# Interim Summary Report

# EURL-*Salmonella* Proficiency Test Cluster Analysis 2022

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

20 July 2023
Z&O letter report 2023-0035

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 2022, concerning the optional part on Cluster Analysis (CA).
A total of 20 NRLs participated in the cluster analysis; all performed WGS analysis and five participants also performed MLVA analysis.
The evaluations of the individual laboratory results were sent to each of the participants separately. The full results will be reported in more detail in the final report on the EURL-*Salmonella* PT Typing 2022.

## 2. *Salmonella* strains for cluster analysis

A total of six *Salmonella* strains (22SCA01 – 22SCA06) in HI agar transport tubes were sent to the participants in the EURL-*Salmonella* PT Typing 2022, part CA. Background information on these "wet" strains is given in Table 1A. In addition, raw sequence data (fastq.gz files, md5 checksums) on another six *Salmonella* strains (22SCA11 – 22SCA16) were made available to the participants via a secure ftp server for "dry" evaluation (WGS only). 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 2022*

| Strain code | Serovar | ST | Origin | MLVA-profile |
|---|---|---|---|---|
| 22SCA01 a) | Enteritidis | 11 | Human | 3-10-5-3-1 |
| 22SCA02 c) | Enteritidis | 11 | Human | 3-10-4-4-1 |
| 22SCA03 | Enteritidis | 11 | Human | 2-9-7-4-2 |
| 22SCA04 b) REF | Enteritidis | 11 | Human | 3-10-6-3-1 |
| 22SCA05 a) | Enteritidis | 11 | Human | 3-10-5-3-1 |
| 22SCA06 b) | Enteritidis | 11 | Human | 3-10-6-3-1 |

a) Technical duplicates
b) Technical duplicates
c) Biological duplicate strain 21SCA08

*Table 1B. Background information on the "dry"* Salmonella *strains used for cluster analysis in 2022 (WGS only)*

| Strain code | Serovar | ST | Origin |
|---|---|---|---|
| 22SCA11 c) | Enteritidis | 11 | Human |
| 22SCA12 c) | Enteritidis | 11 | Human |
| 22SCA13 * | Enteritidis | n.a. | Unknown |
| 22SCA14 | Enteritidis | 11 | Human |
| 22SCA15 d) | Enteritidis | 11 | Human |
| 22SCA16 | Enteritidis | 11 | Human |

c) Strain 21SCA08, raw data PT 2021 from 2 different participants
d) Biological duplicate strain 22SCA01
* *S.* Enteritidis contaminated with *E. coli* reads
n.a. not applicable (QC not passed)

Strains were selected by the EURL-*Salmonella* to be suitable for analysis by using either MLVA or WGS. In preparation of the PT 2021 on cluster analysis, a set of 15 human surveillance *Salmonella* strains were re-cultured from storage (2019) and submitted for MLVA and WGS analysis both directly and after sub-culturing for ten times. Re-cultured strains were stored both at minus 70°C and in HI transport tubes. Strains were re-cultured again on blood-agar plates from both types of storage in the summer of 2022 and submitted for WGS analysis.

Subsequently, six "wet" strains and six "dry" strains were selected for inclusion in the PT 2022 (also see Figure 1). Two sets of wet technical duplicates were included: strain 22SCA01 and strain 22SCA05 shipment tubes were both prepared from the same blood-agar plate containing strain 22SCA01; strain 22SCA04 and strain 22SCA06 shipment tubes were both prepared from the same blood-agar plate containing strain 22SCA04.
Cluster analysis could be performed up to the choice of the participant by MLVA and/or WGS, using their own routine method(s).
Like the year before, the PT Cluster Analysis 2022 was mimicking an outbreak situation, with a *Salmonella* Enteritidis ST11, MLVA type 3-10-6-3-1 as the reference strain (22SCA-REF). Raw WGS data of this strain (22SCA-REF_R1.fq.gz and 22SCA-REF_R2.fq.gz, as well as md5 checksums) were also made available through the secure ftp server.
For this particular PT 2022 situation, the cgMLST-based cluster definition (WGS) was set at maximum six allelic differences from the reference sequence. For MLVA, the cluster definition was set at no loci with a different number of repeats.
Participants were asked to analyse the six "wet" *Salmonella* strains (MLVA/WGS) and the six "dry" *Salmonella* strains (WGS only), and to report per strain whether a clustering match with the reference strain was found or not.
Evaluation (per methodology) 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 2022).

## 3. Evaluation of the cluster analysis results based on MLVA data

Five participants (Laboratory codes 1, 17, 19, 28, 33) submitted cluster analysis results based on MLVA data.
The allelic profiles as submitted by the participants are given in Annex 1. Laboratory 19 did not report the results in the expected format and therefore these results are regarded as deviating.
Participants were asked to report per strain (Table 2) whether or not a clustering match was found with the reference outbreak strain (22SCA-REF) in the EURL-*Salmonella* PT 2022: *Salmonella* Enteritidis ST11, MLVA type 3-10-6-3-1.
The MLVA cluster definition for the PT 2022 was set at no loci with a different number of repeats. Based on this cluster definition, MLVA-based results were expected to indicate strains 22SCA04 (reference strain) and 22SCA06 (technical duplicate of the reference strain) to be a clustering match with the reference outbreak strain.
Four participants reported the MLVA-based cluster analysis results completely as expected (Table 2).
The fifth participant reported the allelic profiles in a deviating format, therefore the evaluation of their results cannot be done according to the PT Typing 2022 Protocol.

*Table 2. Expected cluster analysis results and the cluster analysis results as reported by the 5 MLVA participants*

| | **Strain code** | | | | | |
|---|---|---|---|---|---|---|
| **Labcode** | **22SCA01** | **22SCA02** | **22SCA03** | **22SCA04** | **22SCA05** | **22SCA06** |
| **Expected** | **No** | **No** | **No** | **Yes** | **No** | **Yes** |
| 1 | No | No | No | Yes | No | Yes |
| 17 | No | No | No | Yes | No | Yes |
| 19* | Yes | Yes | Yes | Yes* | Yes | Yes* |
| 28 | No | No | No | Yes | No | Yes |
| 33 | No | No | No | Yes | No | Yes |

In blue: Deviation from the expected result.
*The allelic profiles were not reported in the expected format, therefore the evaluation of the cluster analysis results cannot be done according to the PT Typing 2022 Protocol.

**4. Evaluation of the cluster analysis results based on WGS data**

Twenty participants (Table 3) submitted a total of 26 cluster analysis results based on WGS data; four participants submitted both cgMLST-based and SNP-based data results, and one participant submitted two cgMLST-based and one SNP-based data analyses. Some details on the sequencing procedures as performed by the participants are given in Annex 2.
WGS-based pre-test results as well as the PT 2022 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+, using the cgMLST Enterobase v2.0 scheme and visualised in a minimum spanning tree (MST, Figure 1).
Stable and consistent cgMLST analysis result were obtained for both the minus 70°C-stored and the HI-tubes-stored strains (Figure 1, ELt5a and ELt5b). Subsequently, the "wet" strains selected to be included for the PT 2022 (Figure 1, ELt6) were freshly prepared from the minus 70°C stocks (2021).

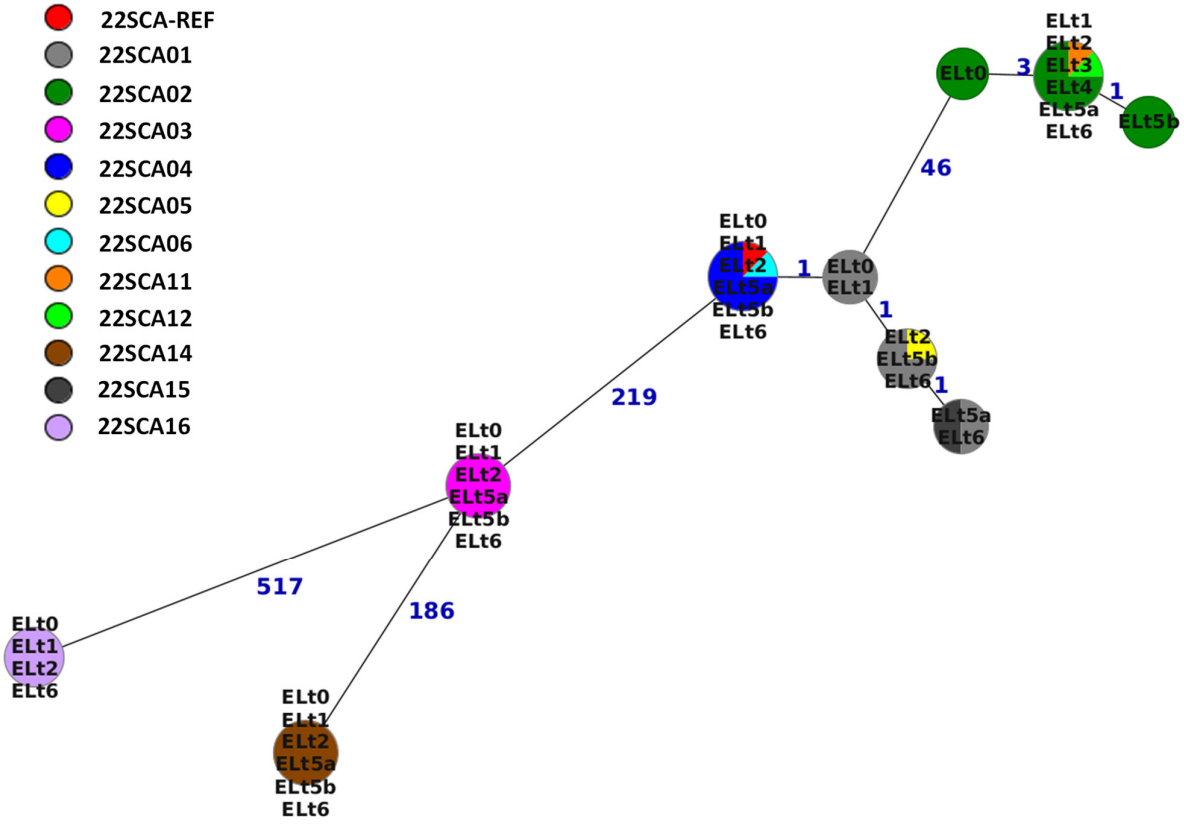



*Figure 1. MST of the EURL-*Salmonella *(EL) pre-tests and PT 2022 results, (RidomSeqSphere+, cgMLST (3002), pairwise ignoring missing values).*

*ELt0: Original WGS data from the stored human surveillance* Salmonella *strains (2019);*
*ELt1: WGS data from initial pre-testing for PT 2021 (8 July 2021);*
*ELt2: WGS data after ten times sub-culturing (blood-agar/BPW) for PT 2021 (17 August 2021);*
*ELt3: PT 2021 data at the start of the PT (November 2021);*
*ELt4: PT 2021 data at the end of the PT (February 2022);*
*ELt5a: WGS data after one-year storage at minus 70°C (September 2022);*
*ELt5b: WGS data after one-year storage in HI agar transport tubes (September 2022);*
*ELt6: PT 2022 data at the start of the PT (November 2022).*

Fourteen compressed paired-end fastq files (strains 22SCA11 - 22SCA16 plus 22SCA-REF) had to be downloaded for analysis from the secure ftp server. The md5 checksums for these files were available on the server as well (Annex 3). Participants were asked whether they checked the md5sum values after downloading, and 16 participants indicated that they did this. Participants were also asked to copy/paste in the result form "your md5 output for all your strains". Regrettably, this was not a very clear question, but 8 of 16 participants entered their md5 checksums for the files that they had to upload to the secure ftp server (for strains 22SCA01 - 22SCA06), which was inclined with this question. After downloading the raw data files from the participants at the EURL-*Salmonella*, this was checked to be correct, indicating that also this transfer of data via the secure ftp site went alright.

All participants' raw data (compressed fastq files) for the six "wet" strains (22SCA01 – 22SCA06) were successfully processed through the Juno-assembly pipeline as mentioned. The *de novo* assembled genomes (fasta files) were analysed in Ridom SeqSphere+, using the cgMLST Enterobase v2.0 and visualised in a MST, which also includes the "dry" strain data (22SCA-REF, 22SCA11 – 22SCA16) (Figure 2). Data per "wet" strain are given in Annex 4. Results for Laboratory 26 indicate a swap between their results for strains 22SCA02, 22SCA03, and 22SCA04.

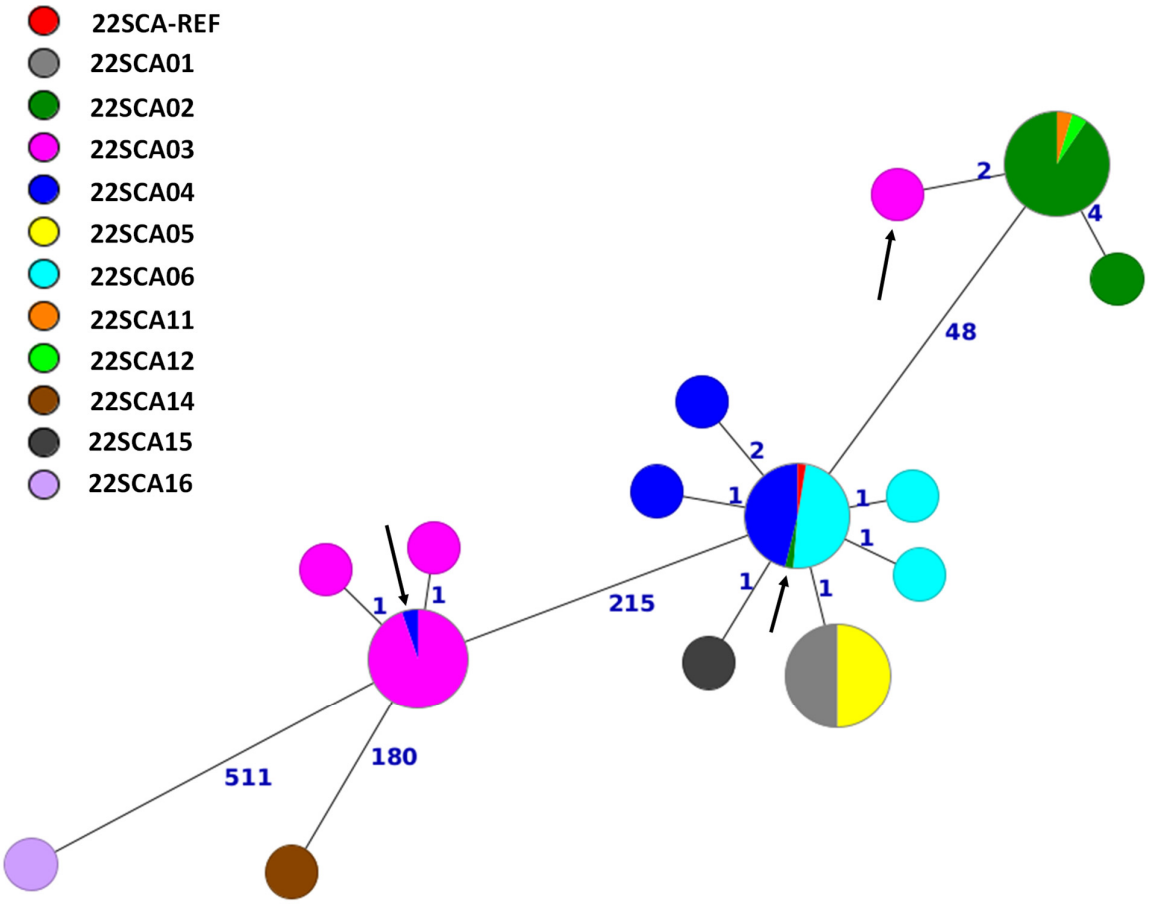



*Figure 2. MST of the strains from the participants' processed raw data plus the "dry" strain data (22SCA-REF, 22SCA11 – 22SCA16) (Ridom SeqSphere+, cgMLST (3002), pairwise ignoring missing values)*

*Three arrows are indicating the swap between results for strains 22SCA02, 22SCA03, and 22SCA04.

Participants were asked to report per strain:
-whether the data passed their Quality Control (QC) criteria or not,
-whether a clustering match with the reference strain in the EURL-*Salmonella* PT Typing 2022 (22SCA-REF) was found or not.
Apart from the reference cluster, any further clusters could be reported optionally.

Strain 22SCA13 was expected not to pass the QC of the participants, because the data files of this strain also contained numerous *E. coli* reads. The PT Typing 2022 Protocol indicated to exclude strains from the cluster analysis if the data did not pass the QC.
Strain 22SCA13 was reported to be excluded from cluster analysis by 17 of the 20 participants, although it was still included in the distance matrix of one of their submissions.
Reasons for (not) excluding strain 22SCA13 are given in Annex 5.
Annex 6 shows per submission the participants' distance matrix data for their comparison of the 22SCA-REF with the final 11 strains (strain 22SCA13 expected to be excluded from the cluster analysis).
The cluster definition for this particular PT Typing 2022 situation was set at maximum allelic differences from the reference sequence. Based on this (cgMLST-based) cluster definition, WGS-based results were expected to indicate the "wet" strains 22SCA04 (reference strain), 22SCA06 (technical duplicate of the reference strain), 22SCA01 (clustering with the reference strain), 22SCA05 (technical duplicate of strain 22SCA01) and the "dry" strain 22SCA15 (ELt5a data of strain 22SCA01) to be a clustering match with the provided reference outbreak strain 22SCA-REF data (also see Figure 1).
Nineteen of the 26 submissions (five participants with multiple submissions) reported the WGS-based cluster analysis results completely as expected (Table 3).

*Table 3. Expected cluster analysis results and the cluster analysis results as reported per data analysis method by the 20 WGS participants*

| | Strain code | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **Labcode-method** | **22 SCA01** | **22 SCA02** | **22 SCA03** | **22 SCA04** | **22 SCA05** | **22 SCA06** | **22 SCA11** | **22 SCA012** | **22 SCA13** | **22 SCA14** | **22 SCA15** | **22 SCA16** |
| **Expected** | **Yes** | **No** | **No** | **Yes** | **Yes** | **Yes** | **No** | **No** | **n.a.** | **No** | **Yes** | **No** |
| 1-cgMLST | Yes | No | No | Yes | Yes | Yes | No | No | n.a. | No | Yes | No |
| 2-cgMLST | Yes | No | No | Yes | Yes | Yes | No | No | n.a. | No | Yes | No |
| 3-cgMLST | Yes | No | No | Yes | Yes | Yes | No | No | n.a. | No | Yes | No |
| 7-cgMLST | Yes | No | No | Yes | Yes | Yes | No | No | n.a. | No | Yes | No |
| 8-SNPr | Yes | No | No | Yes | Yes | Yes | No | No | n.a. | No | Yes | No |
| 8-cgMLST | Yes | No | No | Yes | Yes | Yes | No | No | n.a. | No | Yes | No |
| 9-SNPr | Yes | No | No | Yes | Yes | Yes | No | No | n.a. | No | Yes | No |
| 10-SNPr | Yes | No | No | Yes | Yes | Yes | No | No | n.a. | No | Yes | No |
| 14-cgMLST | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| 14-SNPr | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| 16-SNPr | Yes | No | No | Yes | Yes | Yes | No | No | n.a. | No | Yes | No |
| 17-cgMLST | Yes | No | No | Yes | Yes | Yes | No | No | n.a. | No | Yes | No |
| 19-cgMLST | Yes | No | Yes | Yes | Yes | Yes | No | No | n.a. | No | Yes | No |
| 23-cgMLST | Yes | No | No | Yes | Yes | Yes | No | No | n.a. | No | Yes | No |
| 24-cgMLST | Yes | No | No | Yes | Yes | Yes | No | No | n.a. | No | Yes | No |
| 26-SNPa | Yes | Yes | No | No | Yes | Yes | No | No | n.a. | No | Yes | No |
| 27-cgMLST1 | Yes | No | No | Yes | Yes | Yes | No | No | n.a. | No | Yes | No |
| 27-cgMLST2 | Yes | No | No | Yes | Yes | Yes | No | No | n.a. | No | Yes | No |
| 27-SNPr | Yes | No | No | Yes | Yes | Yes | No | No | n.a. | No | Yes | No |
| 28-cgMLST | Yes | No | No | Yes | Yes | Yes | No | No | No | No | Yes | No |
| 28-SNPr | Yes | No | No | Yes | Yes | Yes | No | No | No | No | Yes | No |
| 29-SNPr | Yes | No | No | Yes | Yes | Yes | No | No | No | No | Yes | No |
| 30-SNPa | Yes | No | No | Yes | Yes | Yes | No | No | n.a. | No | Yes | No |
| 30-cgMLST | Yes | No | No | Yes | Yes | Yes | No | No | n.a. | No | Yes | No |
| 32-cgMLST | Yes | No | No | Yes | Yes | Yes | No | No | n.a. | No | Yes | No |
| 33-cgMLST | Yes | No | No | Yes | Yes | Yes | No | No | n.a. | No | Yes | No |

In blue: Deviation from the expected result.

Technical duplicates 22SCA01 and 22SCA05 were reported within one cluster in all 26 submissions. Technical duplicates 22SCA04 and 22SCA06 were reported within one cluster in all but one of the submissions.

Some observations on the interpretation of Table 3 are given below:
Laboratory 14: there may have been a misunderstanding in the way to report the results of the cluster analysis. Based on the submitted distance matrix, and the analyses shown in Figure 2/Annex 4, data are in line with the expected results except for strain 22SCA13 (Annex 5 and 6).
Laboratory 19: this deviation may have been a mistake in filling the result form, because this answer is not supported by the distance matrix that was submitted, nor by Figure 2/Annex 4.
Laboratory 26: there may have been a swap of strain numberings: Data of strain 22SCA02 reported as strain 22SCA04; Data of strain 22SCA03 reported as strain 22SCA04; Data of strain 22SCA04 reported as strain 22SCA02. Based on the submitted distance matrix (with the wrong numbering), and the analysis shown in Figure 2/Annex4, data seem to be in line with the expected results.
Laboratory 28 and laboratory 29: strain 22SCA13 was expected to be excluded from the cluster analysis (also see Annex 5 and 6).

Apart from the cluster with the reference strain, a second cluster was optionally to be identified: 22SCA02, 22SCA11, and 22SCA12. This second cluster was reported in 21 of the 26 submissions, three of these were deviating from the expected results. Laboratory 26 reported 22SCA11 and 22SCA12 to be a second cluster, without 22SCA02. Laboratory 29 reported the second cluster correctly, but also considered the four remaining strains as a third cluster, "although quite divergent". Laboratory 33 reported the second cluster correctly, but reported strains 22SCA04 and 22SCA14 as a third cluster.

## List of abbreviations

| | |
|---|---|
| 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* |
| 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-*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 |
| WGS | Whole Genome Sequencing |

## References

European Centre for Disease Prevention and Control. Laboratory standard operating procedure for multiple-locus variable-number tandem repeat analysis of *Salmonella enterica* serotype Enteritidis. Stockholm:ECDC; 2016. doi 10.2900/973540. Available online: https://www.ecdc.europa.eu/sites/default/files/media/en/publications/Publications/Salmonella-Enteritidis-Laboratory-standard-operating-procedure.pdf

## 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. Expected and reported MLVA results for all five participants

Loci were asked to be reported in the order: SENTR7-SENTR5-SENTR6-SENTR4-SE-3.

| | Strain code | | | | | |
|---|---|---|---|---|---|---|
| **Lab code** | **22SCA01** | **22SCA02** | **22SCA03** | **22SCA04** | **22SCA05** | **22SCA06** |
| **Expected** | **3-10-5-3-1** | **3-10-4-4-1** | **2-9-7-4-2** | **3-10-6-3-1** | **3-10-5-3-1** | **3-10-6-3-1** |
| 1 | 3-10-5-3-1 | 3-10-4-4-1 | 2-9-7-4-2 | 3-10-6-3-1 | 3-10-5-3-1 | 3-10-6-3-1 |
| 17 | 3-10-5-3-1 | 3-10-4-4-1 | 2-9-7-4-2 | 3-10-6-3-1 | 3-10-5-3-1 | 3-10-6-3-1 |
| 19 | 5-2-3-7-6-6-11 | 5-2-3-7-6-6-11 | 5-2-3-7-6-6-11 | 5-2-3-7-6-6-11 | 5-2-3-7-6-6-11 | 5-2-3-7-6-6-11 |
| 28 | 3-10-5-3-1 | 3-10-4-4-1 | 2-9-7-4-2 | 3-10-6-3-1 | 3-10-5-3-1 | 3-10-6-3-1 |
| 33 | 3-10-5-3-1 | 3-10-4-4-1 | 2-9-7-4-2 | 3-10-6-3-1 | 3-10-5-3-1 | 3-10-6-3-1 |

In blue: Deviation from the expected result.

## Annex 2. Sequencing details as reported by the 20 WGS participants

| Labcode | Wet lab[a)] | WGS platform used | Data analysis | Tool for analysis | Method for cluster analysis |
|---|---|---|---|---|---|
| 1 | In-In-In | Illumina MiSeq | cgMLST-based | Ridom SeqSphere | Minimum Spanning Tree (MST) |
| 2 | In-In-In | Illumina NextSeq | cgMLST-based | Ridom SeqSphere | Minimum Spanning Tree (MST) |
| 3 | In-In-In | Illumina MiSeq | cgMLST-based | BioNumerics | Minimum Spanning Tree (MST) |
| 7 | In-In-In | Illumina MiSeq | cgMLST-based | linux command line | Neighbor joining (NJ) |
| 8-cgMLST | In-Out-Out | Illumina NovaSeq | cgMLST-based | Ridom SeqSphere | Minimum Spanning Tree (MST) |
| 8-SNPr | In-Out-Out | Illumina NovaSeq | SNP-based - reference-based | SNIPPY | Maximum likelihood (ML) |
| 9 | In-In-In | Illumina MiSeq | SNP-based - reference-based | Python script | Minimum Spanning Tree (MST) |
| 10 | In-In-Out | Illumina NextSeq | SNP-based - reference-based | https://cge.food.dtu.dk/services/CSIPhylogeny/ | Maximum likelihood (ML) |
| 14-cgMLST | In-In-In | Illumina MiSeq | cgMLST-based | galaxy.sciensano | Maximum likelihood (ML) |
| 14-SNPr | In-In-In | Illumina MiSeq | SNP-based - reference-based | Galaxy Sciensano | Maximum likelihood (ML) |
| 16 | In-In-In | Illumina MiSeq | SNP-based - reference-based | CSIPhylogeny (https://cge.food.dtu.dk/services/CSIPhylogeny/) | Maximum likelihood (ML) |
| 17 | In-In-In | Illumina MiSeq | cgMLST-based | in-house galaxy | MSTreeV2 |
| 19 | In-In-In | Illumina Miniseq | cgMLST-based | Ridom SeqSphere | Minimum Spanning Tree (MST) |
| 23 | In-In-In | Illumina NextSeq 2000 | cgMLST-based | Ridom SeqSphere | Minimum Spanning Tree (MST) |
| 24 | In-In-In | Illumina MiSeq | cgMLST-based | Ridom SeqSphere | Minimum Spanning Tree (MST) |
| 26 | In-In-In | Illumina MiSeq | SNP-based - assembly-based | CSIPhylogeny 1.4 | Maximum likelihood (ML) |
| 27-cgMLST1 | In-In-In | Illumina NextSeq | cgMLST-based | inhouse chewieSnake pipeline (Enterobase scheme) | single linkage hierarchical clustering |
| 27-cgMLST2 | In-In-In | Illumina NextSeq | cgMLST-based | Ridom SeqSphere+Enterobase scheme | single linkage hierarchical clustering |
| 27-SNPr | In-In-In | Illumina NextSeq | SNP-based - reference-based | SNP-analysis using SnippySnake pipeline | single linkage hierarchical clustering |
| 28-cgMLST | In-In-In | Illumina MiSeq | cgMLST-based | PyMLST v1 | Minimum Spanning Tree (MST) |
| 28-SNPr | In-In-In | Illumina MiSeq | SNP-based - reference-based | BWA, bcftools, RAxML | Maximum likelihood (ML) |
| 29 | In-In-In | Illumina MiSeq | SNP-based - reference-based | Snippy, Snapper DB, Gubbins, RAxML, iToL | ML and SNP address analysis |
| 30-cgMLST | In-In-In | Illumina MiSeq | cgMLST-based | chewBBACA using the scheme from Enterobase | Calculated AD based on output chewBBACA |
| 30-SNPa | In-In-In | Illumina MiSeq | SNP-based - assembly-based | In house pipeline[b)] | Maximum likelihood (ML) |
| 32 | In-In-In | MiniSeq | cgMLST-based | Ridom SeqSphere | Distance matrrix only |
| 33 | In-In-In | Illumina MiSeq | cgMLST-based | ChewBBaCa | Minimum Spanning Tree (MST) |
| EURL-Salm | In-In-In | Illumina NextSeq | cgMLST-based | Ridom SeqSphere | Minimum Spanning Tree (MST) |

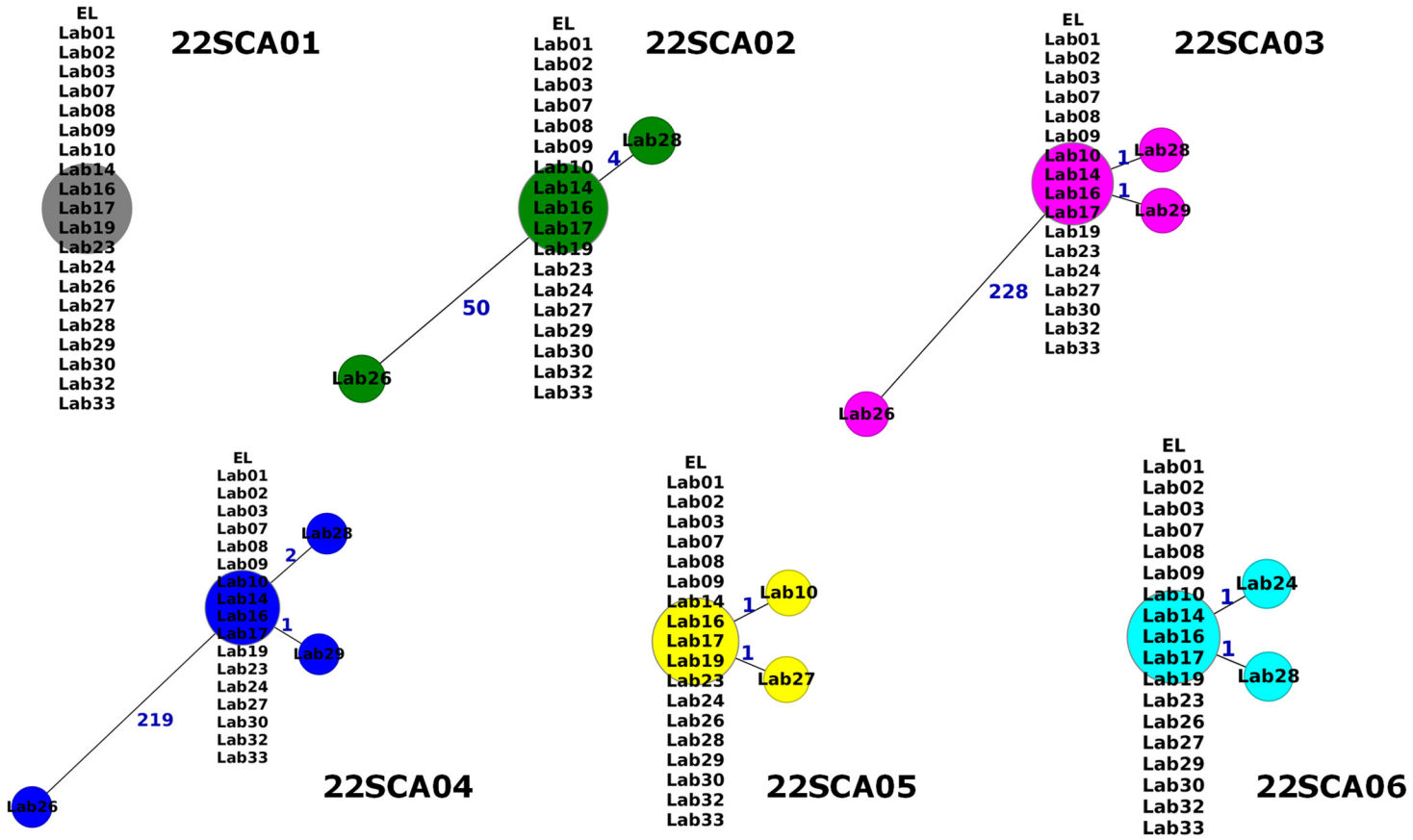
a) Wet lab preparations: DNA extraction, Library preparation, sequencing. In: In-house, Out: Outsourced.
b) Based on parSNP, Gubbins, creating a ML tree in IQ Tree, creating a SNP distance matrix with snp-dists.

## Annex 3. Md5 checksums of the 14 files that had to be downloaded from the secure ftp server for further analysis

| | |
|---|---|
| ef67c2f8a661568dd6ea6b416b31c935 | 22SCA11_R1.fastq.gz |
| a1faf3e3d910d3ffa7626e7f2133d657 | 22SCA11_R2.fastq.gz |
| 7df616cd89c6ec530eee347c812950cf | 22SCA12_R1.fastq.gz |
| 77dc74c4e5b45f55f58d7032b24bc8b0 | 22SCA12_R2.fastq.gz |
| 0ea86d67a119bf13acbf67d75d46ebd8 | 22SCA13_R1.fastq.gz |
| f8fdd64d6d8bdfef56545007f03105f6 | 22SCA13_R2.fastq.gz |
| e61775a2192d0fcf5e51256c56a6ac90 | 22SCA14_R1.fastq.gz |
| 9b6758df047758026294e018e0a6c139 | 22SCA14_R2.fastq.gz |
| e93b28a1abe7099909851421b657e0d7 | 22SCA15_R1.fastq.gz |
| e798974e3fd4ab4b63cc98b5562a3a71 | 22SCA15_R2.fastq.gz |
| 04d9e7263606fe78ec625cce8c2c536f | 22SCA16_R1.fastq.gz |
| 4f6339282180b90db737e551a60dde49 | 22SCA16_R2.fastq.gz |
| 65f572c91b90478c144d599e3035e432 | 22SCA-REF_R1.fastq.gz |
| 3004180d62c0bf76115a61129e858119 | 22SCA-REF_R2.fastq.gz |

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

Results for Laboratory 26 indicate a swap between their results for strains 22SCA02, 22SCA03, and 22SCA04.



## Annex 5. Reasons for (not) excluding strain 22SCA13 from cluster analysis

| Labcode | Strain 22SCA13 excluded from cluster analysis | Reason(s) not passing QC |
|---|---|---|
| 1 | Yes | Contamination Check Result: Potential contamination by second species above 10% detected: Escherichia coli; genome size too big (12.7 MB); strain excluded for MST and matrix |
| 2 | Yes | total length too high, # contigs too high, GC% too low, contamination too high: mostly E. coli reads |
| 3 | Yes | Failed on assembly size, contig number and N50 values. Contamination confirmed using Kmer Finder. |
| 7 | Yes | Contamination with E.coli, number of contigs over 3400, total lenght is oversize |
| 8 | Yes | Contamination by Escherichia 42.26%: coli(39.96%) - Salmonella 37.83%: enterica(37.57%) |
| 9 | Yes | Contamination and failed assembly |
| 10 | Yes | purity of culture; CG%; No. contigs; genome size |
| 16 | Yes | Total length 10,7Mbp, GC% 50,9, 3490 contigs, contamination confirmed by KmerFinder |
| 17 | Yes | contaminated with E. coli, low N50, total length 2x, number of contigs too high, <90% MLST loci detected |
| 19 | Yes | final assembly lenght too large |
| 23 | Yes | Contaminated with E. coli, assembly size too big for Salmonella |
| 24 | Yes | Contamination with E. coli |
| 26 | Yes | N. contigs >500; Total length higher than expected; N50<15000 |
| 27 | Yes | Fail: Total length 12,718,480 bp; Read Fraction Majority Genus 0.488; Contam SNVs 2508 (inter and intra contamination); Warning: # Contigs 5,602; N50 11,563; Single copy orthologs 0.500; Duplication Rate 1.415; GC 51.07 |
| 30 | Yes | Contamination with other species (E. coli) |
| 32 | Yes | Potential contamination by second species above 10% detected: Escherichia coli |
| 33 | Yes | contamination status = True |
| 14 | No | |
| 28 | No* | Sample 22SCA13 was contaminated (only about 35% of the reads were classified as belonging to the Salmonella taxon). We select those reads removing that way the contamination. Thus, we continue the analysis just with the reads classified as Salmonella. |
| 29 | No | |

In blue: Deviation from the expected result.

*The PT Typing 2022 Protocol indicated to exclude strains from the cluster analysis if the data did not pass the QC, therefore the approach by Laboratory 28 was considered as deviating.

## Annex 6. Per submission, the participants' distance matrix data for their comparison to the 22SCA-REF with the 11 (or 12) strains.

| | Strain code | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Labcode-method | 22 SCA-REF | 22 SCA01 | 22 SCA02 | 22 SCA03 | 22 SCA04 | 22 SCA05 | 22 SCA06 | 22 SCA11 | 22 SCA12 | 22 SCA13* | 22 SCA14 | 22 SCA15 | 22 SCA16 |
| 1-cgMLST | 0 | 2 | 50 | 218 | 0 | 2 | 0 | 50 | 50 | | 249 | 2 | 544 |
| 2-cgMLST | 0 | 2 | 50 | 220 | 0 | 2 | 0 | 50 | 50 | | 251 | 3 | 546 |
| 3-cgMLST | 0 | 1 | 51 | 225 | 0 | 1 | 0 | 51 | 51 | | 259 | 1 | 561 |
| 7-cgMLST | 0 | 4 | 54 | 223 | 1 | 4 | 1 | 54 | 54 | | 251 | 4 | 552 |
| 8-cgMLST | 0 | 2 | 50 | 218 | 0 | 2 | 0 | 50 | 50 | | 250 | 2 | 544 |
| 14-cgMLST | 0 | 6 | 103 | 411 | 0 | 6 | 0 | 103 | 103 | 414 | 472 | 6 | 1220 |
| 17-cgMLST | 0 | 6 | 58 | 229 | 1 | 6 | 1 | 58 | 58 | | 262 | 5 | 567 |
| 19-cgMLST | 0 | 3 | 51 | 219 | 1 | 3 | 1 | 51 | 52 | | 250 | 3 | 545 |
| 23-cgMLST | 0 | 2 | 50 | 220 | 0 | 2 | 0 | 50 | 51 | | 251 | 2 | 546 |
| 24-cgMLST | 0 | 2 | 49 | 218 | 0 | 2 | 0 | 50 | 50 | | 249 | 2 | 543 |
| 27-cgMLST1 | 0 | 4 | 53 | 220 | 1 | 4 | 1 | 53 | 54 | 213[a)] | 250 | 4 | 544 |
| 27-cgMLST2 | 0 | 2 | 51 | 220 | 0 | 3 | 0 | 51 | 51 | | 251 | 2 | 546 |
| 28-cgMLST | 0 | 3 | 53 | 212 | 2 | 5 | 1 | 48 | 49 | 232[b)] | 246 | 3 | 537 |
| 30-cgMLST | | 4 | | | 1 | 4 | 1 | | 54 | | | 5 | |
| 32-cgMLST | 0 | 2 | 50 | 219 | 0 | 2 | 0 | 50 | 50 | | 248 | 2 | 545 |
| 33-cgMLST | 0 | 4 | 64 | 246 | 0 | 4 | 0 | 64 | 65 | | 270 | 4 | 606 |
| EURL-Salm-cgMLST | 0 | 2 | 50 | 220 | 0 | 2 | 0 | 50 | 50 | | 251 | 3 | 546 |
| 26-SNPa | 0 | 7 | 1 | 126 | 455 | 9 | 3 | 108 | 110 | | 521 | 9 | 1413 |
| 30-SNPa | 0 | 10 | 111 | 453 | 4 | 10 | 4 | 111 | 109 | | 516 | 11 | 1220 |
| 8-SNPr | 0 | 6 | 123 | 501 | 0 | 6 | 0 | 117 | 118 | | 596 | 6 | 1321 |
| 9-SNPr | 0 | 6 | 98 | 447 | 0 | 6 | 0 | 98 | 98 | | 500 | 6 | 1223 |
| 10-SNPr | 0 | 6 | 101 | 443 | 0 | 6 | 0 | 101 | 101 | | 506 | 6 | 1447 |
| 14-SNPr | 0 | 6 | 109 | 449 | 0 | 6 | 0 | 109 | 109 | 468 | 504 | 6 | 1327 |
| 16-SNPr | 0 | 9 | 111 | 451 | 3 | 9 | 2 | 109 | 112 | 403[c)] | 520 | 8 | 1420 |
| 27-SNPr | 0 | 6 | 108 | 479 | 0 | 6 | 0 | 108 | 108 | | 531 | 6 | 1683 |
| 28-SNPr | 0 | 7 | 112 | 495 | 2 | 7 | 1 | 114 | 112 | 516[b)] | 636 | 6 | 1765 |
| 29-SNPr | 0 | 6 | 102 | 418 | 1 | 6 | 1 | 102 | 102 | 418 | 461 | 6 | 1127 |

In blue: Deviation from the expected result.

* Strain 22SCA13 was expected not to be included in the cluster analysis (due to not passing QC).
a) 22SCA13 QC failed and will not be included for reporting. However we checked the allelic differences for own interests.
b) See Annex 5.
c) Reported to be excluded from the cluster analysis.