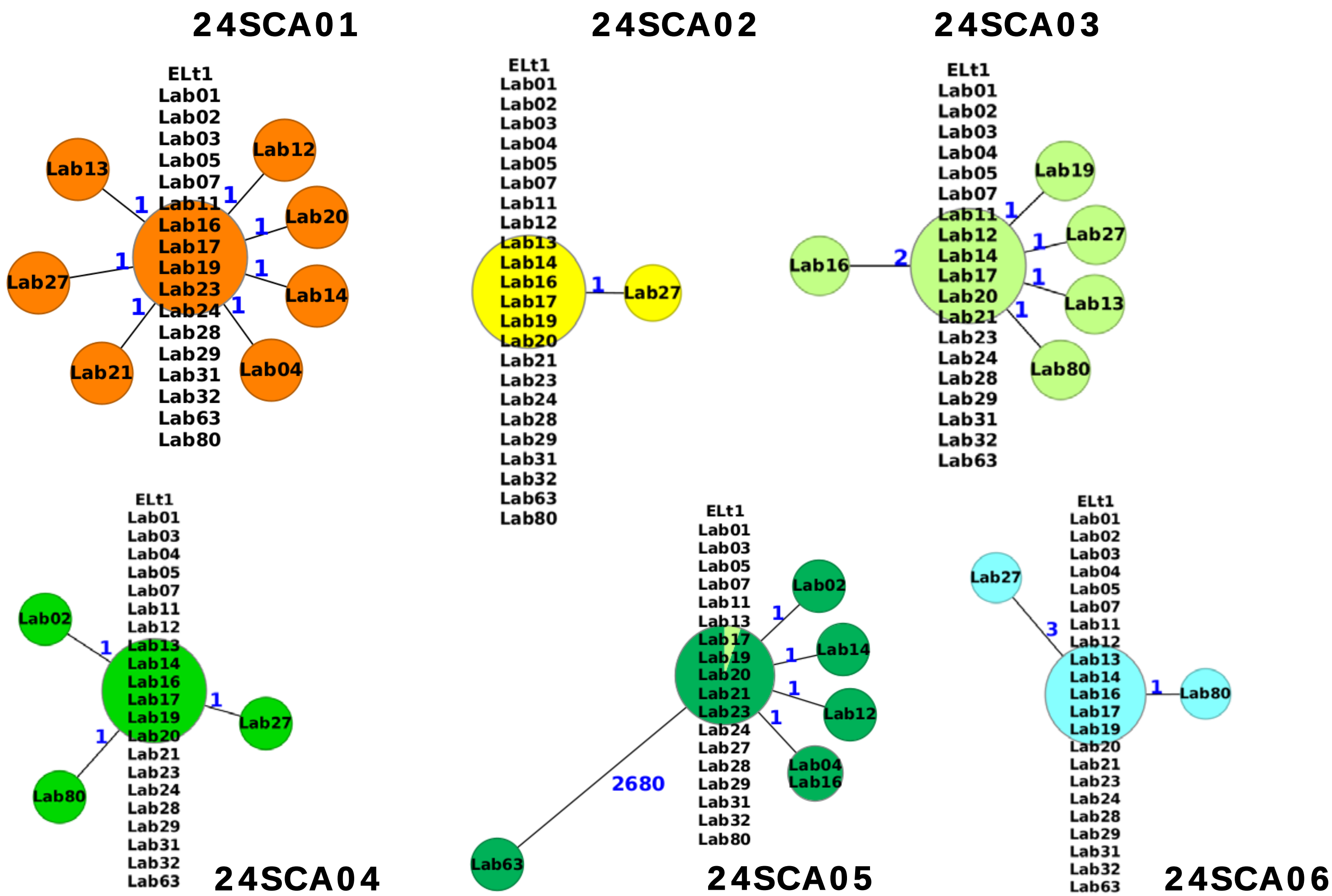
```
6bf05a577ba9f88494866573553bdb6c    24SCA11_R1.fastq.gz
5a71cd0c82cca8aae46e6f8a1def44e0    24SCA11_R2.fastq.gz
19add44955492a91cccc138439f22c09    24SCA12_R1.fastq.gz
ac240654acd8a1a615d82b5a94f6777c    24SCA12_R2.fastq.gz
2d1b5a7240c8a5ca80a59eff558700a8    24SCA13_R1.fastq.gz
06fc5d9751305b0288d7ab036d680d73    24SCA13_R2.fastq.gz
fdecdbdeb0a5a6000871ea474cbd8534    24SCA14_R1.fastq.gz
5f84a601e9b6f5d63ca2d588cdf70b5e    24SCA14_R2.fastq.gz
02b6951c938daa0b563342d9fcfa9b3f    24SCA15_R1.fastq.gz
44ec5575e2cd84b607994b8d2098be59    24SCA15_R2.fastq.gz
d9e3a11c52344e5317f6e54c79e5a2be    24SCA16_R1.fastq.gz
7b4b4de7c0ea5dab00bbe60316323b08    24SCA16_R2.fastq.gz
383599f27b18687e9ebd88069d40182c    24SCA-REF_R1.fastq.gz
db3b146bb2ecb3dfe9db8450b37fe73b    24SCA-REF_R2.fastq.gz
```

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

	'Dry' strains 24SCA11 - 24SCA16			'Wet' strains 24SCA01 - 24SCA06	
Labcode	**Did you check the md5sum values of the downloaded files**	**Lab's md5sum data submitted**	**Check by the EURL-*Salmonella***	**Lab's md5sum data submitted**	**Md5sum determination and check by the EURL-*Salmonella* after downloading**
1	Yes	Yes	OK	Yes	OK
2	Yes	Yes	OK	Yes	OK
3	Yes	Yes	OK	Yes	OK
4	Yes	Yes	OK	Yes	OK
5	Yes	Yes	No*	Yes	No*
7	Yes	Yes	OK, but missing data or mismatch for 5/14 files	Yes	OK
11	No	No	n.a.	No	n.a.
12	Yes	Yes	OK	Yes	OK
13	Yes	Yes	OK	Yes	OK
14	Yes	Yes	OK	Yes	OK
16	Yes	Yes	OK	Yes	OK
17	Yes	No	n.a.	Yes	OK
19	Yes	Yes	OK	Yes	OK
20	Yes	Yes	OK, but mismatch for 24SCA16_R1	Yes	OK
21	Yes	No	n.a.	Yes	OK
23	Yes	Yes	OK	Yes	OK
24	Yes	Yes	OK	Yes	OK
27	Yes	Yes	OK	Yes	OK
28	Yes	Yes	OK	Yes	OK
29	Yes	Yes	OK	Yes	OK
31	Yes	No	n.a.	No	n.a.
32	Yes	No	n.a.	Yes	OK
63	Yes	Yes	OK	Yes	OK, but data for strain 24SCA01 only
80	Yes	Yes	OK	Yes	OK

n.a.: not applicable, OK: matching data.
No*: md5sum values of 40 digits were given, therefore these were not comparable to the EURL-*Salmonella* md5sum values of 32 digits.

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 4, 11, and 24.

In blue: Deviation from the expected result.

		Strain code					
Lab code	**Serotyping method(s) used**	**24SCA01**	**24SCA02**	**24SCA03**	**24SCA04**	**24SCA05**	**24SCA06**
EL	**In-house pipeline**	**Infantis**	**Infantis**	**Infantis**	**Infantis**	**Infantis**	**Infantis**
1	SISTR	Infantis	Infantis	Infantis	Infantis	Infantis	Infantis
2	Classical serology, SISTR v 1.1.1	Infantis	Infantis	Infantis	Infantis	Infantis	Infantis
3	Seqsero2 (v1.1.1)	Infantis	Infantis	Infantis	Infantis	Infantis	Infantis
5	RIdom SeqSphere+	Infantis, Virchow	Infantis, Virchow	Infantis, Virchow	Infantis, Virchow	Infantis, Virchow	Infantis, Virchow
7	not applicable						
12	Classical serology, Sistr	Infantis	Infantis	Infantis	Infantis	Infantis	Infantis
13	Classical serology	Infantis	Infantis	Infantis	Infantis	Infantis	Infantis
14	SeqSero2	Infantis	Infantis	Infantis	Infantis	Infantis	Infantis
16	NGS, sistr	Infantis	Infantis	Infantis	Infantis	Infantis	Infantis
17	WGS - SeqSero2, SISTR Geno-serotyping	Infantis	Infantis	Infantis	Infantis	Infantis	Infantis
19	Classical serology, SSP and Galaxy sistr	Infantis O7:r:1,5	Infantis O7:r:1,5	Infantis O7:r:1,5	Infantis O7:r:1,5	Infantis O7:r:1,5	Infantis O7:r:1,5
20	Seqsero2 + SISTR1	Infantis	Infantis	Infantis	Infantis	Infantis	Infantis
21	SeqSero 1.2	Infantis	Infantis	Infantis	Infantis	Infantis	Infantis
23	SeqSero_v2	Infantis	Infantis	Infantis	Infantis	Infantis	Infantis
27	MOST, SeqSero2, SISTR	Infantis	Infantis	Infantis	Infantis	Infantis	Infantis
28	WGS SeqSero serotype finder	Infantis	Infantis	Infantis	Infantis	Infantis	Infantis
29	SeqSero2 and Sistr	Infantis	Infantis	Infantis	Infantis	Infantis	Infantis
31	Galaxy Sciensano	Infantis	Infantis	Infantis	Infantis	Infantis	Infantis
32	Based on genome, with sistr software	Intis	Intis	Intis	Intis	Intis	Intis
63	SeqSero2	Infantis	Infantis	Infantis	Infantis	Typhimurium	Infantis
80	SeqSero2 + Sistr	Infantis	Infantis	Infantis	Infantis	Infantis	Infantis

		Strain code					
Lab code	**Serotyping method(s) used**	**24SCA11**	**24SCA12**	**24SCA13**	**24SCA14**	**24SCA15**	**24SCA16**
EL	**In-house pipeline**	**Infantis**	**failed QC**	**Infantis**	**Infantis**	**Infantis**	**failed QC**
1	SISTR	Infantis	Infantis (but with failed QC status in SISTR pipeline)	Infantis	Infantis	Infantis	Contaminated, NT
2	SISTR v 1.1.1	Infantis	QC Fail	Infantis	Infantis	Infantis	QC Fail
3	Seqsero2 (v1.1.1)	Infantis	Lagos	Infantis	Infantis	Infantis	Paratyphi B var. Java
5	Ridom SeqSphere+	Infantis, Virchow	QC failed, >30 cgMLST330 loci missing	Infantis, Virchow	Infantis, Virchow	Infantis, Virchow	QC failed, >30 cgMLST330 loci missing
7	Ridom SeqSphere+	Infantis	NA (Failed QC)	Infantis	Infantis	Infantis	NA (Failed QC)
12	Sistr	Infantis	Infantis	Infantis	Infantis	Infantis	?
13	SeqSero	Infantis	Infantis	Infantis	Infantis	Infantis	Paratyphi B
14	Seqsero2	Infantis	contaminated	Infantis	Infantis	Infantis	contaminated
16	NGS, sistr	Infantis		Infantis	Infantis		
17	SeqSero2, SISTR Geno-serotyping	Infantis	Contaminated	Infantis	Infantis	Infantis	Contaminated
19	SSP and Galaxy sistr	Infantis O7:r:1,5	not applicable	Infantis O7:r:1,5	Infantis O7:r:1,5	Infantis O7:r:1,5	not applicable
20	Seqsero2 + SISTR1	Infantis	/	Infantis	Infantis	Infantis	/
21	SeqSero 1.2	Infantis	-	Infantis	Infantis	Infantis	-
23	SeqSero_v2	Infantis	N/A - failed QC	Infantis	Infantis	Infantis	N/A - failed QC
27	MOST, SeqSero2, SISTR	Infantis	Infantis and Lagos	Infantis	Infantis	Infantis	Brazzaville|Edinburg|Atento|Leeuwarden and Paratyphi B var. L(+) tartrate+ and Java
28	SeqSero serotype finder	Infantis	Bradford	Infantis	Infantis	Infantis	4:r:-
29	SeqSero2 and Sistr	Infantis	Infantis	Infantis	Infantis	Infantis	-
31	Galaxy Sciensano	Infantis	Infantis (contaminated)	Infantis	Infantis	Infantis	Paratyphi B III (contaminated)
32	Based on genome, with sistr software	Intis	Intis	Intis	Intis	Intis	Paratyphi B Variant Java
63	SeqSero2	Infantis		Infantis	Infantis	Infantis	
80	SeqSero2 + Sistr	Infantis	QC not passed	Infantis	Infantis	Infantis	QC not passed

Annex 6. Reasons for (not) passing participant's QC and (not) excluding strains 24SCA12 and 24SCA16 from cluster analysis

In blue: Deviation from the expected result.

Lab code	Strain 24SCA12 passed QC	Reason(s) not passing QC	Strain 24SCA12 excluded from cluster analysis[a)]
Expected	**No**	**Wet mixture of 24SCA02 REF *S.* Infantis and 19SCA03 *S.* Typhimurium**	**Yes**
1	No	Too many contigs (3104 contigs), total length of assembly too long (6,7Mbp), N50 too low (3952 bp). Suspect intra species contamination.	Yes
2	No	Total length (> 5326000); NumContamSNVs (x>7);	Yes
3	No	> 300 contigs, bad N50 (6782), 19,48% contamination, multiple serotypes detected	Yes
4	No	total length > 5.3 Mb, Confindr 37 SNP; number of contigs >300	Yes
5	No	cgMLST good targets percentage <90%	Yes
7	No	Low Core %, high genome size, N50 & Contig counts	Yes
11	No	No contigs =2952, size = 6 Mb = mixed culture	Yes
12	No	percentage good targets too low - 96.6; genome size too high - 5.5 MB	Yes
13	No	Contaminated with other Salmonella, based on MLST2.0 and ConFindr	Yes
14	No	Contamination detected by CheckM is above the quality threshold of 2	Yes
16	No	Intraspecies contamination with 363 SNVs and N50 value is less than 150000pb	Yes
17	No	too many contigs, too big assembly length, wrong GC%, too many missassemblies, too small N50, too large L50	Yes
19	No	N. contigs>500; Total length higher than expected	Yes
20	No	Contamination, total length of assembly, total number of contigs, N50, largest contig	Yes
21	No	checkM2 contamination 4,97; total number of contigs 2360; total length of assembled genome 5,8 Mb; N50 9901; Kmer intraspecies contamination indication	Yes
23	No	Poor assembly, incorrect genome size, possibly mix of strains/serovars	Yes
24	Yes		No
27	No	Contamination: multiple serotypes (Infantis, Lagos); large assembly (> 6 Mbp)	Yes
28	No	contamination	Yes
29	No	Did not pass cgMLST loci requirement. Found 94,7%.	Yes
31	No	Contamination	Yes
32	No	too litle core genomes genes/too many contigs	Yes
63	No	Likely a mixture of multiple strains; confindr suggests contamination, assembly too large and too fragmented, N50 too low.	Yes
80	No	Out of the thresholds recommended by EFSA : Too much contigs (4590); length > 5.4Mb; N50=3528 &Contamination detected (Intra specie)	Yes

a) Based on the submitted distance matrix.

Lab code	Strain 24SCA16 passed QC	Reason(s) not passing QC	Strain 24SCA16 excluded from cluster analysis[a)]
Expected	**No**	**Wet mixture of 24SCA03 *S.* Infantis and 2013S16 *S.* Paratyphi B, var. Java**	**Yes**
1	No	Too many contigs (3043 contigs), total length of assembly too long (6,5Mbp), N50 too low (3965 bp). Suspect intra species contamination.	Yes
2	No	Total length (> 5326000); NumContamSNVs (x>7);	Yes
3	No	only 46% cgMLST targets identified, > 300 contigs, bad N50, 50% contamination with (1751), multiple serotypes detected	Yes
4	No	total length <4.3 Mb, number of contigs >300, N50 too small, <90% cgMLST loci called	Yes
5	No	cgMLST good targets percentage <90%	Yes
7	No	low core %, high genome size, N50 & contig count.	Yes
11	No	No of contigs = 8181, total size = 6.5 Mb = mixed culture	Yes
12	No	percentage good targets too low - 43.0; genome size too high - 6.6 MB	Yes
13	No	% of good cgMLST Targets 88,51 %	Yes
14	No	Contamination detected by CheckM is above the quality threshold of 2	Yes
16	No	Intraspecies contamination with 359 SNVs, N50 value is less than 150000pb and less than 90% bases are covered with a sequencing depth higher than 20X	Yes
17	No	too many contigs, too big assembly length, wrong GC%, too many missassemblies, too small N50, too large L50	Yes
19	No	N. contigs>500; Total length higher than expected; N50<1500	Yes
20	No	Contamination, total length of assembly, total number of contigs, N50, largest contig	Yes
21	No	checkM2 contamination 5,67; total number of contigs 3361; total length of assembled genome 5,9 Mb; N50 6166; Kmer intraspecies contamination indication	Yes
23	No	Poor assembly, incorrect genome size, possibly mix of strains/serovars	Yes
24	No	N50 too small and too many contigs	Yes
27	No	Contamination: multiple serotypes (Java, Paratyphi B var. L(+) tartrate and Brazzaville|Edinburg|Atento|Leeuwarden); large assembly (> 6 Mbp)	Yes
28	No	contamination; N50 to low; nr of contigs to high	Yes
29	No	Did not pass cgMLST loci requirement. Found 37,1%.	Yes
31	No	Contamination	Yes
32	No	too litle core genomes genes/too many contigs	Yes
63	No	Likely a mixture of multiple strains; confindr suggests contamination, assembly too large and too fragmented, N50 too low.	Yes
80	No	Out of the thresholds recommended by EFSA: Too much contigs (8840); length > 5.3Mb; N50=1692 & Contamination detected (Intra specie)	Yes

a) Based on the submitted distance matrix.

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

In blue: Deviation from the expected result.

	Strain code												
Labcode-method	**24 SCA-REF**	**24 SCA01**	**24 SCA02**	**24 SCA03**	**24 SCA04**	**24 SCA05**	**24 SCA06**	**24 SCA11**	**24 SCA12**	**24 SCA13**	**24 SCA14**	**24 SCA15**	**24 SCA16**
	Ref Inf	Inf	Inf=REF	Inf	Inf	Inf	Inf	Inf	Inf+STM	Inf	Inf=REF	Inf	Inf+Java
EL-cgMLST	0	130	**0**	**1**	144	**1**	35	84		146	**0**	86	
1-cgMLST	0	131	**0**	**1**	145	**1**	35	87		147	**0**	89	
2-cgMLST	0	132	**0**	**2**	146	**2**	37	88		148	**0**	88	
3-cgMLST	0	35	130	**0**	**1**	144	**1**	84		146	**0**	86	
4-cgMLST	0	138	**0**	**1**	152	**2**	38	94		154	**0**	96	
5-cgMLST	0	129	**0**	**1**	143	**1**	35	83		145	**0**	85	
7-cgMLST	0	130	**0**	**1**	144	**1**	35	84		146	**0**	86	
11-cgMLST	0	130	**0**	**1**	144	**1**	35	84		146	**0**	86	
12-cgMLST[a)]	0	131	**0**	**1**	144	**1**	35	84		146	**0**	86	
13-cgMLST[b)]	0	128	**0**	**1**	142	**1**	35	84		146		86	
14-cgMLST	0	153	**0**	**2**	170	**3**	39	102		170	**0**	100	
16-cgMLST[c)]	0	155	**0**	**3**	170	**3**	41	103		171	**0**		
17-cgMLST	0	132	**0**	**0**	145	**1**	14	88		150	**0**	84	
20-cgMLST	0	130	**0**	**1**	144	**1**	35	84		146	**0**	85	
23-cgMLST	0	124	**1**	**1**	133	**1**	32	86		146	**1**	84	
24-cgMLST	0	130	**0**	**1**	144	**1**	35	84	**1**	146	**0**	86	
28-cgMLST	0	127	**0**	**1**	141	**1**	35	85		143	**0**	86	
29-cgMLST	0	138	**0**	**2**	149	**3**	37	88		152	**2**	93	
32-cgMLST	0	129	**2**	**1**	144	**1**	35	84		146	**0**	85	
80-cgMLST	0	130	**0**	**2**	145	**1**	35	84		146	**0**	85	
17-wgMLST	0	165	**0**	**1**	189	**1**	26	108		197	**0**	103	
19-SNPa[d)]	0	293	10	12	312	**7**	63	177		317	15	195	
21-SNPr	0	275	**0**	**3**	297	**3**	58	171		298	**0**	178	
23-SNPr	0	279	**3**	**5**	296	**5**	60	167		296	**1**	179	
27-SNPr	0	259	**0**	**3**	282	**3**	55	163		282	**0**	171	
31-SNPr	0	300	**0**	**2**	549	**2**	121	210		550	**0**	194	
63-SNPr[e)]	0	259	**0**	**3**	263		56	160		282	**0**	173	
80-SNPa	0	303	**1**	**5**	954	**3**	344	343		954	**0**	320	

The 4 clustering strains indicated in green. The 2 strains expected to be excluded indicated in orange.
Empty cell: strain not included in the lab's distance matrix.
Numbers in **bold** were reported to be clustering with the REF strain.

a) Comment Lab 12: We would have repeated NGS of strain 24SCA12 as it has an AD=3 to the reference strain, but failed QC.
b) Comment Lab 13: 24SCA12 included *S*. Infantis which clustered with the REF strain but was contaminated with other salmonella. 24SCA14 included *S*. Infantis which clustered with REF strain but was contaminated slightly with *E. coli*. 24SCA16 included *S*. Paratyphi B and some other salmonella. All of them were excluded from the final cluster analysing.
c) Comment Lab 16: Strain 24SCA15: Intraspecies contamination with 37 SNVs, excluded from analysis/distance matrix.
d) Comment Lab 19: Strain 24SCA02, Strain 24SCA03 and 24SCA0314 shared 11 SNPs.
e) Comment Lab 63: Data set passed QC criteria for *Salmonella*, but was excluded from reference-based SNP-analysis as it is a more distant sequence type.