National Institute for Public Health
and the Environment
Ministry of Health, Welfare and Sport

EURL-*Salmonella* Proficiency Test Typing 2019

RIVM report 2020-0084
W.F. Jacobs-Reitsma et al.

National Institute for Public Health
and the Environment
Ministry of Health, Welfare and Sport

EURL-*Salmonella* Proficiency Test Typing 2019

RIVM report 2020-0084

Colophon

RIVM attaches a great deal of importance to the accessibility of its products. However, it is at present not yet possible to provide this document in a completely accessible form. If a part is not accessible, it is mentioned as such. Also see www.rivm.nl/accessibility.

DOI 10.21945/RIVM-2020-0084

W.F. Jacobs-Reitsma (author), RIVM
A. Verbruggen (author), RIVM
R.E. Diddens (author), RIVM
A.H.A.M. van Hoek (author), RIVM
K.A. Mooijman (author), RIVM

Contact:
W.F. Jacobs-Reitsma
cZ&O Centre for Zoonoses and Environmental Microbiology
wilma.jacobs@rivm.nl

This investigation was performed by order and for the account of the European Commission, Directorate General for Health and Food Safety (DG-SANTE), within the framework of RIVM project number E/114506/19 European Reference Laboratory for *Salmonella.*

Published by:
National Institute for Public Health and the Environment, RIVM
P.O. Box1 | 3720 BA Bilthoven
The Netherlands
www.rivm.nl/en

Synopsis

EURL-*Salmonella* Proficiency Test Typing 2019

From 1992, National Reference Laboratories (NRLs) of European Union (EU) Member States have been obliged to participate in annual quality control 'Proficiency' Tests (PTs). NRLs from countries outside the EU occasionally participate in these tests on a voluntary basis. One of the PTs is on typing of *Salmonella* bacteria. The NRLs of all 28 EU Member States performed well in the 2019 quality control test on *Salmonella* typing. Overall, the participating laboratories were able to assign the correct name to 97% of the strains tested.

Laboratories are obliged to type *Salmonella* with the reference method (serotyping). In 2019, they could also perform additional typing at DNA level, for example by using Whole Genome Sequencing (WGS). These more detailed typing methods are sometimes needed to trace the source of a contamination. Valuable information and typing data were obtained which can be used to improve the organisation of future quality control tests on typing.

Each Member State designates a specific laboratory within their national boundaries to be responsible for the detection and identification of *Salmonella* strains in animals and/or food products. These laboratories are referred to as the National Reference Laboratories (NRLs). The performance of these NRLs in *Salmonella* typing is assessed annually by testing their ability to identify 20 *Salmonella* strains.

The EU candidate countries Albania, Republic of North Macedonia, and Serbia, as well as the European Free Trade Association (EFTA) countries Iceland, Norway and Switzerland took part in the 2019 assessment.

The annual Proficiency Test on *Salmonella* typing is organised by the European Union Reference Laboratory for *Salmonella* (EURL-*Salmonella*). The EURL-*Salmonella* is located at the National Institute for Public Health and the Environment (RIVM) in the Netherlands.

Keywords: EURL-*Salmonella*, *Salmonella*, serotyping, molecular typing, PFGE, MLVA, WGS, cluster analysis, Proficiency Test

Publiekssamenvatting

EURL-*Salmonella* ringonderzoek typering 2019

Sinds 1992 zijn de Nationale Referentie Laboratoria (NRL's) van de Europese lidstaten verplicht om elk jaar hun kwaliteit te laten toetsen met zogeheten ringonderzoeken. Daarnaast doen soms NRL's van landen buiten de Europese Unie (EU) vrijwillig mee. Een van de ringonderzoeken is de typering van *Salmonella*-bacteriën. In 2019 scoorden alle NRL's van de 28 EU lidstaten goed bij de kwaliteitscontrole op typering van *Salmonella*. Als groep konden de deelnemende laboratoria aan 97 procent van de geteste stammen de juiste naam geven.

De laboratoria zijn verplicht om *Salmonella* met een standaardmethode te typeren (serotypering). Daarnaast mochten zij in 2019 zelf aangeven of ze extra typeringen op DNA-niveau wilden doen, bijvoorbeeld met Whole Genome Sequencing (WGS). Deze preciezere typering kan soms nodig zijn om de bron van een besmetting op te sporen. Dit leverde veel informatie op om ook de kwaliteit van de typeringen op DNA-niveau te kunnen toetsen, en waar nodig te verbeteren.

Voor de kwaliteitstoetsen wijst elke lidstaat een laboratorium aan, het Nationale Referentie Laboratorium (NRL). Dit NRL is namens dat land verantwoordelijk om *Salmonella* in monsters van levensmiddelen of dieren aan te tonen en te typeren. Om te controleren of de laboratoria hun werk goed doen, moeten zij onder andere twintig *Salmonella*-stammen de juiste naam kunnen geven.

In 2019 deden zes landen buiten de Europese Unie mee: de EU kandidaat lidstaten Albanië, Republiek Noord-Macedonië en Servië, en de European Free Trade Association (EFTA) landen IJsland, Noorwegen en Zwitserland.

Het Europese Unie Referentie Laboratorium voor *Salmonella* (EURL-*Salmonella*) organiseert het jaarlijkse ringonderzoek Salmonella-typering. Dit laboratorium is gevestigd bij het RIVM in Nederland.

Kernwoorden: EURL-*Salmonella*, *Salmonella*, serotypering, moleculaire typering, PFGE, MLVA, WGS, cluster analyse, ringonderzoek

Contents

Summary — 9

1 Introduction — 11

2 Participants — 13

3 Materials and methods — 15
3.1 Design of the Proficiency Test (PT) — 15
3.1.1 Laboratory codes — 15
3.1.2 Protocol and test report — 15
3.1.3 Transport — 15
3.2 Serotyping part of the PT — 15
3.2.1 Salmonella strains for serotyping — 15
3.2.2 Evaluation of the serotyping results — 16
3.3 Cluster analysis part of the PT — 17
3.3.1 Salmonella strains for cluster analysis — 17
3.3.2 Evaluation of the cluster analysis results in general — 17
3.3.3 Evaluation of the cluster analysis results based on PFGE data — 18
3.3.4 Evaluation of the cluster analysis results based on MLVA data — 19
3.3.5 Evaluation of the cluster analysis results based on WGS data — 20

4 Results and Discussion — 23
4.1 Technical data — 23
4.1.1 General — 23
4.1.2 Accreditation — 24
4.1.3 Transport of samples — 24
4.2 Serotyping results — 24
4.2.1 General — 24
4.2.2 Biochemical testing — 24
4.2.3 Use of PCR for confirmation — 25
4.2.4 General comments on the PT 2019 serotyping evaluation — 25
4.2.5 Serotyping results per laboratory — 25
4.2.6 Performance of the participants — 27
4.2.7 Serotyping results per strain — 28
4.2.8 Trend analysis of the serotyping results of the EU NRLs — 28
4.3 Cluster analysis results — 30
4.3.1 General — 30
4.3.2 Results cluster analysis based on PFGE data — 30
4.3.3 Results cluster analysis based on MLVA data — 31
4.3.4 Results cluster analysis based on WGS data — 33
4.3.5 Investigations on the variability in the results of strains used in the cluster analysis — 39

5 Conclusions — 45
5.1 Serotyping — 45
5.2 Cluster analysis — 45

List of abbreviations — 47

References — 49

Annex 1 Example of an individual laboratory evaluation report on serotyping results — 51

Annex 2 Serotyping results per strain and per laboratory — 53

Annex 3 Details of strains that caused problems in serotyping — 57

Annex 4 Details of serotyping results for strain S21 — 58

Annex 5 Example of an individual laboratory evaluation report on cluster analysis results — 59

Annex 6 Serotyping results cluster analysis part — 63

Annex 7 PFGE results cluster analysis part — 64

Annex 8 MLVA results cluster analysis part — 69

Annex 9 WGS results cluster analysis part, methods used by the participants — 70

Annex 10 WGS results cluster analysis part, QC criteria as indicated by the participants — 71

Annex 11 WGS results cluster analysis part, Minimum Spanning Tree per strain — 75

Annex 12 Results QC parameters on the *de novo* assembled genomes, per participant — 77

Summary

In November 2019, the annual *Salmonella* typing Proficiency Test (PT) was organised by the European Union Reference Laboratory for *Salmonella* (EURL-*Salmonella*, Bilthoven, the Netherlands). The study's main objective was to evaluate whether the typing of *Salmonella* strains by the National Reference Laboratories for *Salmonella* (NRLs-*Salmonella*) in the European Union was carried out uniformly, and whether comparable results were obtained.

A total of 29 NRLs-*Salmonella* of the 28 Member States of the European Union participated, supplemented by the NRLs of the EU candidate countries Albania, Republic of North Macedonia, and Serbia, and the EFTA countries Iceland, Norway and Switzerland.

All 35 laboratories performed serotyping. A total of twenty obligatory *Salmonella* strains plus one optional *Salmonella* strain were selected by the EURL-*Salmonella* for serotyping. The strains had to be typed according to the method routinely used in each laboratory, following the White-Kauffmann-Le Minor scheme (Grimont and Weill, 2007). The laboratories were allowed to send strains for serotyping to another specialised laboratory in their country if this was part of their usual procedure.

Overall, 99% of the strains were typed correctly for the O-antigens, 97% of the strains were typed correctly for the H-antigens, and 97% of the strains were correctly named by the participants.
In 2007, criteria for 'good performance' with regard to serotyping were defined (Mooijman, 2007). Using these criteria, 33 participants achieved good performance. Two participants did not meet the level of good performance at the first stage of the study. The online follow-up on this performed by the first participant was satisfactory, but the proposed training session for the second participant had to be put on hold due to the ongoing COVID-19 pandemic.

Eighteen laboratories also performed additional typing at DNA level by investigating an additional set of ten *Salmonella* strains for the presence of *"cluster(s) of closely related isolates".* This cluster analysis could be optionally performed by PFGE and/or MLVA and/or WGS, using the participants' own routines.

No performance criteria were set for this pilot study. As a minimum, it was expected that the participants would report the technical duplicate strains SCA03 and SCA06 to be (part of) one cluster. This was found in 5/6 (PGFE), 7/7 (MLVA), and 12/14 (WGS) of the cases. Cluster definition, however, was interpreted in a variety of ways, which made an appropriate evaluation difficult.

Valuable information and typing data were obtained from this first pilot study on cluster analysis; this will be used in organising similar future studies.

1 Introduction

This report describes the 2019 Proficiency Test (PT) on typing of *Salmonella* organised by the European Union Reference Laboratory for *Salmonella* (EURL-*Salmonella*, Bilthoven, the Netherlands) in November 2019.

According to EC Regulation No. 2017/625 (EC, 2017), one of the tasks of the EURL-*Salmonella* is to organise PTs for the National Reference Laboratories for *Salmonella* (NRLs-*Salmonella*) in the European Union (EU). The main objectives for typing *Salmonella* strains are that this should be carried out uniformly in all Member States, and that comparable results should be obtained. The implementation of PTs on typing started in 1995.

A total of 35 laboratories participated in this study. These included 29 NRLs-*Salmonella* in the 28 EU Member States, three NRLs in EU candidate countries, and three NRLs in EFTA countries. The main objective of this study was to check the performance of the NRLs in serotyping *Salmonella*. All NRLs performed serotyping of the twenty obligatory strains, and all but five of the participants serotyped the optional 21st strain. Any NRLs of EU Member States that do not achieve the defined level of good performance for serotyping have to participate in a follow-up study.

This typing study included a pilot of an optional part on cluster analysis. The cluster analysis involved ten *Salmonella* strains and participants could perform it using either PFGE and/or MLVA and/or WGS or any combination of these methods, using their own routines.

A total of eighteen NRLs participated in the cluster analysis, with six (PFGE analysis), eight (MLVA analysis), and fourteen (WGS analysis).

2 Participants

Country	City	Institute
Albania	Tirana	Food Safety and Veterinary Institute
Austria	Graz	AGES
Belgium	Brussels	Sciensano
Bulgaria	Sofia	NDRVMI
Croatia	Zagreb	Croatian Veterinary Institute
Cyprus	Nicosia	Cyprus Veterinary Services
Czech Republic	Prague	State Veterinary Institute Prague
Denmark	Ringsted	Danish Veterinary and Food Administration (DVFA)
Estonia	Tartu	Veterinary and Food Laboratory
Finland	Kuopio	Finnish Food Authority
France	Maisons-Alfort	ANSES (Laboratoire de Sécurité des Aliments)
Germany	Berlin	German Federal Institute for Risk Assessment (BFR)
Greece	Chalkida	Veterinary Laboratory of Chalkis
Hungary	Budapest	National Food Chain Safety Office, Food Chain Safety Laboratory Directorate, Microbiological NRL
Iceland	Reykjavík	Landspítali University Hospital, Dept. of Clinical Microbiology
Ireland	Celbridge	Central Veterinary Research Laboratory
Italy	Legnaro	Istituto Zooprofilattico Sperimentale delle Venezie
Latvia	Riga	Institute of Food Safety, Animal Health and Environment (BIOR)
Lithuania	Vilnius	National Food and Veterinary Risk Assessment Institute
Luxembourg	Dudelange	Laboratoire National de Santé
Malta	Valletta	Malta Public Health Laboratory
Netherlands	Bilthoven	National Institute for Public Health and the Environment (RIVM), Centre for Infectious Diseases Research, Diagnostics and Screening (IDS)
North Macedonia Republic of	Skopje	Faculty of Veterinary Medicine – Microbiology of food and feed
Norway	Oslo	Norwegian Veterinary Institute
Poland	Pulawy	National Veterinary Research Institute
Portugal	Oeiras	INIAV-Instituto Nacional de Investigação Agrária e Veterinária
Romania	Bucharest	Institute for Diagnosis and Animal Health
Serbia	Belgrade	NIVS Institute of Veterinary Medicine of Serbia
Slovak Republic	Bratislava	State Veterinary and Food Institute
Slovenia	Ljubljana	UL, Veterinary Faculty, NVI
Spain	Algete-Madrid	Laboratorio Central de Veterinaria
Sweden	Uppsala	National Veterinary Institute (SVA)
Switzerland	Bern	Institute of Veterinary Bacteriology (ZOBA)
United Kingdom	Addlestone	Animal and Plant Health Agency (APHA)
United Kingdom	Belfast	AFBI – Northern Ireland

3 Materials and methods

3.1 Design of the Proficiency Test (PT)

3.1.1 Laboratory codes

Each NRL-*Salmonella* was randomly assigned a laboratory code between 1 and 35.

3.1.2 Protocol and test report

Three weeks before the start of the PT, the NRLs received the protocol by email. Web-based result forms were used to report results. Instructions for the completion of these result forms and data-entry were sent to the NRLs on 6 and 11 November 2019, in emails for serotyping and for the pilot study on cluster analysis, respectively.
The protocol and blank result forms can be found on the EURL-*Salmonella* website: https://www.eurlsalmonella.eu/proficiency-testing/typing-studies

3.1.3 Transport

The parcels containing the strains for serotyping and cluster analysis were sent by the EURL-*Salmonella* on 4 November 2019. All samples were packed and transported as Biological Substance Category B (UN 3373) and transported by a door-to-door courier service.

3.2 Serotyping part of the PT

3.2.1 Salmonella *strains for serotyping*

A total of twenty *Salmonella* strains (coded S1–S20) had to be serotyped by the participants. As agreed at the 24th EURL-*Salmonella* Workshop in Amersfoort, the Netherlands (Mooijman, 2019), a less common strain (S21) was additionally included. Testing this strain was optional and results were not included in the evaluation.

The *Salmonella* strains used for the part on serotyping originated from the National *Salmonella* Centre collection in the Netherlands. The strains were verified by the Centre before distribution. The complete antigenic formulas of the 21 serovars, in accordance with the most recent White-Kauffmann-Le Minor scheme (Grimont & Weill, 2007), are shown in Table 1. However, participants were asked to report only those results on which the identification of serovar names was based. Two strains (S10 and S18) represented serovars included in the EURL-*Salmonella* serotyping studies for the first time.

Table 1 Antigenic formulas of the 21 Salmonella *strains according to the White-Kauffmann-Le Minor scheme used in the EURL-*Salmonella *PT serotyping 2019*

Strain code	O-antigens	H-antigens		Serovar	Origin
		phase 1	phase 2		
S1	6,8	z_{10}	e,n,x	Hadar	Ringtrial
S2	1,4,[5],12	y	1,2	Coeln	Laying hen
S3	1,13,22	z	1,6	Poona	Human
S4	9,46	z_{38}	-	Fresno	Animal feed
S5	1,9,12	g,m	-	Enteritidis	Unknown
S6	3,{10}{15}{15,34}	e,h	l,w	Meleagridis	Animal feed
S7	1,4,[5],12,[27]	d	1,2	Stanley	Human
S8	1,4,[5],12	f,g,s	[1,2]	Agona	Laying hen
S9[a)]	1,4,[5],12	i	1,2	Typhimurium	Human
S10[b)]	6,7	r	1,6	Nigeria	Human
S11	3,{10},{15}	l,z_{13}	1,5	Uganda	Human
S12	11	r	e,n,x	Rubislaw	Non-human
S13	{6,7,14}{54}	g,m,[p],s	[1,2,7]	Montevideo	Human
S14[c)]	1,4,[5],12	i	-	1,4,[5],12:i:-	Human
S15	6,8	d	1,2	Muenchen	Human
S16	6,7,14	r	1,2	Virchow	Human
S17	6,7,14	r	1,5	Infantis	Human
S18[b)]	16	y	1,5	Saphra	Non-human
S19	1,13,23	i	l,w	Kedougou	Chicken
S20	1,4,[5],12,[27]	g,s,t	[1,2]	Kingston	Animal feed
S21[d)]	48	g,z_{51}	-	IV 48:g,z_{51}:-	Human

[a)] Potentially contaminated with *E. coli*. Results from strain S9 were excluded from the evaluation.
[b)] Represented in a EURL-*Salmonella* PT serotyping for the first time.
[c)] Typhimurium, monophasic variant as determined by PCR.
[d)] *Salmonella enterica* subspecies *houtenae* (optional strain).

3.2.2 *Evaluation of the serotyping results*

The evaluation of the various serotyping errors in this report is presented in Table 2.

Table 2 Evaluation of serotyping results

Results	Evaluation
Auto-agglutination or, Incomplete set of antisera (outside range of antisera)	Not typable
Partly typable due to incomplete set of antisera or, Part of the formula (for the name of the serovar) or, No serovar name	Partly correct
Wrong serovar or, Mixed sera formula	Incorrect

In 2007, the following criteria for 'good performance' in PTs on serotyping were defined (Mooijman, 2007). Penalty points are given for the incorrect typing of strains, but a distinction is made between the five most

important human health-related *Salmonella* serovars (as indicated in EU legislation, also sometimes referred to as 'top-5'), and all other strains:

- four penalty points: incorrect typing of *S*. Enteritidis, *S*. Typhimurium (including the monophasic variant), *S*. Hadar, *S*. Infantis or *S*. Virchow, or assigning the name of one of these five serovars to another strain;
- one penalty point: incorrect typing of all other *Salmonella* serovars.

The total number of penalty points is calculated for each NRL-*Salmonella*. The criterion for good performance is set at less than four penalty points. All EU Member State NRLs not meeting the criterion of good performance (four penalty points or more) have to participate in a follow-up study.

3.3 Cluster analysis part of the PT

3.3.1 Salmonella *strains for cluster analysis*

A total of ten *Salmonella* strains (coded SCA01 – SCA10) were included in this first pilot on cluster analysis. Background information on the strains is given in Table 3.
Strains had to be selected by the EURL-*Salmonella* to be suitable for analysis using either PFGE, MLVA, or WGS data. Based on their original PFGE (and MLVA) patterns in 2016 or before, a set of fifteen strains were re-cultured from storage and submitted for WGS analysis. Subsequently nine strains were selected for inclusion in the PT (see section 3.3.5, Figure 3). The tenth strain was to be a technical duplicate; strain SCA03 and strain SCA06 shipment tubes were both prepared from the same blood-agar plate containing original strain SCA03.

Table 3 Background information on the Salmonella *strains used for cluster analysis in 2019*

Strain code	Serovar	ST	MLVA-profile	Origin
SCA01 [a)]	4,5,12:i:-	34	3-13-9-NA-211	Human
SCA02	Typhimurium	19	3-16-17-18-311	Human
SCA03 [b)]	Typhimurium	19	3-16-7-17-311	Human
SCA04	Typhimurium	19	2-20-8-11-212	Human
SCA05 [a)]	4,5,12:i:-	34	3-11-9-NA-211	Human
SCA06 [b)]	Typhimurium	19	3-16-7-17-311	Human
SCA07	Typhimurium	19	5-9-14-9-211	Human
SCA08	Typhimurium	19	3-14-17-25-311	Human
SCA09 [a)]	4,5,12:i:-	34	3-13-9-NA-211	Human
SCA10	Typhimurium	19	2-12-7-9-212	Human

[a)] Typhimurium, monophasic variant as determined by PCR.
[b)] Technical duplicates (in bold).

3.3.2 *Evaluation of the cluster analysis results in general*

Cluster analysis was performed up to the choice of the participant by PFGE and/or MLVA and/or WGS (or any combination of these methods), using their own routines. Details of the method(s) used and the outcome of the cluster analysis had to be reported in the electronic result form. Additionally, specific data for PFGE and WGS had to be sent by email or uploaded to a secure ftp server.

Evaluation (per methodology, see sections 3.3.3 – 3.3.5) of the participants' cluster analysis results was done on the ability to correctly identify cluster(s) of genetically closely related isolates, as pre-defined by the EURL-*Salmonella*, and referred to as "expected results". However, cluster definitions may vary depending on the situation or the specific research question, e.g., in outbreak investigations or surveillance. The participants of this pilot PT were free to use their own interpretation of "cluster(s) of closely related isolates". Therefore, no performance criteria were set for this pilot PT on cluster analysis. As a minimum, it was expected that the participants would report the technical duplicate strains SCA03 and SCA06 to be (part of) one cluster.

3.3.3 Evaluation of the cluster analysis results based on PFGE data

Data submission for PFGE results included:

- **Electronic result form**: protocol used, position of the lanes, total number of bands per profile, cluster identification.
- **The PFGE gel image** had to be emailed as *an uncompressed 8-bit grey-scale TIFF file* to the EURL-*Salmonella*. The laboratory code had to be included in the name of the .tif file, for example: Lab01_PFGE2019.tif.
- The ZIP (BioNumerics 7) or XML export files (BioNumerics 6 or below) were prepared from the analysis in BioNumerics, *including all test strains and reference strains, as well as the TIFF image*. **The BN analysis data had to be emailed in a ZIP file** to the EURL-*Salmonella*. The zip file had to include the laboratory code in the name, for example: Lab01_PFGE2019.zip.

The original EURL-*Salmonella* reference (REF) cluster analysis based on PFGE data (2016 or before) is shown in Figure 1. PFGE data were analysed in BioNumerics, and similarity was calculated using the Dice coefficient, with both tolerance and optimization at 1,5%. The expected results therefore would be to find two clusters of closely related strains, consisting of the four strains SCA02, SCA03/SCA06 (technical duplicates, so only one PFGE profile available), SCA08 and of the three strains SCA01, SCA05, SCA09.
However, based on the combined results of the 6 participants partaking in the PFGE part of the study, the REF cluster analysis was updated using the appropriate profiles from several participants (Figure 2 and Annex 7). Strain SCA03 and its duplicate SCA06 especially showed a change in the PFGE profile, leading to a separate cluster. The three strains in cluster SCA01/SCA05/SCA09 showed more variability than originally and no longer formed a closely related cluster. The expected results were updated to find two clusters of closely related strains, consisting of the SCA02 and SCA08 strains, and the SCA03/SCA06 strains (technical duplicates) (Figure 2).

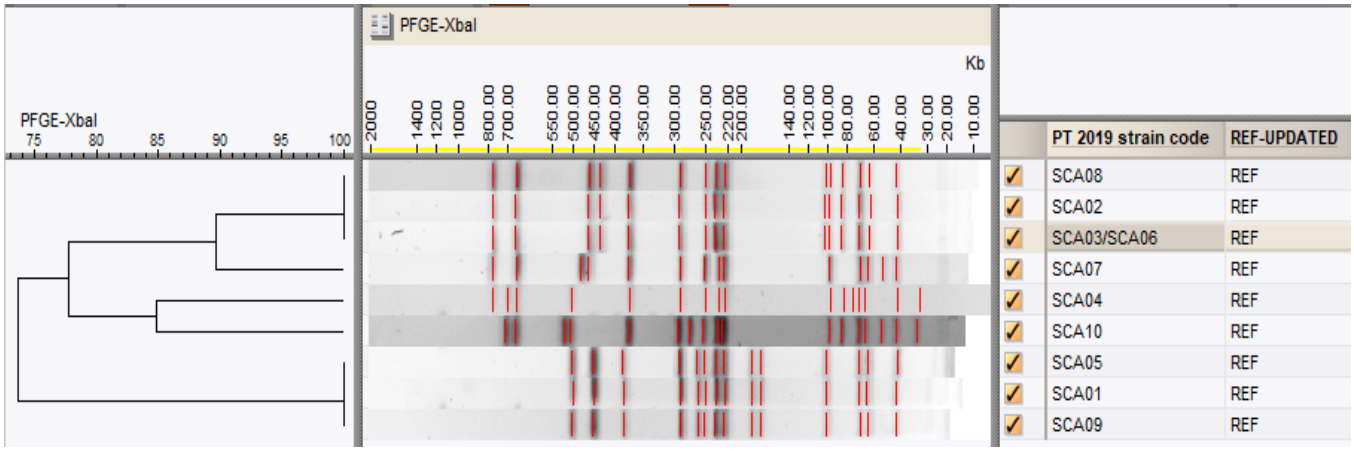


*Figure 1 EURL-*Salmonella *(REF) cluster analysis based on original PFGE data*

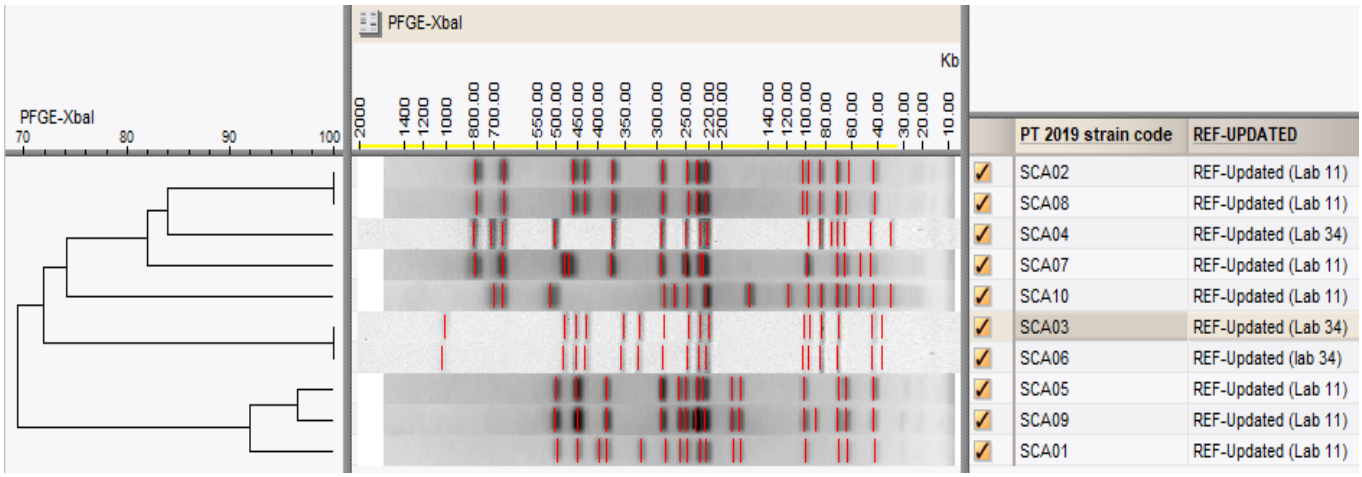


Figure 2 Updated REF cluster analysis based on participants' PFGE data

3.3.4 Evaluation of the cluster analysis results based on MLVA data

Data submission for MLVA results included:

- **Electronic result form**: scheme/loci used, the allelic profile, cluster identification.

The EURL-*Salmonella* reference (REF) cluster analysis based on MLVA data is shown in Table 4.
Expected results were two clusters of closely related strains, consisting of the SCA01 and SCA09 strains and the SCA03 and SCA06 strains (technical duplicates).

*Table 4 EURL-*Salmonella *(REF) cluster analysis based on MLVA data*

MLVA-profile	Strain code	REF cluster identification
2-12-7-9-212	SCA10	
2-20-8-11-212	SCA04	
3-11-9-NA-211	SCA05	
3-13-9-NA-211	SCA01	Cluster 2, MLVA-based analysis
3-13-9-NA-211	SCA09	Cluster 2, MLVA-based analysis
3-14-17-25-311	SCA08	
3-16-7-17-311	SCA03	Cluster 1, MLVA-based analysis
3-16-7-17-311	SCA06	Cluster 1, MLVA-based analysis
3-16-17-18-311	SCA02	
5-9-14-9-211	SCA07	

3.3.5 Evaluation of the cluster analysis results based on WGS data

Data submission for WGS results included:

- ***Electronic result form****:* background information on the wet-lab and dry-lab methods used, cluster identification (SNP-based or cgMLST/wgMLST-based).
- ***Raw reads*** *(fastq files) uploaded to the secure ftp server according to the instructions.* The file names had to include the laboratory code and strain code in the name, for example: Lab01_SCA01_R1.fastq, Lab01_SCA01_R2.fastq, etc.

The original EURL-*Salmonella* reference (REF) cluster analysis based on WGS data is shown in Figure 3. REF sequencing was performed externally on an Illumina NovaSeq platform. Raw data were processed via an in-house developed pipeline (https://github.com/Papos92), which includes the SPAdes assembler. Cluster analysis was done in Ridom SeqSphere^{+}, using the cgMLST Enterobase v2.0 scheme and visualised in a Minimum Spanning Tree (MST, Figure 3).
The expected results therefore would have been to find one cluster of closely related strains, consisting of the SCA02, and SCA03/SCA06 strains (technical duplicates, so only 1 initial WGS profile available). Regrettably, strain SCA03 and its technical duplicate SCA06 showed far more variety than expected (see section 4.3.4, Figure 10 and Figure 11). This was observed during the PT in November 2019, and after storage (room temperature) of the transport tubes containing the strains till February 2020 (Figure 11: EL and EL2, respectively). Nevertheless, the cluster definition for the WGS-based cluster 1 was retained and included strains SCA02, SCA03, and SCA06.

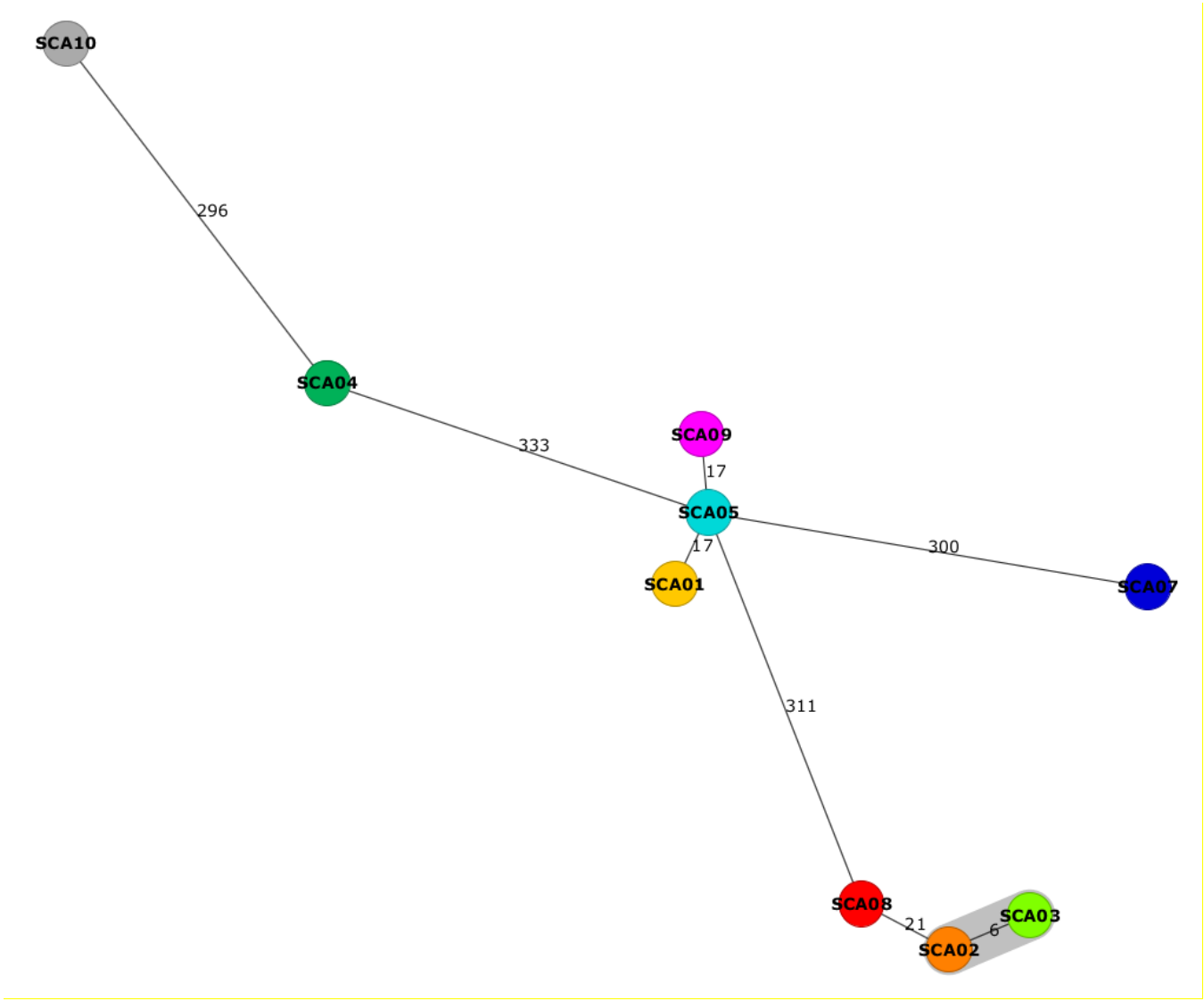


*Figure 3 EURL-*Salmonella *(REF) cluster analysis based on WGS data, Cluster Alert (in grey background) in this study set at 7 allelic differences*

4 Results and Discussion

4.1 Technical data

4.1.1 General

A total of 35 laboratories participated in this study (Chapter 2). These included 29 NRLs-*Salmonella* in the 28 EU Member States, three NRLs in EU candidate countries, and three NRLs in EFTA countries.
The frequency of *Salmonella* serotyping at the participating laboratories and the number of strains serotyped in 2019 are summarised in Table 5.

Table 5 Frequency and number of Salmonella *strains serotyped in 2019*

Laboratory code	Serotyping frequency in 2019	No. of strains serotyped in 2019
12	Daily	70
15	Daily	140
10	Daily	169
24	Daily	200
9	Daily	302
6	Daily	350
3	Daily	400
33	Daily	410
1	Daily	470
21	Daily	500
5	Daily	600
13	Daily	600
35	Daily	600
30	Daily	800
29	Daily	900
20	Daily	950
25	Daily	1000
16	Daily	1600
14	Daily	2500
31	Daily	3300
22	Daily	4000
34	Daily	4500
28	Daily	5000
11	Daily	7000
32	Daily	7300
8	Twice a week	52
18	Twice a week	200
27	Twice a week	750
23	Thrice a week	30
7	Thrice a week	130
19	Thrice a week	130
17	Thrice a week	250
2	Once a week	1300
26	Once a week	2500
4	Monthly	30
n=35		49033

4.1.2 *Accreditation*

Of the 35 participants, 33 are accredited for serotyping *Salmonella*, mainly according to EN ISO/IEC 17025, and in some cases (also) according to EN ISO 15189. One laboratory mentioned ISO 6579-1 only. One laboratory noted that it will be seeking accreditation of *Salmonella* serotyping in the near future, it is known that the other laboratory will not do so.
One laboratory is accredited for serotyping *S*. *enterica* subsp. *enterica*. All other laboratories stated that they are accredited for all *Salmonella* serovars.

4.1.3 *Transport of samples*

All but one participant received their package in the same week as sent (week 45, 2019). The exception was received on 18 November 2019, due to a delay at customs. All packages were received in good condition.

4.2 Serotyping results

4.2.1 *General*

The twenty obligatory strains were all tested by the NRLs-*Salmonella* in the participating countries. Classical serology was used by 28 participants; another six mentioned the combined use of classical serology and Luminex assays (3) or multiplex/real time PCR (3). One participant used Whole Genome Sequencing.
Details on the number and the source of the sera used by the participants are summarised in Tables 6 and 7.

Table 6 Number of laboratories using sera from various manufacturers

Manufacturer	Number of NRLs (n=33)
Biorad	16
Own preparation	1
Pro-Lab	6
Remel	3
Sifin	18
Statens Serum Institute (SSI)	28
Other	3

Table 7 Number of laboratories using sera from one or more manufacturers and/or in-house prepared sera

Number of manufacturers from which sera are obtained (including in-house preparations)	Number of NRLs (n=33)
1	8
2	13
3	8
4	3
5	1

4.2.2 *Biochemical testing*

Twenty-six participants indicated the use of biochemical tests for confirmation. Sometimes these were used routinely for all strains, but mostly they used appropriate tests for some specific strains.
Twelve participants used a variety of biochemical tests (most often malonate and ONPG) on the optional strain S21 only.

4.2.3 *Use of PCR for confirmation*

Fifteen laboratories used PCR to confirm strain S14, the monophasic variant of *S*. Typhimurium 1,4,[5],12:i:-, and five of these also used PCR to confirm strain S9, *S*. Typhimurium. The majority of laboratories mentioned using the reference by Tennant et al., 2010.

4.2.4 *General comments on the PT 2019 serotyping evaluation*

Selection, preparation and shipment of the strains to the participants is always carried out with upmost care, and includes various quality control steps, including purity and typeability. This year however, at least ten participants reported that strain S9 (*S*. Typhimurium) was a mixed culture. Internal investigations confirmed that this strain was contaminated with *E. coli* at a very low level. Therefore, results for strain S9 were excluded from the evaluations.

4.2.5 *Serotyping results per laboratory*

The evaluation of the type of errors for O- and H-antigens and for identification of the strains are shown in Figures 4, 5 and 6. The percentages of correct results per laboratory are shown in Figure 7. The O-antigens were typed correctly by 32 of the 35 participants (91%). This corresponds to 99% of the total number of strains. The H-antigens were typed correctly by 27 of the 35 participants (77%), corresponding to 97% of the total number of strains. As a result, 26 participants (74%) gave the correct serovar names, corresponding to 97% of all strains evaluated.

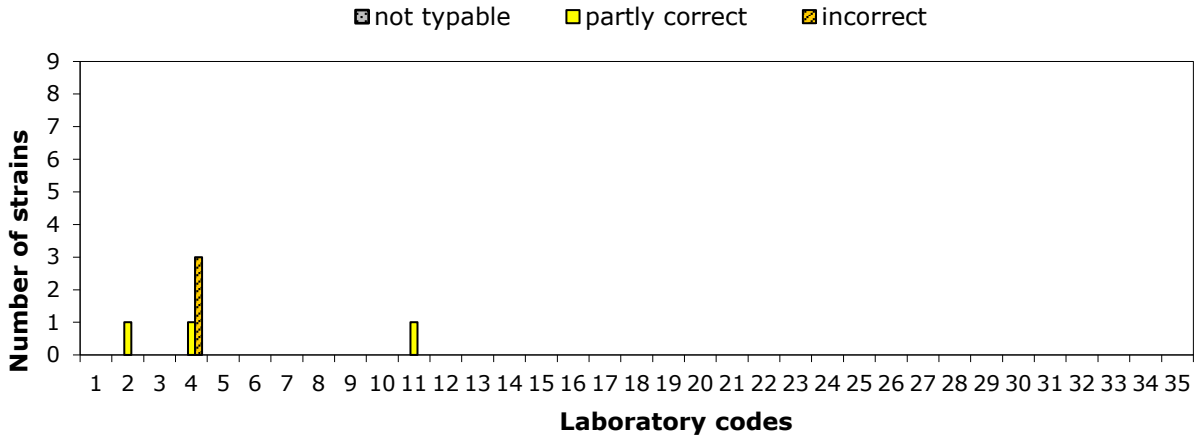


Figure 4 Evaluation of type of errors for O-antigens, per NRL

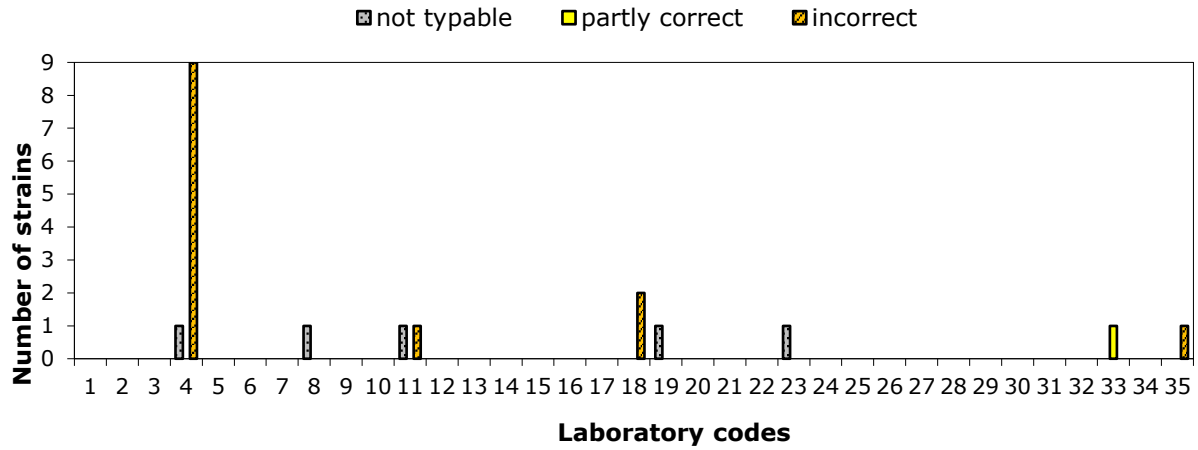


Figure 5 Evaluation of type of errors for H-antigens, per NRL

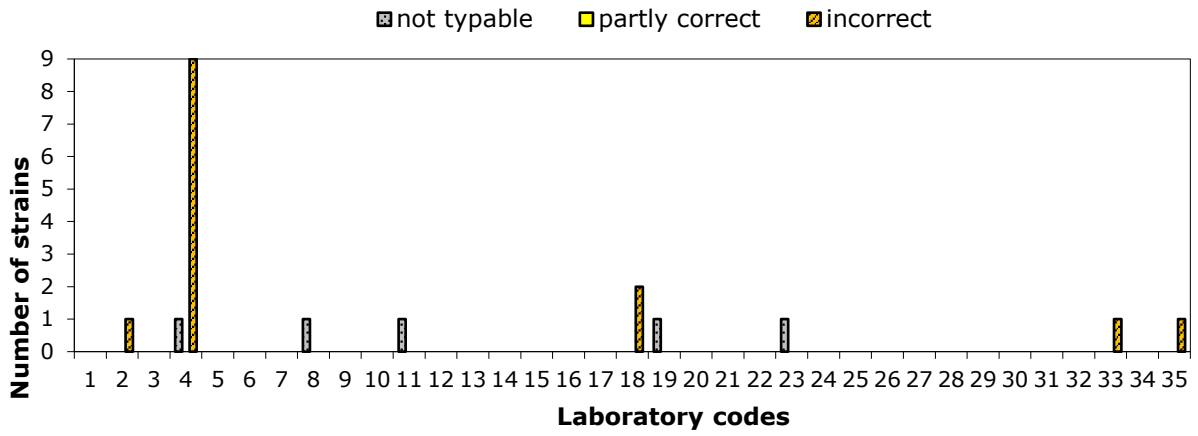


Figure 6 Evaluation of the type of errors in the identification of the serovar names, per NRL

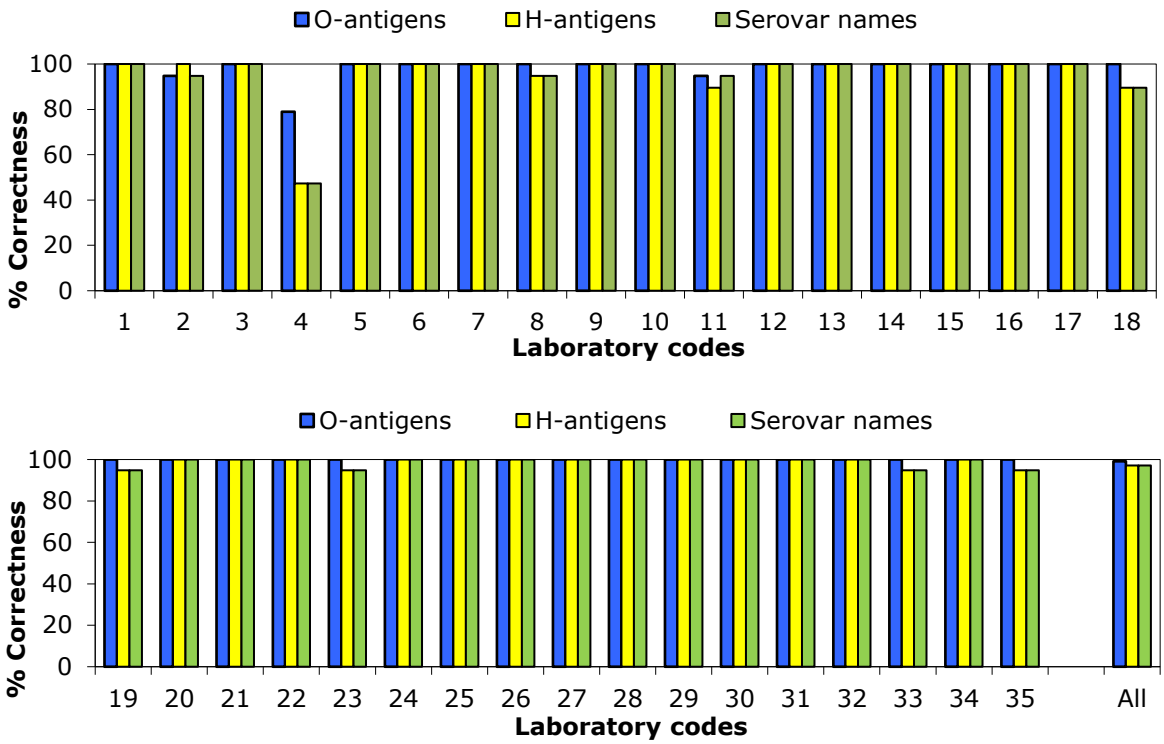


Figure 7 Percentages of correct serotyping results, per NRL

4.2.6 *Performance of the participants*

The number of penalty points was determined for each NRL using the guidelines described in Section 3.2.2. Table 8 shows the number of penalty points for each NRL and indicates whether the level of good performance was achieved (yes or no).

Overall, the performance of the NRLs in the PT Serotyping 2019 was very good. Notably, Lab 4 was an exception (eighteen penalty points), but this NRL is new and still learning about serotyping; this was its second participation.

Two participants (Lab 4 and Lab 35) did not meet the level of good performance at the first stage of the study. The online follow-up on this performed by the first participant was satisfactory, but the proposed training session for the second participant had to be put on hold due to the ongoing COVID-19 pandemic.

All participants received their individual laboratory evaluation report as well as the interim summary report on serotyping on 25 February 2020. An example of an individual laboratory evaluation report on serotyping results is given in Annex 1. The interim summary report is available on the EURL-*Salmonella* website: www.eurlsalmonella.eu/publications/proficiency-test-reports.

Table 8 Evaluation of serotyping results per NRL

Laboratory code	Penalty points	Good performance	Laboratory code	Penalty points	Good performance
1	0	yes	19	0	yes
2	1	yes	20	0	yes
3	0	yes	21	0	yes
4	18	NO	22	0	yes
5	0	yes	23	0	yes
6	0	yes	24	0	yes
7	0	yes	25	0	yes
8	0	yes	26	0	yes
9	0	yes	27	0	yes
10	0	yes	28	0	yes
11	0	yes	29	0	yes
12	0	yes	30	0	yes
13	0	yes	31	0	yes
14	0	yes	32	0	yes
15	0	yes	33	1	yes
16	0	yes	34	0	yes
17	0	yes	35	4	NO
18	2	yes			

4.2.7 *Serotyping results per strain*

The overall results reported per strain (S1 – S20) and per laboratory are given in Annex 2. Details on the strains that caused problems in serotyping are shown in Annex 3.
A completely correct identification was obtained for seven *Salmonella* serovars: Poona (S3), Enteritidis (S5), Montevideo (S13), Virchow (S16), Infantis (S17), Saphra (S18), and Kingston (S20). Another eight serovars would have been completely correct, should the results of Lab 4 (new participant) have been excluded: Coeln (S2), Meleagridis (S6), Stanley (S7), Agona (S8), Nigeria (S10), Rubislaw (S12), 1,4,[5],12:i:- (S14), and Kedougou (S19).
Five participants did not have access to the required but less common antiserum z38 for strain S4 (S. Fresno, 9,46:z_{38}:-) leading to the five 'not typable' results for the H-antigens and the serovar names (Figures 5 and 6).
The variety in reported serovar names for strain 1,4,[5],12:i:- (S14) are shown in Annex 2. Fifteen participants used a PCR method to confirm this strain as a monophasic *S*. Typhimurium strain.

Details on the additional and optional strain S21 are given in Annex 4. All but five participants tried to serotype strain S21, a *Salmonella enterica* subsp. *houtenae* (IV). Some laboratories did not have access to the required antisera to finalise this (48:g,z_{51}:-).

4.2.8 *Trend analysis of the serotyping results of the EU NRLs*

Historical data for all participants of the EURL-*Salmonella* PTs on the serotyping of *Salmonella* can be found on the EURL-*Salmonella* website: www.eurlsalmonella.eu/publications/proficiency-test-reports.

The historical data on the EU NRLs only are visualised in Figure 8, showing the percentages of correctly typed strains, and in Figure 9, showing the number of Penalty Points and non-Good Performance. The percentages of correctly typed strains are stable over time, usually showing a better performance for the O-antigens than for the H-antigens. The number of Penalty Points has clearly declined, from 35 points when this system started in 2007, to four points in the 2019 study. The rise seen in the 2018 study was mainly caused by the relatively large number of seven EU NRLs that made a mistake in typing an *S.* Cannstatt strain. Moreover, the number of EU NRLs with a non-Good Performance is low: two in the period 2010–2013, one in the 2014, 2015 and 2018 studies, and none in the 2016, 2017 and 2019 studies.

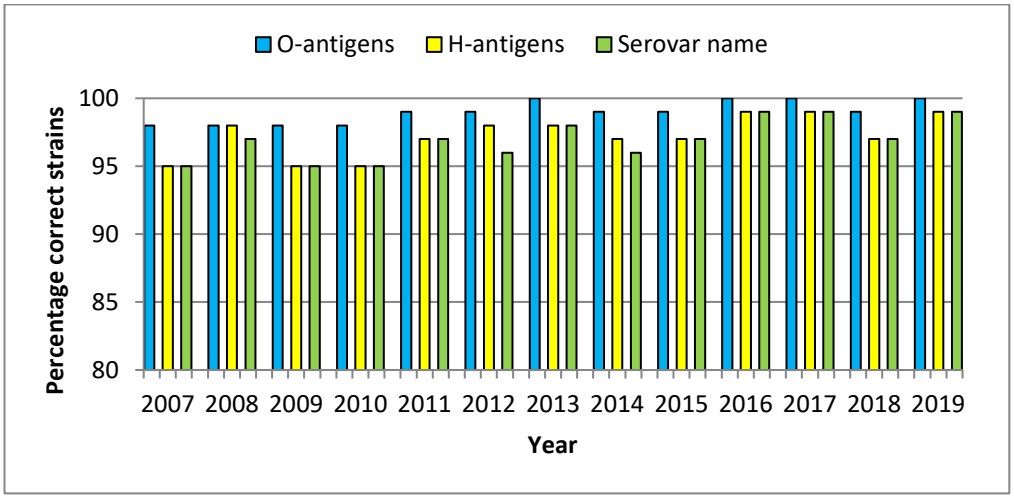


Figure 8 Serotyping results of the EU NRLs, showing the percentages of correctly typed strains

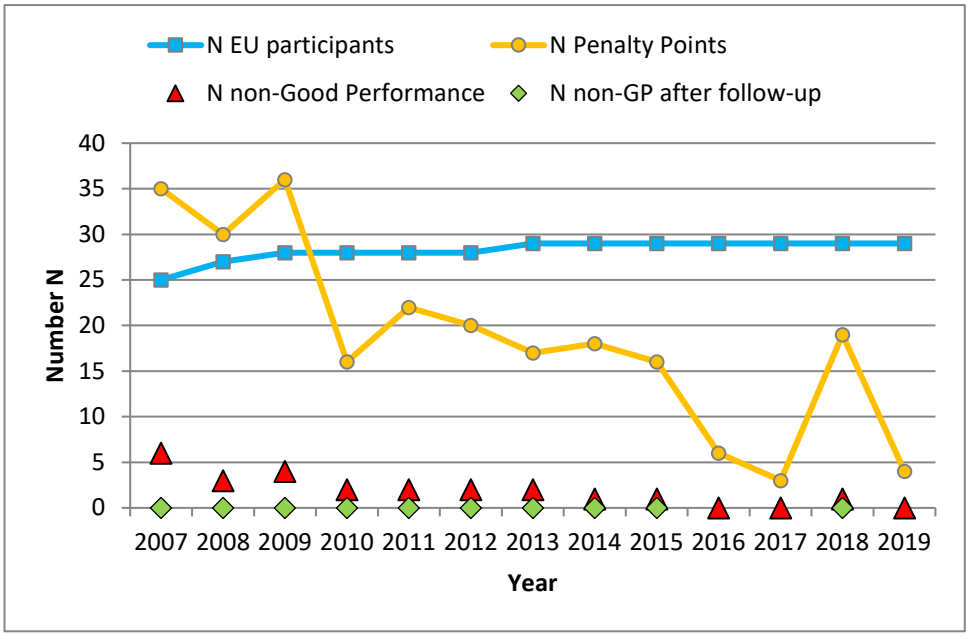


Figure 9 Serotyping results of the EU NRLs, showing the number of Penalty Points and non-Good Performance (non-GP)

4.3 Cluster analysis results

4.3.1 *General*

Participants conducted the cluster analysis using either PFGE and/or MLVA and/or WGS (or any combination of these methods), using their own routines.
A total of eighteen NRLs participated in the cluster analysis;
six participants used PFGE analysis, eight used MLVA, and fourteen used WGS (Table 9).

All participants received their individual laboratory evaluation of this first pilot on cluster analysis on 16 June 2020, together with the interim summary report on the overall results. An example of an individual laboratory evaluation report on cluster analysis results is given in Annex 5. The interim summary report is available on the EURL-*Salmonella* website:
https://www.eurlsalmonella.eu/publications/proficiency-test-reports.

As a general question, the participants were asked if they serotyped the ten strains. Results received from 12 participants are given in Annex 6, for information purposes only.

Table 9 Participation in Cluster analysis in 2019, per method or combination of methods used

Method used:			Number of participants	Laboratory codes
PFGE			2	8, 25
	MLVA		1	13
		WGS	8	9, 15, 16, 17, 27, 28, 29, 32
PFGE	MLVA		1	21
	MLVA	WGS	3	20, 22, 26
PFGE	MLVA	WGS	3	11, 14, 34
Total PFGE:	Total MLVA:	Total WGS:	Total overall:	
6	8	14	18	

4.3.2 *Results cluster analysis based on PFGE data*

Six participants (Laboratory codes 8, 11, 14, 21, 25, 34) submitted cluster analysis results based on PFGE data. One participant (Laboratory code 25) submitted the PFGE image only, the other five participants also included their results of the gel analysis in BioNumerics (Annex 7).

The expected results were updated to find **two** clusters of closely related strains, consisting of the SCA02 and SCA08 strains, and of the SCA03/SCA06 strains (technical duplicates) (see Figure 2).

The number of clusters, and their identification reported by the six PFGE participants are given in Table 10.

Table 10 Number of clusters, and their identification as reported by the 6 PFGE participants

		Cluster					
Lab code	**Number of clusters reported**	**1**	**2**	**3**	**4**	**5**	**Cluster identification as expected**
REF Updated	**2**	**SCA03 SCA06**	**SCA02 SCA08**				**REF Updated**
11	1		SCA02 SCA08				No
8	2	SCA03 SCA06	SCA02 SCA08				Yes
25	2	SCA02 SCA08 SCA07 SCA03 SCA06 SCA04 SCA10		SCA01 SCA09 SCA05			No
21	3	SCA03 SCA06	SCA02 SCA08	SCA01 SCA09 SCA05			No
34	3	SCA03 SCA06	SCA02 SCA08	SCA01 SCA09 SCA05			No
14	5	SCA03 SCA06	SCA02 SCA08 SCA07	SCA01 SCA09 SCA05	SCA04	SCA10	No

SCA03 SCA06	Cluster 1, PFGE-based analysis
SCA02 SCA08	Cluster 2, PFGE-based analysis

Note: the order of the clusters reported by participants has been rearranged for easier reading and comparison.

One participant (Laboratory code 8) reported the cluster identification completely as expected. Four participants also indicated the three SCA01, SCA05, SCA09 strains as a cluster, which may be due to a slight difference in their settings or criteria for defining a PFGE cluster. Laboratories 14 and 25 interpreted the cluster analysis differently; they assigned all strains into clusters.

The technical duplicates SCA03/SCA06 were expected to be assigned within one cluster; this was reported by 5 of the 6 participants (Table 10).

4.3.3 *Results cluster analysis based on MLVA data*

Eight participants (Laboratory codes 11, 13, 14, 20, 21, 22, 26, 34) submitted MLVA data; seven also reported cluster analysis results.

All eight participants basically used the "Lindstedt" scheme (ECDC, 2011; Larsson et al., 2009; Lindstedt et al., 2004), reporting the loci in the order: STTR9, STTR5, STTR6, STTR10, STTR3. The allelic profiles submitted by the participants are given in Annex 8.
The EURL-*Salmonella* reference (REF) cluster analysis based on MLVA data is shown in Table 4. The expected results were to report two clusters of closely related strains, consisting of the two strains SCA01 and SCA09 and of the two strains SCA03 and SCA06 (technical duplicates).
The number of clusters, and their identification reported by the eight MLVA participants are given in Table 11.

Table 11 Number of clusters, and their identification reported by the 8 MLVA participants

		Cluster					
Lab code	**Number of clusters reported**	**1**	**2**	**3**	**4**	**5**	**Cluster identification as expected**
REF	**2**	**SCA03 SCA06**	**SCA01 SCA09**				**REF**
11	1	SCA03 SCA06					No
13	2	SCA03 SCA06	SCA01 SCA09 SCA05				No
20	2	SCA03 SCA06	SCA01 SCA09				Yes
22	2	SCA03 SCA06	SCA01 SCA09				Yes
26	2	SCA03 SCA06	SCA01 SCA09				Yes
34	3	SCA03 SCA06	SCA01 SCA05 SCA09	SCA02 SCA08			No
14	5	SCA08 SCA02 SCA03 SCA06	SCA05 SCA01 SCA09	SCA07	SCA10	SCA04	No
21	Not done						not applicable

SCA03 SCA06	Cluster 1, MLVA-based analysis
SCA01 SCA09	Cluster 2, MLVA-based analysis

Note: the order of the clusters reported by participants has been rearranged for easier reading and comparison.

Three participants (Laboratory codes 20, 22, 26) of the seven cluster analysis submissions reported the cluster identification as expected. Laboratories 13, 14, and 34 appeared to use slightly wider criteria for cluster 2, and also included strain SCA05.
Laboratory 11 correctly reported cluster 1, but did not report cluster 2, probably due to the unexpected MLVA typing result for strain SCA09 (Annex 8).
Laboratory 14 interpreted the cluster analysis differently, assigning all strains into clusters.

The technical duplicates SCA03/SCA06 were expected to be assigned in one cluster and this was reported by all 7 reporting participants (Table 11).

4.3.4 Results cluster analysis based on WGS data

Fourteen participants (Laboratory codes 9, 11, 14, 15, 16, 17, 20, 22, 26, 27, 28, 29, 32, 34) submitted cluster analysis results based on WGS data. For extended comparison, data of one additional WGS participant (Laboratory code 91) are also shown in Figures 10 and 11.

General details on the wet-lab and dry-lab protocols performed by the participants are given in Annex 9. All participants performed DNA extraction, library preparation, and sequencing in-house, except for participant 9 (sequencing outsourced), participant 28 (library preparation and sequencing outsourced), and the EURL-*Salmonella* (all outsourced). The Illumina MiSeq platform was used most often (10x), followed by the Illumina NextSeq (3x), and Illumina MiniSeq and NovaSeq (1x each). Including the EURL-*Salmonella*, 10 participants used cgMLST for data analysis and 5 used SNP-based analysis (4x reference-based and 1x assembly-based). Tools used for this analysis varied from in-house pipelines to commercial ones, most commonly Ridom SeqSphere (7x). Both Neighbor joining (NJ, 7x) and Minimum Spanning Tree (MST, 5x) were commonly used for the analysis of clusters.

All participants' Quality Criteria (QC) parameters reported for the evaluation of their data are listed in Annex 10. A variety in naming these QC parameters and the thresholds used, was observed. An overview of the most widely used (names of) parameters is given in Table 12.

Table 12 Participants' most widely used QC parameters

X[a)]	QL/ QT[b)]	Criterion as mentioned	Threshold (if applicable)
1x:	QL	Assembled Genome Length	Approx. 4.9 ± 10%
1x:	QT	Assembly length	~ 5 Mbases for *Salmonella*
1x:	QT	Contigs assembly size	>4.5 Mbases AND <6.0 Mbases
1x:	QT	Length of contig assembly	< reference genome + 10%
2x:	QT	Total length	4.5-5.5 kb for *Salmonella*
6x:	QL	Confirmation of genus (e.g., MLST/Python, K-mer finder, KRAKEN2)	
4x:	QL	Serotyping (e.g., SISTR, SeqSero)	

X[a)]	QL/ QT[b)]	Criterion as mentioned	Threshold (if applicable)
7x:	QL	Contamination check (e.g., Kmer ID, Confindr, KRAKEN)	
2x:	QL	GC content	%/similar between strains
1x:	QL	GC percentage	51.9-52.2%
1x:	QT	GC%	>49% AND <53%
1x:	QL	MLST (Bionumerics)	7 loci Achtman scheme
1x:	QL	Perc. Good cgMLST Targets	>99%
1x:	QL	Percent matching targets in S. enterica cgMLST scheme	More than 90%
1x:	QT	% of good cgMLST targets	>90% of 3002 targets
1x:	QT	7-loci MLST mean coverage	30x
1x:	QT	Allele calling	cgMLST found and called >95%
1x:	QT	Allele calling result	percentage of good targets ~98%
1x:	QT	cgMLST genes found	>95%
1x:	QT	Core Genome	At least 98%
1x:	QT	Reference coverage	>90%
1x:	QT	Average assembly coverage	more than 10 reads
1x:	QT	Coverage	10x
1x:	QT	Coverage	minimum 20-30x
1x:	QT	Coverage	>25x
4x:	QT	Coverage (depth)	>30x
1x:	QT	Coverage	min 30x, max 100x
2x:	QT	Coverage	50x
1x:	QT	Mean coverage	10
1x:	QT	Median coverage	>20x
1x:	QT	raw reads theoretical coverage	>30x
1x:	QT	N50	>10 000
1x:	QT	N50	>15 000
1x:	QT	N50	>30 000
2x:	QT	N50	50 000
1x:	QT	N50	80 000
1x:	QT	N50	>100 000
1x:	QT	N50	>400 000
1x:	QT	Number of contigs assembly	<60
1x:	QT	Number of contigs	<200
1x:	QT	Number of contigs	200 bases (contigs shorter than 200 bases have to be ignored)
1x:	QT	Number of contigs	250
1x:	QT	Number of contigs	<300
1x:	QT	Number of contigs	400 or less
1x:	QT	Number of contigs	<500

[a)] X: number of participants mentioning this criterion.
[b)] QL/QT: Qualitative or Quantitative reported criterion.

All but one of the participants' raw data (fastq files) were successfully processed through the in-house assembly pipeline as discussed in section 3.3.5. Raw data from participant 28 were processed using the CLC Genomic Workbench software package (Qiagen) because of unspecified technical problems using the other pipeline. All de novo assembled genomes (fasta files) were analysed in Ridom SeqSphere+ using the cgMLST Enterobase v2.0 and visualised in an MST (Figure 10). Data per strain are given in Annex 11. Figure 11 shows the MST for the WGS-based cluster 1 strains SCA02, SCA03/SCA06 (technical duplicates) and the close strain SCA08, indicating the participants' laboratory codes in numbers.

Note that results for Laboratory 17 indicate a swap between the SCA01 and SCA02 strains, as well as for SCA09 and SCA10. Results for Laboratory 16 indicate a swap between the SCA02 and SCA03 strains. Laboratory 11 most likely tested/analysed strain SCA01 twice: both as SCA01 and SCA02 (Figures 10 and 11; Annex 10).

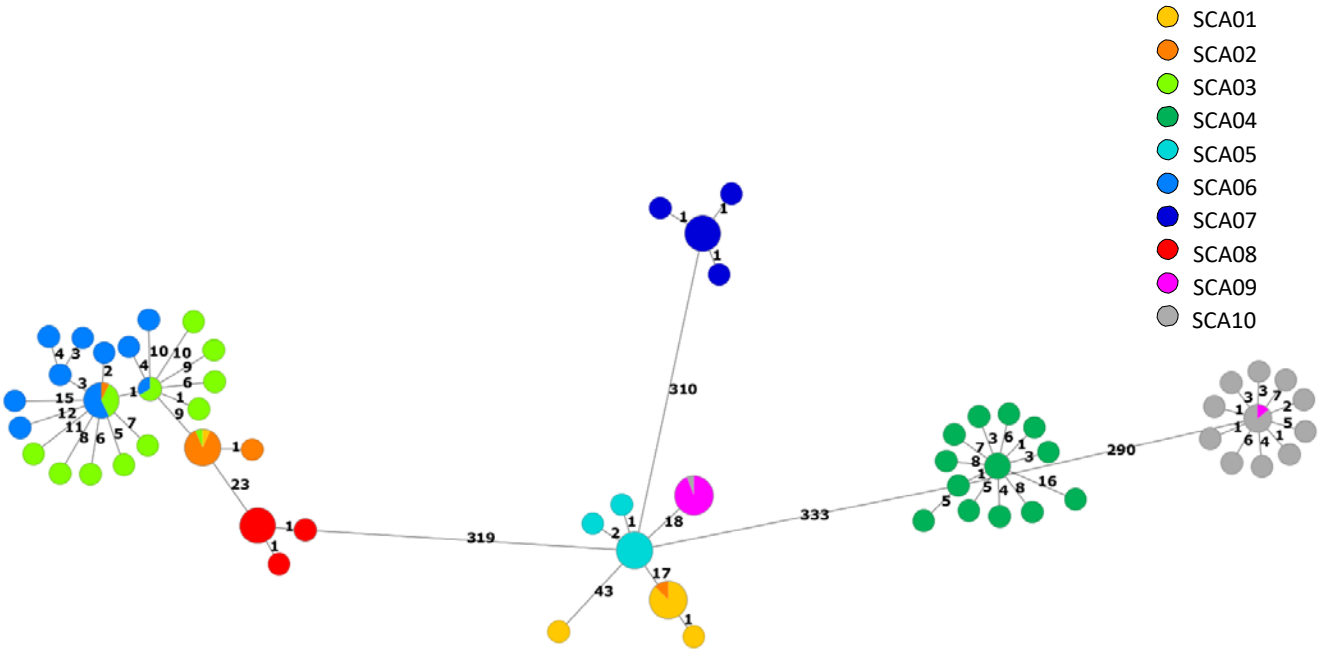


Figure 10 MST of all strains from all participants' processed raw data (Ridom SeqSphere⁺, cgMLST including all 3002 targets, pairwise ignoring missing values)

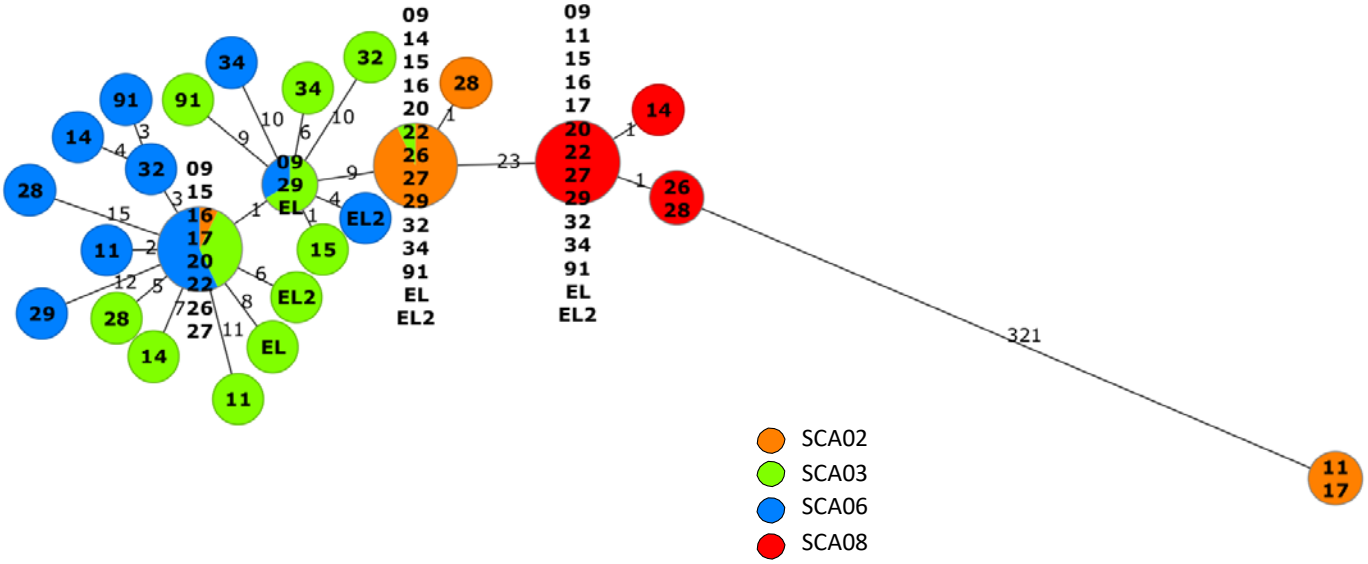


Figure 11 MST of strains SCA02, SCA03, SCA06, SCA08 from all participants' processed raw data (Ridom SeqSphere+, cgMLST including all 3002 targets, pairwise ignoring missing values)

An overview of the main QC parameter results on all in-house *de novo* assembled genomes (fasta files) is given in Table 13, summarised per participant and in Table 14, summarised per strain. Detailed data per participant are given in Annex 12.

Table 13 Results QC parameters on the de novo *assembled genomes, average per participant*

Laboratory code	Average # contigs	Average Largest contig	Average Total length	Average N50	Average Coverage
9	77	746285	4966195	252849	64
11	95	772101	4977863	269210	73
14	132	280709	4951139	91549	62
15	75	541405	4945110	220412	58
16	69	745894	4962087	255760	82
17	135	647108	4939261	212311	96
20	106	541993	4958067	156049	106
22	82	582170	4962178	195220	96
26	75	526131	4945792	212587	339
27	62	730576	4958246	255200	95
28	62	614189	4947408	231572	320
29	72	720350	4962176	265244	138
32	84	602429	4974861	217966	316
34	61	763205	4958744	266594	73
91	76	657373	4945441	209909	89
EL0(7/10/2019)	63	727410	4948463	234375	516
EL1(25/11/2019)	70	692066	4949153	204745	253
EL2(10/2/2020)	66	709503	4948085	237410	172

Table 14 Results QC parameters on the de novo *assembled genomes, average per strain*

Strain number	Average # contigs	Average Largest contig	Average Total length	Average N50	Average Coverage
SCA01	79	726547	4945761	222491	154
SCA02	85	621120	4976039	199719	151
SCA03	89	566499	4963979	186213	166
SCA04	112	532857	4906870	198586	171
SCA05	81	770596	5028401	258252	178
SCA06	80	580958	4962743	198571	172
SCA07	63	731355	4893399	268010	178
SCA08	75	599542	4972543	207024	168
SCA09	78	768716	4986241	258316	142
SCA10	70	552610	4918047	220172	158

The expected results for WGS-based cluster analysis were retained to find one cluster of closely related strains, consisting of the three strains SCA02, SCA03, and SCA06 (section 3.3.5; Figures 10 and 11).

The number of clusters, and their identification as reported by the fourteen WGS participants are given in Table 15.

One participant (Laboratory code 27) reported the cluster identification completely as expected and noted that they considered the three strains related in terms of outbreak investigation.
However, participants were free to use their own definition of cluster analysis and "closely related strains", so especially with WGS-derived data, many differences were noted regarding interpretation of the cluster identification.
Five participants reported strains SCA03/SCA06 as the two strains belonging to one cluster. Seven participants reported strains SCA03/SCA06 belonging to one cluster, but also included one or more strains in this cluster. These latter participants also reported strains SCA01, SCA05, and SCA09 as a second cluster of closely related strains. Laboratories 14, 28 and 32 interpreted the cluster analysis differently, by assigning all strains into clusters.

The technical duplicates SCA03/SCA06 were expected to be assigned within 1 cluster and this was reported by 12 of the 14 participants (Table 15). Nevertheless, this may have been hampered by the unexpected but variable nature of the original strain. For example, Laboratory code 11 noted: "WGS: strain LAB11_SCA03; LAB11_SCA06 were very similar, but did not meet the threshold, in reality we would take other factors, such as epidemiological data in consideration."

Table 15 Number of clusters, and their identification as reported by the 14 WGS participants

Lab code (method used)	Number of clusters reported	Cluster 1	Cluster 2	Cluster 3	Cluster identification as expected
REF (cg-MLST)	**1**	**SCA02 SCA03 SCA06**			**REF**
29 (cg-MLST)	0				No
11 (cg-MLST)	1	SCA01 SCA02			No
15 (cg-MLST)	1	SCA03 SCA06			No
17 (cg-MLST)	1	SCA03 SCA06			No
20 (SNP-reference)	1	SCA03 SCA06			No
22 (cg-MLST)	1	SCA03 SCA06			No
26 (cg-MLST)	1	SCA03 SCA06			No
27 (cg-MLST)	2	SCA02 SCA03 SCA06			Yes
9 (SNP-reference)	2	SCA02 SCA03 SCA06	SCA01 SCA05 SCA09		No
16 (SNP-reference)	2	SCA08 SCA03 SCA02 SCA06 and SCA07 SCA10 SCA04	SCA01 SCA05 SCA09		No
34 (cg-MLST)	2	SCA02 SCA03 SCA06 SCA08	SCA01 SCA05 SCA09		No
14 (SNP-assembly)	3	SCA08 SCA02 SCA06 SCA03 (SCA07)	SCA01 SCA05 SCA09	SCA10 SCA04	No
28 (cg-MLST)	3	SCA08 SCA02 SCA03 SCA06	SCA01 SCA05 SCA09	SCA07 SCA04 SCA10	No
32 (SNP-reference)	4	SCA08 SCA02 SCA03 SCA06	SC07 SC05 SC01 SC09	SC04 SC10	No

SCA02 SCA03 SCA06 Cluster 1, WGS-based analysis

11 Analysis of Lab11 WGS data shows that most likely strain SCA01 was processed twice: as SCA01 and as SCA02

Note: the order of the clusters as reported by participants have been rearranged, for easier reading and comparison.

4.3.5 *Investigations on the variability in the results of strains used in the cluster analysis*

Interestingly, several strains showed far more variety than expected. This was observed during the PT in November 2019 with all participants (Figures 10 and 11, Annex 11: participants' laboratory codes in numbers), and also after storage of the transport tubes at room temperature containing the strains till February 2020 (Figures 10 and 11, Annex 11: EL and EL2, respectively).
No specific explanations for this variability were found after the quality assessment of the WGS data (e.g., on participants' wet-lab or dry-lab protocols [Annex 9], or on QC data [Tables 13 and 14, Annex 12]).

An SNP (single nucleotide polymorphisms) analysis against a reference genome was conducted to investigate the variability observed among strain SCA03 and its technical duplicate SCA06. In addition to these two diverging strains, the less variable SCA02 and SCA08 were also included for comparison.
A closely related reference was selected from the public DNA databases by screening available complete genomes at the public DNA database (NCBI) with cgMLST analysis, together with the four strains.
S. Typhimurium str. DT104 (Accession number HF937208.1) resulted as the closest match to SCA03/SCA06.
For identification of the SNPs, all available short paired-end reads of the four strains from the PT participants and the EURL-*Salmonella* (EL0, EL, EL2) were mapped to the reference genome using BWA-MEM (Li and Durbin, 2009). Next, the SAM files generated by this aligner software package were sorted and filtered into BAM files using SAMtools (Li et al., 2009). VCF files with high quality SNPs (>90% consensus, minimum depth 15x, GQ > = 20) were created using BCFtools (Li et al., 2009). The SNPs found were visualised (Figure 12) with the integrative genomics viewer IGV (Robinson et al, 2011).

The potential micro-evolution of strains was also wet-lab investigated. A new set of strains was selected to be tested for use in the PT2020 Cluster Analysis. Four strains from the PT2019 set were also included in the pre-testing. Test strains 20SCAT01 - 20SCAT20 were freshly cultured on blood-agar plates and directly submitted for WGS analysis. Strains SCA02, SCA03, SCA07 and SCA10 were cultured from the -70°C stock, prepared from the transport tubes in November 2019. Approximately every other day, all test strains were sub-cultured ten times, using alternately liquid (BPW) and solid (blood agar plates) media, and then resubmitted for WGS analysis (n=15). Results are visualised in an MST in Figure 13.

All strains showed identical WGS results after ten times sub-culturing, except for strains 20SCAT01 and 20SCAT05. The two exceptions are strains SCA03 and SCA10, two variable strains from the PT2019, and these showed a notable difference after the sub-culturing challenge.

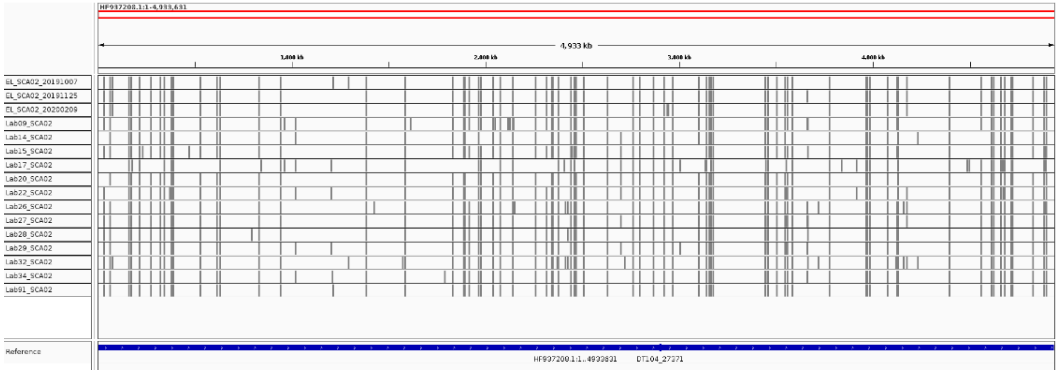
Another wet-lab experiment was conducted in November 2020, one year after the preparation of the original transport tubes for the PT2019 strains. A subset was freshly cultured on blood agar plates (from the transport tubes, kept at room temperature) to yield single colonies. Several single colonies per strain were separately sub-cultured and

submitted for WGS analysis. Results are visualised in an MST in Figure 14.

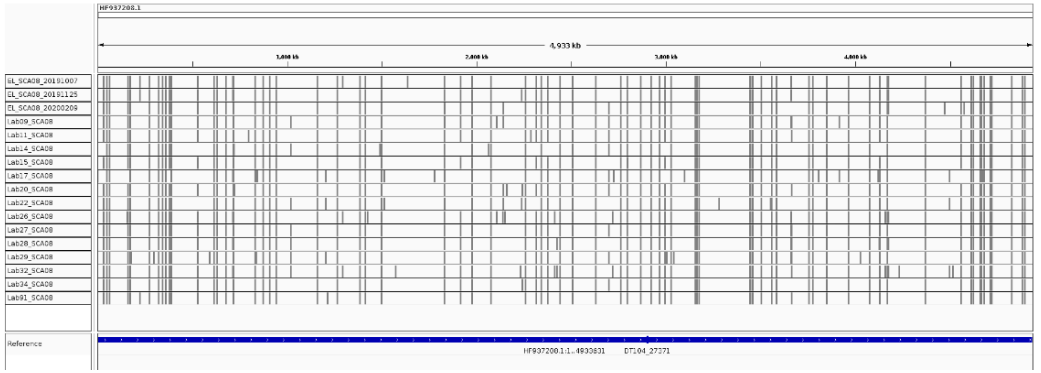
After one year of storage in transport tubes, all strains except SCA03/06, SCA04 and SCA10, still showed comparable WGS results, this was also true for multiple colonies per plate (e.g., see strain SCA02 with six colonies tested).

On the other hand, strains SCA03/06, SCA04 and SCA10 showed variability in time (EL0, EL, EL2, EL3) as well as in 2-6 colonies originating from the same blood agar plate (A-F).

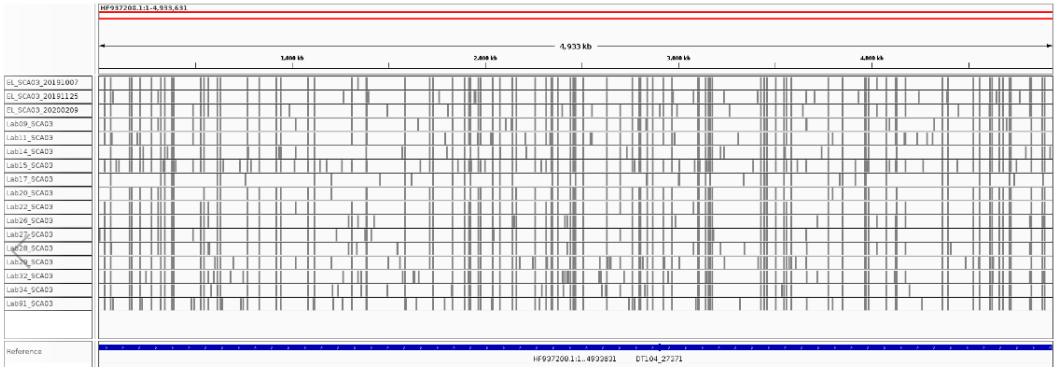
19SCA02



19SCA08



19SCA03



19SCA06

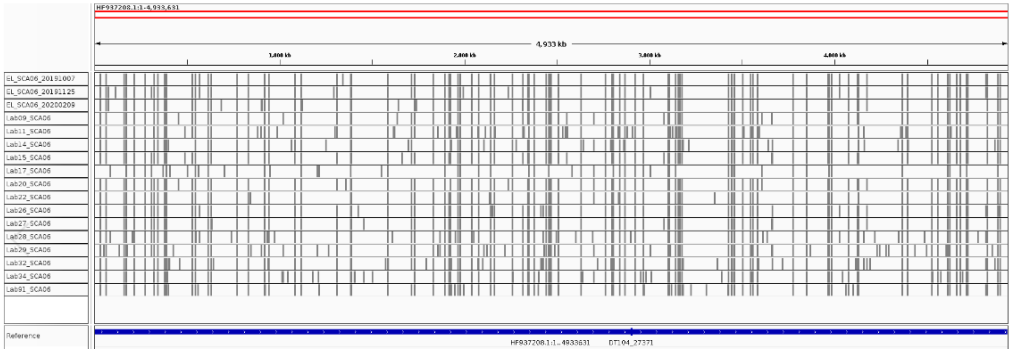


Figure 12 Visualisation of the SNPs analysis results for strains 19SCA02, 19SCA03/06, and 19SCA08

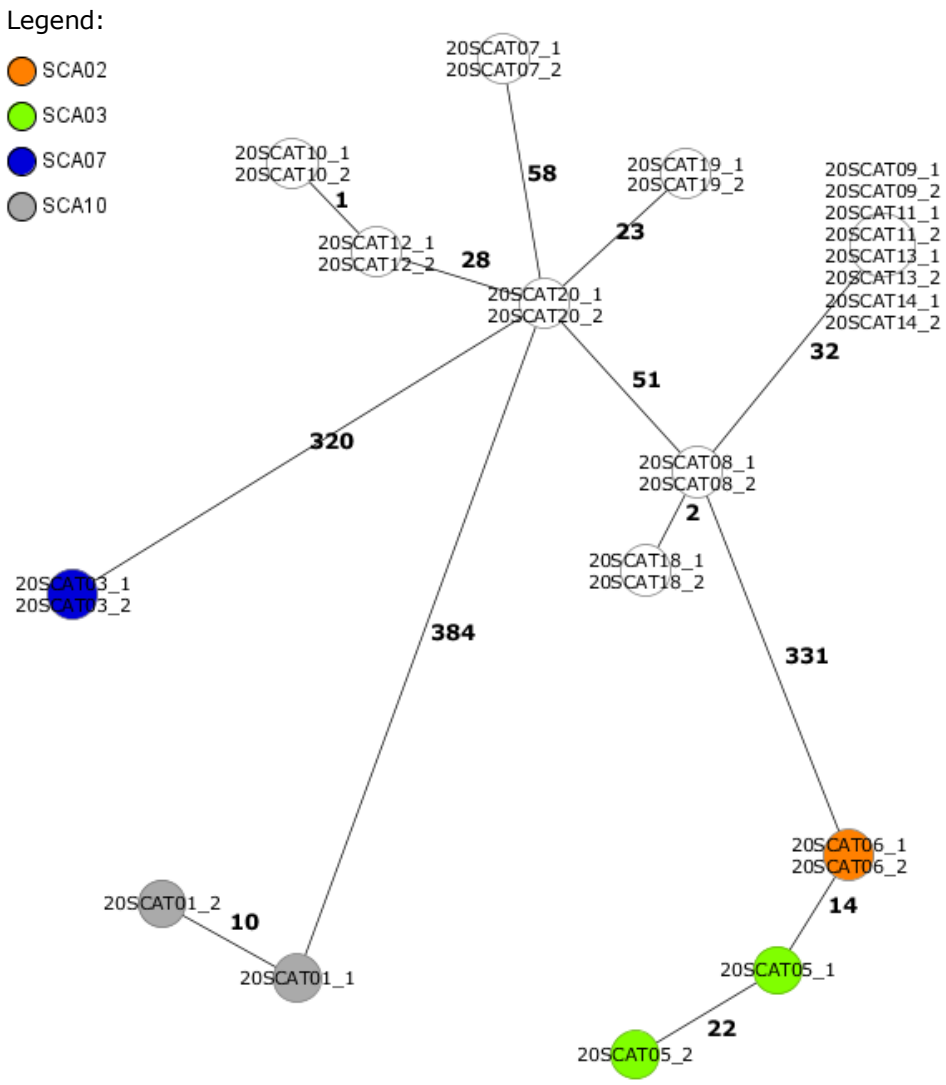


Figure 13 MST of 15 test strains, before (_1) and after (_2) 10 times sub-culturing

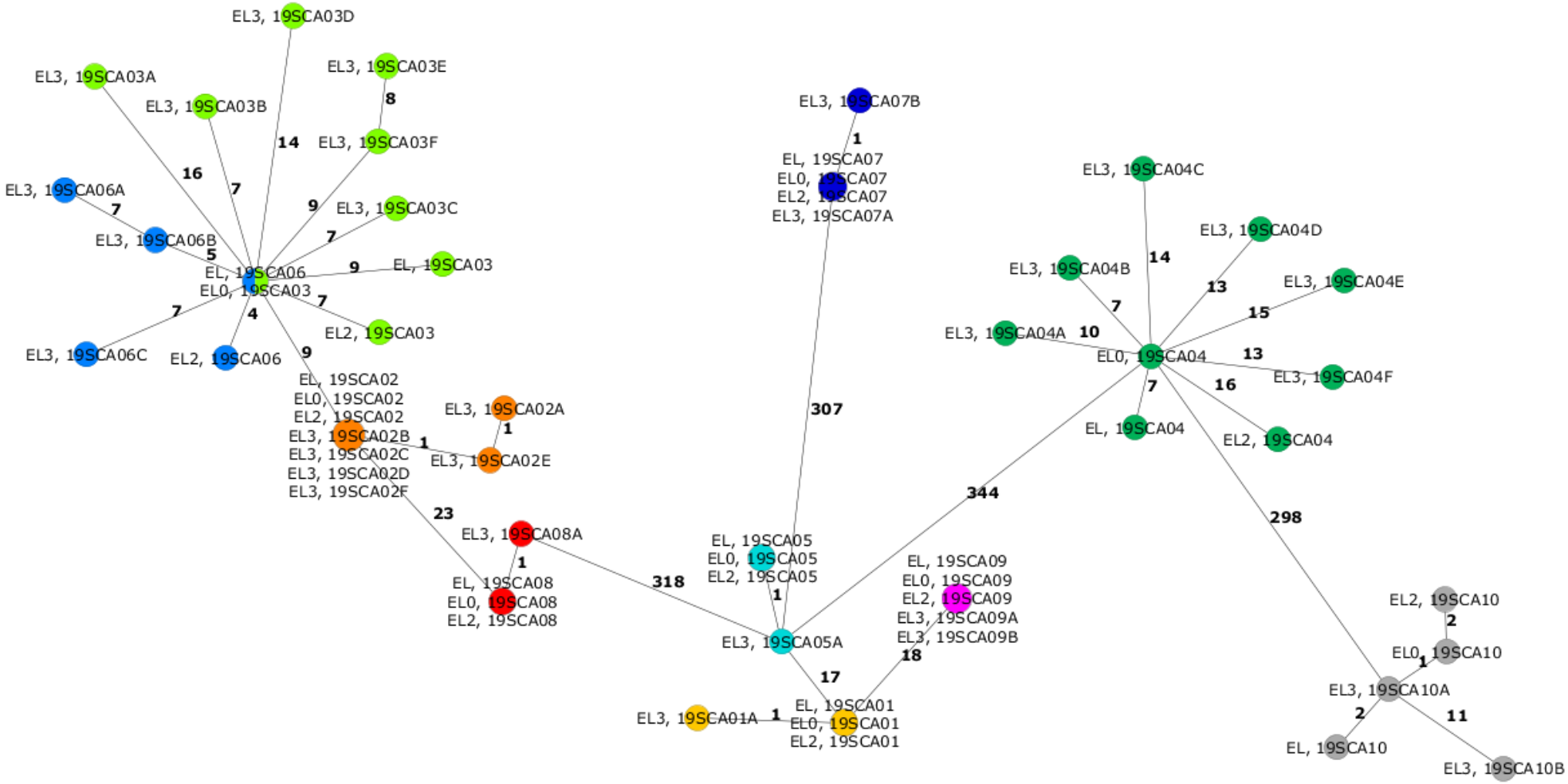


Figure 14 MST of the 10 PT2019 strains, results from the EURL-Salmonella (EL) over time: EL0 (7/10/2019), EL=EL1=PT2019 (25/11/2019), EL2 (10/2/2020) and EL3 (16/11/2020). Different letters (A-F) indicate different single colonies originating from 1 blood agar plate

5 Conclusions

5.1 Serotyping

- Overall results for the 35 evaluated NRLs are:
  - 99% of the strains were typed correctly for the O-antigens.
  - 97% of the strains were typed correctly for the H-antigens.
  - 97% of the strains were correctly named.
- All 29 EU NRLs and four non-EU NRLs achieved the defined level of good performance after the first stage of the study.
- Two non-EU NRLs initially did not achieve the defined level of good performance:
  - The online follow-up on this performed by the first participant was satisfactory.
  - The proposed training session for the second participant had to be put on hold due to the ongoing COVID-19 pandemic.

5.2 Cluster analysis

- A total of eighteen NRLs participated in the cluster analysis pilot, with six participants using PFGE analysis, eight using MLVA and fourteen using WGS.
- Valuable information and data were obtained from this first cluster analysis pilot.
- Participants interpreted *"cluster(s) of closely related isolates"* in a variety of ways, which complicated an appropriate evaluation. No performance criteria were set for this pilot study.
- As a minimum however, it was expected that participants would report the technical duplicate strains SCA03 and SCA06 to be (part of) one cluster. This was found in 5/6 (PGFE), 7/7 (MLVA), and 12/14 (WGS) of the reported cases.
- An unexpected variability on some of the strains was observed, especially in the WGS results. Extended investigations revealed that this was more likely to be due to the biological origin (sub-culturing, long-term storage) than the technical origin (participants' wet-lab/dry-lab protocols, QC data assessments).
- Selection of suitable (stable) PT strains, primarily based on WGS analysis, will be improved by including more pre-testing on the strains.
- The second pilot on (optional) Cluster Analysis in 2020 will be based on the simulation of an outbreak-related request to the NRL-network from the EURL-*Salmonella* (EFSA/ECDC), including a description of the cluster definition.

List of abbreviations

BN	BioNumerics
BPW	Buffered Peptone Water
cgMLST	core genome Multilocus Sequence Typing
DG-SANTE	Directorate General for Health and Food Safety
ECDC	European Centre for Disease prevention and Control
EFSA	European Food Safety Authority
EFTA	European Free Trade Association
EL	EURL-*Salmonella* Laboratory
EU	European Union
EURL-*Salmonella*	European Union Reference Laboratory for *Salmonella*
MLVA	Multiple-Locus Variable number of tandem repeat Analysis
MST	Minimum Spanning Tree
n.a.	not applicable
NCBI	National Center for Biotechnology Information
NRL-*Salmonella*	National Reference Laboratory for *Salmonella*
ISO	International organization for standardization
PCR	Polymerase Chain Reaction
PFGE	Pulsed Field Gel Electrophoresis
PT	Proficiency Test
QC	Quality Control
REF	Reference
RIVM	National Institute for Public Health and the Environment (Bilthoven, The Netherlands)
SNP	Single Nucleotide Polymorphism
SSI	Statens Serum Institut (Copenhagen, Denmark)
ST	Sequence Type
TIFF	Tagged Image File Format
wgMLST	whole genome Multilocus Sequence Typing
WGS	Whole Genome Sequencing

References

EC, 2017. Regulation (EU) 2017/625 of the European Parliament and of the Council of 15 March 2017 on official controls and other official activities performed to ensure the application of food and feed law, rules on animal health and welfare, plant health and plant protection products. Official Journal of the European Union L95: 7 April 2017. http://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32017R0625&rid=3 (accessed 17/9/2021).

ECDC (European Centre for Disease Prevention and Control), 2011. ECDC Technical Report, ECDC 2011-06, version 1.2. Laboratory standard operating procedure for MLVA of *Salmonella enterica* serotype Typhimurium. Available online: https://www.ecdc.europa.eu/sites/default/files/media/en/publications/Publications/1109_SOP_Salmonella_Typhimurium_MLVA.pdf (accessed 17/9/2021).

Grimont, P.A.D. and F.-X. Weill, 2007. Antigenic formulae of the *Salmonella* serovars, 9th ed. WHO Collaborating Centre for Reference and Research on *Salmonella*. Institute Pasteur, Paris, France. https://www.pasteur.fr/sites/default/files/veng_0.pdf (accessed 17/9/2021).

EN ISO 6579-1 Microbiology of the food chain - Horizontal method for the detection, enumeration and serotyping of *Salmonella* - Part 1: Detection of *Salmonella*. International Organization for Standardization, Geneva, Switzerland.

EN ISO 15189 Medical laboratories - Requirements for quality and competence. International Organization for Standardization, Geneva, Switzerland.

EN ISO/IEC 17025 General requirements for the competence of testing and calibration laboratories. International Organization for Standardization, Geneva, Switzerland.

Larsson, J.T., M. Torpdahl, R.F. Petersen, G. Sørensen, B.A. Lindstedt, and E.M. Nielsen, 2009. Development of a new nomenclature for *Salmonella* Typhimurium multilocus variable number of tandem repeats analysis (MLVA). Euro Surveill. 2009;14(15):pii=19174. Available online: http://www.eurosurveillance.org/ViewArticle.aspx?ArticleId=19174 (accessed 17/9/2021).

Li, H., and R. Durbin, 2009. Fast and accurate short read alignment with Burrows–wheeler transform. Bioinformatics 25(14):1754–1760.

Li, H., B. Handsaker, A. Wysoker, T. Fennell, J. Ruan, N. Homer, G. Marth, G. Abecasis, R. Durbin, the 1000 Genome Project Data Processing Subgroup, 2009. The sequence alignment/Map format and SAMtools. Bioinformatics 25(16):2078–2079.

Lindstedt, B.A., T. Vardund, L. Aas, and G. Kapperud, 2004. Multiple-locus variable-number tandem-repeats analysis of *Salmonella enterica* subsp. *enterica* serovar Typhimurium using PCR multiplexing and multicolor capillary electrophoresis. Journal of Microbiological Methods, 59(2),163-72.

Mooijman, K.A., 2007. The twelfth CRL-*Salmonella* Workshop; 7 and 8 May 2007, Bilthoven, the Netherlands. National Institute for Public Health and the Environment, Bilthoven, the Netherlands. RIVM Report no.: 330604006. http://www.eurlsalmonella.eu/Publications/Workshop_Reports (accessed 17/9/2021).

Mooijman, K.A., 2019. The 24th EURL-*Salmonella* workshop; 28-29 May 2019, Amersfoort, The Netherlands. National Institute for Public Health and the Environment, Bilthoven, the Netherlands. RIVM Report no.: 2019-0135. http://www.eurlsalmonella.eu/Publications/Workshop_Reports (accessed 17/9/2021).

Robinson, J.T., H. Thorvaldsdóttir, W. Winckler, M. Guttman, E.S. Lander, G. Getz, and J.P. Mesirov, 2011. Integrative Genomics Viewer. Nature Biotechnology 29(1):24–26.

Tennant, S.M., S. Diallo, H. Levy, S. Livio, S.O. Sow, M. Tapia, P.I. Fields, M. Mikoleit, B. Tamboura, K.L. Kotloff, J.P. Nataro, J.E. Galen and M.M. Levine, 2010. Identification by PCR of non-typhoidal *Salmonella enterica* serovars associated with invasive infections among febrile patients in Mali. PLoS. Negl. Trop. Dis 4(3): 621.

Annex 1 Example of an individual laboratory evaluation report on serotyping results

Results
EURL-*Salmonella* PT Serotyping 2019

Number of penalty points: 0 **Evaluation: Good Performance**

	Reference Results				**Results NRL labcode:**			**1**
Strain	O-antigens	H-antigens (phase 1)	H-antigens (phase 2)	Serovar	O-antigens	H-antigens (phase 1)	H-antigens (phase 2)	Serovar
S1	6,8	z10	e,n,x	Hadar	6,8	z10	e,n,x	Hadar
S2	1,4,[5],12	y	1,2	Coeln	4,5,12	y	1,2	Coeln
S3	1,13,22	z	1,6	Poona	13,22	z	1,6	Poona
S4	9,46	z38	-	Fresno	9,46	z38	-	Fresno
S5	1,9,12	g,m	-	Enteritidis	9,12	g,m	-	Enteritidis
S6	3,{10}{15}{15,34}	e,h	l,w	Meleagridis	3,10	e,h	l,w	Meleagridis
S7	1,4,[5],12,27	d	1,2	Stanley	4,5,12	d	1,2	Stanley
S8	1,4,[5],12	f,g,s	[1,2]	Agona	4,12	f,g,s	-	Agona
S9[a)]	1,4,[5],12	i	1,2	Typhimurium	1,4,5,12	i	1,2	Typhimurium
S10	6,7	r	1,6	Nigeria	6,7	r	1,6	Nigeria
S11	3,{10},{15}	l,z13	1,5	Uganda	3,10	l,z13	1,5	Uganda
S12	11	r	e,n,x	Rubislaw	11	r	e,n,x	Rubislaw
S13	{6,7,14}{54}	g,m,[p],s	[1,2,7]	Montevideo	6,7	g,m,s	-	Montevideo
S14[b)]	1,4,[5],12	i	-	1,4,[5],12:i:-	4,12	i	-	4,12:i:-
S15	6,8	d	1,2	Muenchen	6,8	d	1,2	Muenchen
S16	6,7,14	r	1,2	Virchow	6,7	r	1,2	Virchow
S17	6,7,14	r	1,5	Infantis	6,7	r	1,5	Infantis
S18	16	y	1,5	Saphra	16	y	1,5	Saphra
S19	1,13,23	i	l,w	Kedougou	1,13,23	i	l,w	Kedougou
S20	1,4,[5],12,27	g, s, t	[1,2]	Kingston	1,4,12	g,s,t	-	Kingston
S21[c)]	48	g,z51	-	IV 48:g,z51:-	48	g,z51	-	48:g,z51:-

a) Potentially contaminated with an *E. coli* strain. Results strain S9 were excluded from the evaluation.
b) Typhimurium, monophasic variant as determined by PCR.
c) *Salmonella enterica* subspecies *houtenae*

Results
EURL-*Salmonella* PT Serotyping 2019

For back-ground information, reference results are given completely according to the White-Kauffmann-le Minor scheme (2007).
Participants were asked to report only those results, on which the identification of serovar names was based.

Colour coding:

	remark (e.g. spelling error, or deviations in the results of optional strain S21)
	not typable (e.g. antisera not available, rough strain)
	partly correct; the naming: no penalty points
	incorrect; in the naming: 1 penalty point
	incorrect; in the naming: 4 penalty points

As decided at the 24th EURL-*Salmonella* Workshop (Amersfoort, 2019), Strain S-21 was an additional strain to the study. Testing of this strain was optional and results were not included in the evaluation (remarks in blue or grey only).
The evaluation of the serotyping results was performed as indicated in Table 1 of the Protocol as sent to the participants. In addition to that, Good Performance was evaluated on the basis of penalty points as indicated below.

4 penalty points: Incorrect typing of *S*. Enteritidis, *S*. Typhimurium (including monophasic variant), *S*. Hadar, *S*. Infantis or *S*. Virchow or assigning the name of one of these 5 serovars to another serovar.
1 penalty point: Incorrect typing of all other *Salmonella* serovars.
(no penalty points are given in case a strain was non-typable due to auto-agglutination)

Good Performance is defined as < 4 penalty points.

Annex 2 Serotyping results per strain and per laboratory

Lab:	S1	S2	S3	S4	S5	S6	S7	S8	S9	S10	S11
REF	**Hadar**	**Coeln**	**Poona**	**Fresno**	**Enteritidis**	**Meleagridis**	**Stanley**	**Agona**	**Typhimurium**	**Nigeria**	**Uganda**
1	Hadar	Coeln	Poona	Fresno	Enteritidis	Meleagridis	Stanley	Agona	Typhimurium	Nigeria	Uganda
2	Hadar	Coeln	Poona	Elomrane	Enteritidis	Meleagridis	Stanley	Agona	Typhimurium	Nigeria	Uganda
3	Hadar	Coeln	Poona	Fresno	Enteritidis	Meleagridis	Stanley	Agona	Typhimurium	Nigeria	Uganda
4	Hadar	Typhimurium	Poona	9,46:-:-	Enteriditis	Newlands	Typhimurium	Binche	Typhimurium	Virchow	Uganda
5	Hadar	Coeln	Poona	Fresno	Enteritidis	Meleagridis	Stanley	Agona	Typhimurium	Nigeria	Ouganda
6	Hadar	Coeln	Poona	Fresno	Enteritidis	Meleagridis	Stanley	Agona	Typhimurium	Nigeria	Uganda
7	S. Hadar	S. Coeln	S. Poona	S. Fresno	S. Enteritidis	S. Meleagridis	S. Stanley	S. Agona	S. Typhimurium	S. Nigeria	S. Uganda
8	S. Hadar	S. Coeln	S. Poona	9,46 : ? :-	S. Enteritidis	S. Meleagridis	S. Stanley	S. Agona	S. Typhimurium	S. Nigeria	S. Uganda
9	Hadar	Coeln	Poona	Fresno	Enteritidis	Meleagridis	Stanley	Agona	-:-:-	Nigeria	Uganda
10	Hadar	Coeln	Poona	Fresno	Enteritidis	Meleagridis	Stanley	Agona	Typhimurium	Nigeria	Uganda
11	Hadar	Coeln	Poona	9:-:-	Enteritidis	Meleagridis	Stanley	Agona	Typhimurium	Nigeria	Uganda
12	Hadar	Coeln	Poona	Fresno	Enteritidis	Meleagridis	Stanley	Agona	Typhimurium	Nigeria	Uganda
13	Hadar	Coeln	Poona	Fresno	Enteritidis	Meleagridis	Stanley	Agona	Typhimurium	Nigeria	Uganda
14	Hadar	Coeln	Poona	Fresno	Enteritidis	Meleagridis	Stanley	Agona	Typhimurium	Nigeria	Uganda
15	Hadar	Coeln	Poona	Fresno	Enteritidis	Meleagridis	Stanley	Agona	Typhimurium	Nigeria	Uganda
16	Hadar	Coeln	Poona	Fresno	Enteritidis	Meleagridis	Stanley	Agona	Typhimurium	Nigeria	Uganda
17	Hadar	Coeln	Poona	Frenso	Enteritidis	Meleagridis	Stanley	Agona	Typhimurium	Nigeria	Uganda
18	Hadar	Coeln	Poona	Fresno	Enteritidis	Meleagridis	Stanley	Agona	Typhimurium	Nigeria	London
19	Hadar	Coeln	Poona	9,46:?	Enteritidis	Meleagridis	Stanley	Agona	Typhimurium	Nigeria	Uganda
20	Hadar	Coeln	Poona	Fresno	Enteritidis	Meleagridis	Stanley	Agona	Typhimurium	Nigeria	Uganda
21	Hadar	Coeln	Poona	Fresno	Enteritidis	Meleagridis	Stanley	Agona	Typhimurium	Nigeria	Uganda
22	Hadar	Coeln	Poona	Fresno	Enteritidis	Meleagridis	Stanley	Agona	Typhimurium	Nigeria	Uganda
23	Hadar	Coeln	Poona	9,46:-:-	Enteritidis	Meleagridis	Stanley	Agona	Typhimurium	Nigeria	Uganda
24	Hadar	Coeln	Poona	Fresno	Enteritidis	Meleagridis	Stanley	Agona	Typhimurium	Nigeria	Uganda
25	Hadar	Coeln	Poona	Fresno	Enteritidis	Meleagridis	Stanley	Agona	Typhimurium	Nigeria	Uganda
26	Hadar	Coeln	Poona	Fresno	Enteritidis	Meleagridis	Stanley	Agona	Escherichia coli	Nigeria	Uganda
27	Hadar	Coeln	Poona	Fresno	Enteriditis	Meleagridis	Stanley	Agona	Blockley	Nigeria	Uganda
28	Hadar	Coeln	Poona	Fresno	Enteritidis	Meleagridis	Stanley	Agona	Typhimurium	Nigeria	Uganda

Lab:	S1	S2	S3	S4	S5	S6	S7	S8	S9	S10	S11
REF	**Hadar**	**Coeln**	**Poona**	**Fresno**	**Enteritidis**	**Meleagridis**	**Stanley**	**Agona**	**Typhimurium**	**Nigeria**	**Uganda**
29	Hadar	Coeln	Poona	Fresno	Enteritidis	Meleagridis	Stanley	Agona	Typhimurium	Nigeria	Uganda
30	Hadar	Coeln	Poona	Fresno	Enteritidis	Meleagridis	Stanley	Agona	Typhimurium	Nigeria	Uganda
31	Hadar	Coeln	Poona	Fresno	Enteritidis	Meleagridis	Stanley	Agona	Typhimurium	Nigeria	Uganda
32	Hadar	Coeln	Poona	Fresno	Enteritidis	Meleagridis	Stanley	Agona	Typhimurium	Nigeria	Uganda
33	Hadar	Coeln	Poona	Fresno	Enteritidis	Meleagridis	Stanley	Agona	Typhimurium	Nigeria	Sinstorf
34	Hadar	Coeln	Poona	Fresno	Enteritidis	Meleagridis	Stanley	Agona	Typhimurium	Nigeria	Uganda
35	Bonariensis	Coeln	Poona	Fresno	Enteritidis	Meleagridis	Stanley	Agona	Typhimurium	Nigeria	Uganda
X	1	1	0	1	0	1	1	1	n.d.	1	2

S12	S13	S14	S15	S16	S17	S18	S19	S20	Lab:
Rubislaw	**Montevideo**	**1,4,[5],12:i:-**	**Muenchen**	**Virchow**	**Infantis**	**Saphra**	**Kedougou**	**Kingston**	**REF**
Rubislaw	Montevideo	4,12:i:-	Muenchen	Virchow	Infantis	Saphra	Kedougou	Kingston	1
Rubislaw	Montevideo	4,12:i:-	Muenchen	Virchow	Infantis	Saphra	Kedougou	Kingston	2
Rubislaw	Montevideo	4,5,12:i:-	Muenchen	Virchow	Infantis	Saphra	Kedougou	Kingston	3
Mountmagnet	Montevideo	Typhimurium	Chennai	Virchow	Infantis	Saphra	Idikan	Kingston	4
Rubislaw	Montevideo	Typhimurium monophasic variant	Muenchen	Virchow	Infantis	Saphra	Kedougou	Kingston	5
Rubislaw	Montevideo	4,12:i:-	Muenchen	Virchow	Infantis	Saphra	Kedougou	Kingston	6
S. Rubislaw	S. Montevideo	4 : i : -	S. Muenchen	S. Virchow	S. Infantis	S. Saphra	S. Kedougou	S. Kingston	7
S. Rubislaw	S. Montevideo	4,5,12 : i : -	S. Muenchen	S. Virchow	S. Infantis	S. Saphra	S. Kedougou	S. Kingston	8
Rubislaw	Montevideo	4,12:i:-	Muenchen	Virchow	Infantis	Saphra	Kedougou	Kingston	9
Rubislaw	Montevideo	4,12:i:-	Muenchen	Virchow	Infantis	Saphra	Kedougou	Kingston	10
Rubislaw	Montevideo	Monophasic Typhimurium 4:i:-	Muenchen	Virchow	Infantis	Saphra	Kedougou	Kingston	11
Rubislaw	Montevideo	1,4,5,12:i:-	Muenchen	Virchow	Infantis	Saphra	Kedougou	Kingston	12
Rubislaw	Montevideo	4,5,12:i:-	Muenchen	Virchow	Infantis	Saphra	Kedougou	Kingston	13
Rubislaw	Montevideo	4:i:-	Muenchen	Virchow	Infantis	Saphra	Kedougou	Kingston	14
Rubislaw	Montevideo	4,12:i:-	Muenchen	Virchow	Infantis	Saphra	Kedougou	Kingston	15
Rubislaw	Montevideo	4,12:i:- (monophasic Salmonella Typhimurium)	Muenchen	Virchow	Infantis	Saphra	Kedougou	Kingston	16
Rubislaw	Montevideo	4,5,12;i;-	Muenchen	Virchow	Infantis	Saphra	Kedougou	Kingston	17
Rubislaw	Montevideo	Monophasic Salmonella Typhimurium	Manhattan	Virchow	Infantis	Saphra	Kedougou	Kingston	18
Rubislaw	Montevideo	4,12:i:-	Muenchen	Virchow	Infantis	Saphra	Kedougou	Kingston	19
Rubislaw	Montevideo	4:i:-, Monofasisk Typhimurium	Muenchen	Virchow	Infantis	Saphra	Kedougou	Kingston	20
Rubislaw	Montevideo	monophasic Typhimurium	Muenchen	Virchow	Infantis	Saphra	Kedougou	Kingston	21
Rubislaw	Montevideo	4,5,12:i:-	Muenchen	Virchow	Infantis	Saphra	Kedougou	Kingston	22
Rubislaw	Montevideo	S.Typhimurium monophasic	Muenchen	Virchow	Infantis	Saphra	Kedougou	Kingston	23
Rubislaw	Montevideo	4, 5, 12:i:-	Muenchen	Virchow	Infantis	Saphra	Kedougou	Kingston	24
Rubislaw	Montevideo	1,4,5,12:i:-	Muenchen	Virchow	Infantis	Saphra	Kedougou	Kingston	25
Rubislaw	Montevideo	SI, 4,12:i:-	Muenchen	Virchow	Infantis	Saphra	Kedougou	Kingston	26
Rubislaw	Montevideo	4,12:i:-	Muenchen	Virchow	Infantis	Saphra	Kedougou	Kingston	27
Rubislaw	Montevideo	4,12:i:-	Muenchen	Virchow	Infantis	Saphra	Kedougou	Kingston	28
Rubislaw	Montevideo	4:i:-	Muenchen	Virchow	Infantis	Saphra	Kedougou	Kingston	29
Rubislaw	Montevideo	4,12:i:-	Muenchen	Virchow	Infantis	Saphra	Kedougou	Kingston	30

S12	S13	S14	S15	S16	S17	S18	S19	S20	Lab:
Rubislaw	**Montevideo**	**1,4,[5],12:i:-**	**Muenchen**	**Virchow**	**Infantis**	**Saphra**	**Kedougou**	**Kingston**	**REF**
Rubislaw	Montevideo	4,12:i:-	Muenchen	Virchow	Infantis	Saphra	Kedougou	Kingston	31
Rubislaw	Montevideo	4,12:i:-	Muenchen	Virchow	Infantis	Saphra	Kedougou	Kingston	32
Rubislaw	Montevideo	4,12:i:-	Muenchen	Virchow	Infantis	Saphra	Kedougou	Kingston	33
Rubislaw	Montevideo	4,12:i:-	Muenchen	Virchow	Infantis	Saphra	Kedougou	Kingston	34
Rubislaw	Montevideo	Typhimurium, monophasic (4,12:i:-)	Muenchen	Virchow	Infantis	Saphra	Kedougou	Kingston	35
1	0	1	2	0	0	0	1	0	**X**

remark (e.g., spelling error)
not typable (e.g., antisera not available, rough strain)
partly correct, in the naming: no penalty points
incorrect; in the naming: 1 penalty point
incorrect; in the naming: 4 penalty points

X = number of deviating laboratories (by penalty points) per strain

Results for Strain S21 are given in Annex 4

Annex 3 Details of strains that caused problems in serotyping

Strain code	O-antigens	H-antigens (phase 1)	H-antigens (phase 2)	Serovar	Lab code
S-1	6,8	z10	e,n,x	Hadar	REF
S-1	6,8	i	e,n,x	Bonariensis	35
S-2	1,4,[5],12	y	1,2	Coeln	REF
S-2	4,5,12	i	2	Typhimurium	4
S-3	1,13,22	z	1,6	Poona	REF
S-3	13,22	z	1,2	Poona	11
S-4	9,46	z38	-	Fresno	REF
S-4	9	z38	-	Elomrane	2
S-4	9,46	-	-	9,46:-:-	4
S-4	9,46	?	-	9,46 : ? :-	8
S-4	9	-	-	9:-:-	11
S-4	9,46	?	?	9,46:?	19
S-4	9,46	-	-	9,46:-:-	23
S-6	3,{10}{15}{15,34}	e,h	l,w	Meleagridis	REF
S-6	3,10	e,h	e,n,x	Newlands	4
S-7	1,4,[5],12,[27]	d	1,2	Stanley	REF
S-7	4,5,12	i	2	Typhimurium	4
S-8	1,4,[5],12	f,g,s	[1,2]	Agona	REF
S-8	47	z4,z23	l,w	Binche	4
S-10	6,7	r	1,6	Nigeria	REF
S-10	6,7,14	r	2	Virchow	4
S-11	3,{10}{15}	l,z13	1,5	Uganda	REF
S-11	3,10	l,z13	1,5	Ouganda	5
S-11	3,10	l,v	1,6	London	18
S-11	3,10	l,v	1,5	Sinstorf	33
S-12	11	r	e,n,x	Rubislaw	REF
S-12	21	r	-	Mountmagnet	4
S-14	1,4,[5],12	i	-	1,4,[5],12:i:-	REF
S-14	4,5,12	i	2	Typhimurium	4
S-15	6,8	d	1,2	Muenchen	REF
S-15	4,12	d	z35	Chennai	4
S-15	6,8	d	1,5	Manhattan	18
S-19	1,13,23	i	l,w	Kedougou	REF
S-19	1,13,23	i	5	Idikan	4

reference strain
remark (e.g., spelling error)
not typable (e.g., antisera not available, rough strain)
partly correct; in the naming: no penalty points
incorrect; in the naming: 1 penalty point
incorrect; in the naming: 4 penalty points

Annex 4 Details of serotyping results for strain S21

Strain code	O-antigens	H-antigens (phase 1)	H-antigens (phase 2)	Serovar	Lab code
S-21	**48**	**g,z51**	**-**	**IV 48:g,z51:-**	**REF**
S-21	48	g,z51	-	48:g,z51:-	1
S-21	48	g,z51	-	48:g,z51:-	2
S-21	48	g,z51	-	IV 48:g,z51:-	3
S-21	9,46	g,m,s	-	Macclesfield	4
S-21	48	g,z51	-	48 : g,z51 : - (subsp. houtenae)	5
S-21	48	g,z51	-	Salmonella enterica subsp. houtenae, 48:g,z51:-	6
S-21	48	g,z51	-	Salmonella enterica subspecies arizonae	7
S-21					8
S-21	48	g,z51	-	48:g,z51:-	9
S-21	48	g	-	Houtenae	10
S-21	48	g,z51	-	IV:48:g,z51:-	11
S-21	48	g,z51	-	48:g,z51:-	12
S-21	48	g,z51	-	IV houtenae	13
S-21	48	g,z51	-	IV 48:g,z51:-	14
S-21	48	g,z51	-	48:g,z51:-	15
S-21	48	g,z51	-	S. enterica subsp. houtenae 48:g,z51:-	16
S-21	48	g,z51	-	48;g,z51;-	17
S-21					18
S-21					19
S-21	48	g	-	Subspec IV, Antigenicformula=48:g:-***	20
S-21	48	g	z51	48 : g : z51 (IV)	21
S-21	48	g,z51	-	Salmonella enterica subsp. houtenae 48:g,z51:-	22
S-21	-	-	-	-:-:-	23
S-21					24
S-21	48	g,z51	-	48:g,z51:-	25
S-21	48	g, z51	-	SIV, 48:g,z51:-	26
S-21	48	g,z51	-	48:g,z51:-	27
S-21	48	g,z51	-	S.IV 48:g,z51:-	28
S-21				?	29
S-21	48	g,z51	-	48:g,z51:-	30
S-21	48	g,z51	-	48:g,z51:-	31
S-21	48	g,Z51	-	sg II 48:g,Z51:-	32
S-21	-	g	-	-:g:-	33
S-21	48	g,z51	-	S. IV. 48:g,z51:-	34
S-21	48	g,z51	-	Salmonella enterica subsp. houtenae serovar 48 : gz51 : -	35

reference result
remark (deviations in the results on optional strain S21)
not typable (antisera not available)

Annex 5 Example of an individual laboratory evaluation report on cluster analysis results

Results
EURL-*Salmonella* PT Cluster Analysis 2019

Laboratory code: 34

Evaluation (per methodology) of the participants' cluster analysis results was done on the ability to correctly identify cluster(s) of genetically closely related isolates, as pre-defined by the EURL-*Salmonella*, and referred to as "expected results".
However, cluster definitions may vary depending on the situation or the specific research question, e.g. in outbreak investigations or surveillance.
The participants to this pilot PT were free to use their own interpretation of "cluster(s) of closely related isolates".
Therefore, no performance criteria were set for this pilot PT on cluster analysis.

In general, deviations (of any kind) from the expected (REF) results are indicated in blue:

Background details and overall results can be found in the Interim Summary Report EURL-*Salmonella* PT Cluster Analysis 2019 (www.eurlsalmonella.eu)

Did you serotype the strains: Yes

Strain:	**SCA01**	**SCA02**	**SCA03**	**SCA04**	**SCA05**
Reported results:	monophasic S. Typhimurium	S. Subspec. I Gruppe B	S. Typhimurium	S. Typhimurium	monophasic S. Typhimurium
Expected results:	4,5,12:i:-	Typhimurium	Typhimurium	Typhimurium	4,5,12:i:-

Strain:	**SCA06**	**SCA07**	**SCA08**	**SCA09**	**SCA10**
Reported results:	S. Typhimurium	S. Typhimurium	S. Typhimurium	monophasic S. Typhimurium	S. Typhimurium
Expected results:	Typhimurium	Typhimurium	Typhimurium	4,5,12:i:-	Typhimurium

Submitted PFGE results: Yes

Number of reported clusters detected by PFGE data analysis:	3
Expected result:	2

	PFGE Cluster 1	PFGE Cluster 2	PFGE Cluster 3
Reported IDs for the strains per cluster:	SCA01, SCA05, SCA09	SCA03, SCA06	SCA02, SCA08
Expected result:		SCA03, SCA06	SCA02, SCA08

PFGE-based cluster identification as expected: No
Technical duplicates SCA03 and SCA06 reported within one cluster: Yes

Submitted MLVA results: Yes

Strain:	SCA01	SCA02	SCA03	SCA04	SCA05
Reported results:	03-13-09-NA-0211	03-16-17-18-0311	03-16-07-17-0311	02-20-08-11-0212	03-11-09-NA-0211
Expected results:	3-13-9-NA-211	3-16-17-18-311	3-16-7-17-311	2-20-8-11-212	3-11-9-NA-211

Strain:	SCA06	SCA07	SCA08	SCA09	SCA10
Reported results:	01-16-07-17-0311	05-09-14-09-0211	03-14-17-25-0311	03-13-09-NA-0211	02-12-07-09-0212
Expected results:	3-16-7-17-311	5-9-14-9-211	3-14-17-25-311	3-13-9-NA-211	2-12-7-9-212

Number of reported clusters detected by MLVA data analysis:	3
Expected result:	2

	MLVA Cluster 1	MLVA Cluster 2	MLVA Cluster 3
Reported IDs for the strains per cluster:	SCA01, SCA05, SCA09	SCA03, SCA06	SCA02, SCA08
Expected result:	SCA01, SCA09	SCA03, SCA06	

MLVA-based cluster identification as expected: No
Technical duplicates SCA03 and SCA06 reported within one cluster: Yes

Submitted WGS results:	Yes
WGS platform used:	Illumina Mi-Seq
Analysis used for WGS data:	cgMLST-based
Tool used for analysis:	in-house automated CHEWBACCA based Pipeline
Method used or phylogenetic analysis:	single linkage hierarchical clustering

Number of reported clusters detected by WGS data analysis:	2
Expected result:	1

	WGS Cluster 1	**WGS Cluster 2**
Reported IDs for the strains per cluster:	SCA01, SCA05, SCA09	SCA02, SCA03, SCA06, SCA08
Expected result:		SCA02, SCA03, SCA06

WGS-based cluster identification as expected: No
Technical duplicates SCA03 and SCA06 reported within one cluster: Yes

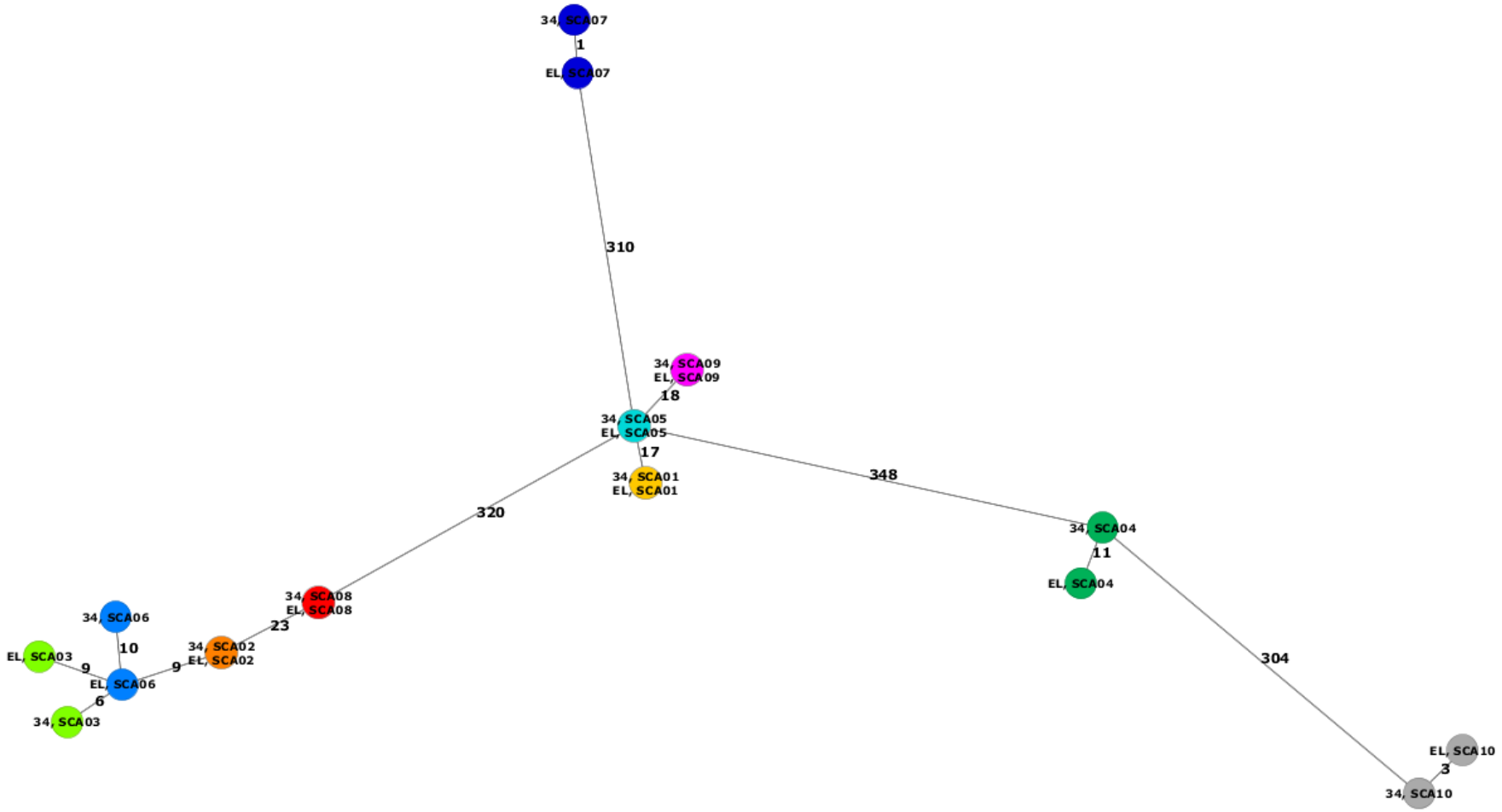


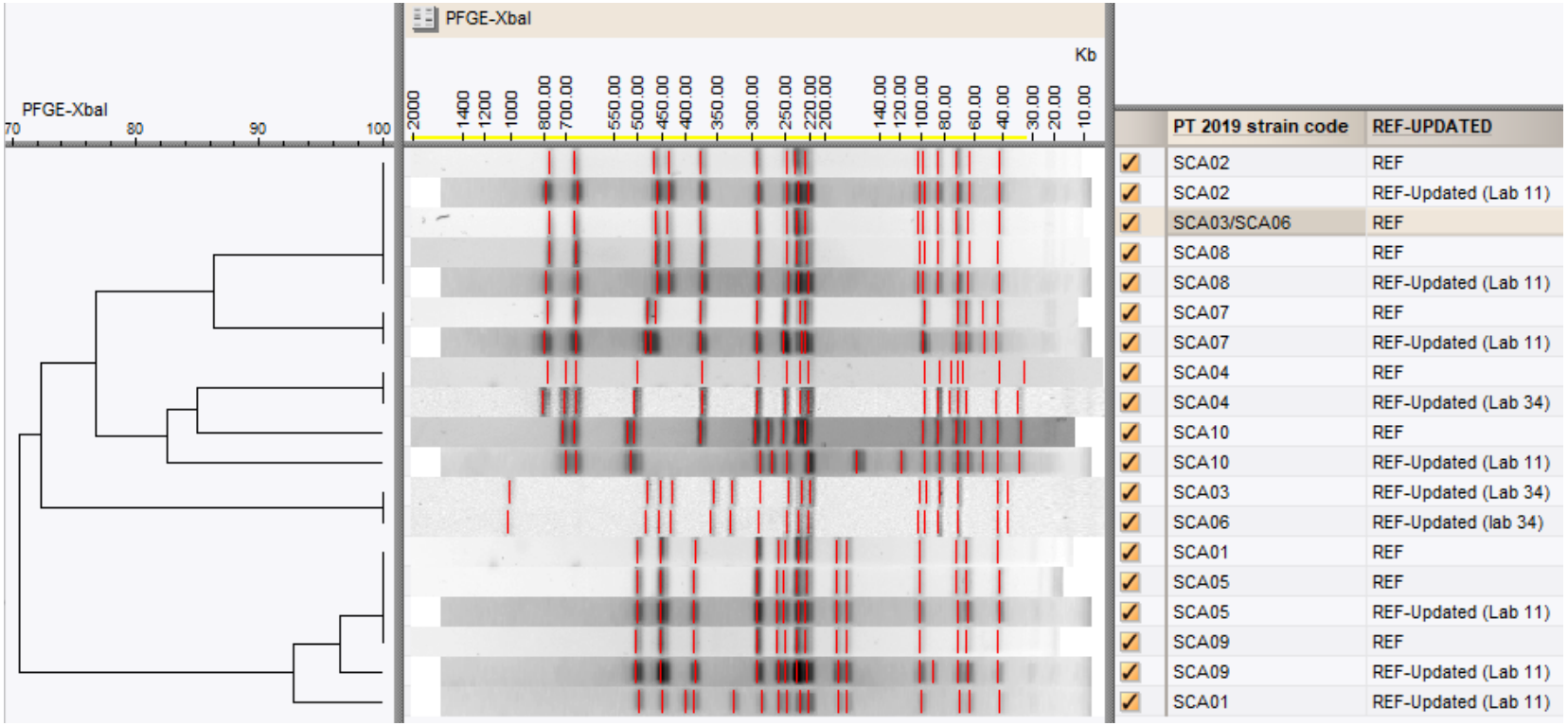
Figure A.5 Minimum Spanning Tree of the participants' results and the EURL-Salmonella (EL) results, analysed in Ridom SeqSphere+, (assembly_pipeline: https://github.com/Papos92), S. enterica cgMLST (3002), pairwise ignoring missing values.

Annex 6 Serotyping results cluster analysis part

Lab code	SCA01	SCA02	SCA03	SCA04	SCA05	SCA06	SCA07	SCA08	SCA09	SCA10
REF	4,5,12:i:-	Typhimurium	Typhimurium	Typhimurium	4,5,12:i:-	Typhimurium	Typhimurium	Typhimurium	4,5,12:i:-	Typhimurium
9	potential monophasic variant of Typhimurium (4:i:-)	Typhimurium (4:i:1,2)	Typhimurium (4:i:1,2)	Typhimurium (4:i:1,2)	potential monophasic variant of Typhimurium (4:i:-)	potential monophasic variant of Typhimurium (4:i:-)	potential monophasic variant of Typhimurium (4:i:-)	potential monophasic variant of Typhimurium (4:i:-)	potential monophasic variant of Typhimurium(O5-)* (4:i:-)	Typhimurium (4:i:1,2)
11	Monophasic Typhimurium	Typhimurium	Typhimurium	Typhimurium	Monophasic Typhimurium	Typhimurium	Typhimurium	Typhimurium	Monophasic Typhimurium	Typhimurium
13	Salmonella Typhimurium	Salmonella Typhimurium	Salmonella Typhimurium	Salmonella Typhimurium	Salmonella Typhimurium	Salmonella Typhimurium	Salmonella Typhimurium monophasic variant	Salmonella Typhimurium	Salmonella Typhimurium	Salmonella Typhimurium
14	4,5:i:-	Salmonella Typhimurium	Salmonella Typhimurium	Salmonella Typhimurium	4,5:i:-	Salmonella Typhimurium	Salmonella Typhimurium	Salmonella Typhimurium	4,5:i:-	Salmonella Typhimurium
16[a)]	I 4,[5],12:i:-, ST34	Typhimurium, ST19	Typhimurium, ST19	Typhimurium, ST19	I 4,[5],12:i:-, ST34	Typhimurium, ST19	Typhimurium, ST19	Typhimurium, ST19	I 4,[5],12:i:-, ST34	Typhimurium, ST19
20	4:i:- monophasic Typhimurium	4:i:1,2 Typhimurium	4:i:1,2 Typhimurium	4:i:1,2 Typhimurium	4:i:- monophasic Typhimurium	4:i:1,2 Typhimurium	4:i:1,2 Typhimurium	4:i:1,2 Typhimurium	4:i:- monophasic Typhimurium	4:i:1,2 Typhimurium
21	monophasic Typhimurium	Typhimurium	Typhimurium	Typhimurium	monophasic Typhimurium	Typhimurium	Typhimurium	Typhimurium	monophasic Typhimurium	Typhimurium
22	4,5,12:i:-	Typhimurium	Typhimurium	Typhimurium	4,5,12:i:-	Typhimurium	Typhimurium	Typhimurium	4,5,12:i:-	Typhimurium
26	1,4,5,12:i:-	Typhimurium	Typhimurium	Typhimurium	1,4,5,12:i:-	Typhimurium	Typhimurium	Typhimurium	4,12:i:-	Typhimurium
29	4,5:i:-, monophasic STM by PCR	Typhimurium	Typhimurium	Typhimurium	4,5:i:-, monophasic STM by PCR	Typhimurium	Typhimurium	Typhimurium	4:i:-, monophasic STM by PCR	Typhimurium
32[b)]	Monophasic variant of Typhimurium	Typhimurium	Typhimurium	Typhimurium	Monophasic variant of Typhimurium	Typhimurium	Typhimurium	Typhimurium	Monophasic variant of Typhimurium	Typhimurium
34	monophasic S. Typhimurium	S. Subspec. I Gruppe B	S. Typhimurium	S. Typhimurium	monophasic S. Typhimurium	S. Typhimurium	S. Typhimurium	S. Typhimurium	monophasic S. Typhimurium	S. Typhimurium

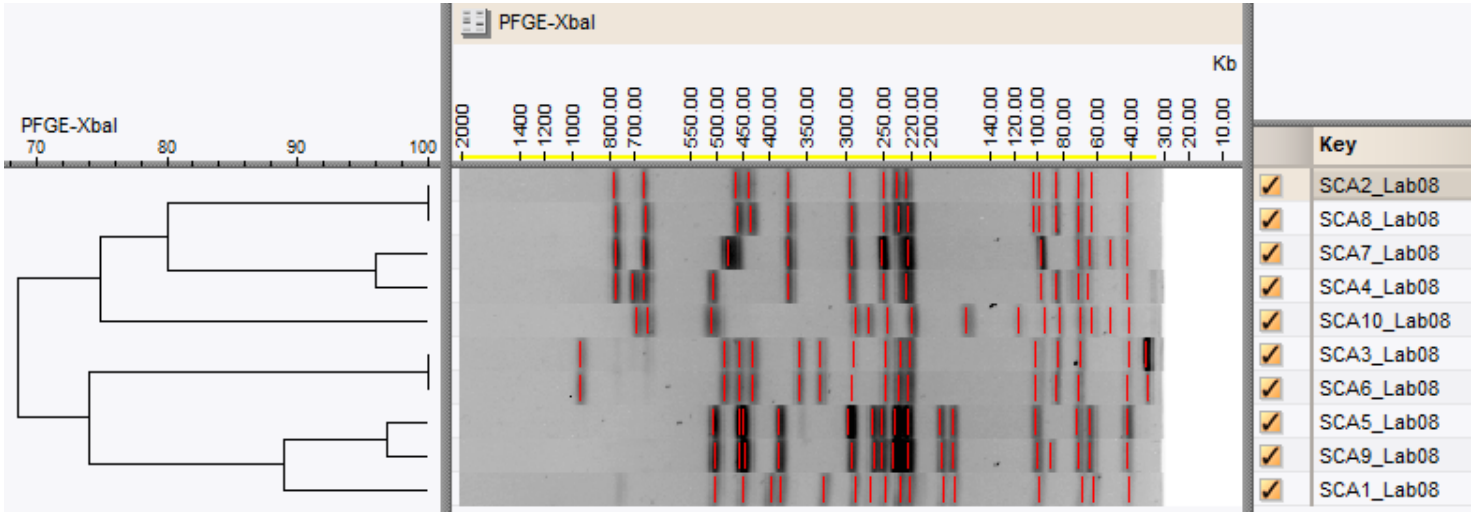
16[a)] The isolates weren't serotyped but a sequences were analysed using SISTR for serovar prediction.
32[b)] For this part of the study we performed WGS based serotyping using in-house Salmonella bioinformatics pipeline
Deviation from the expected result

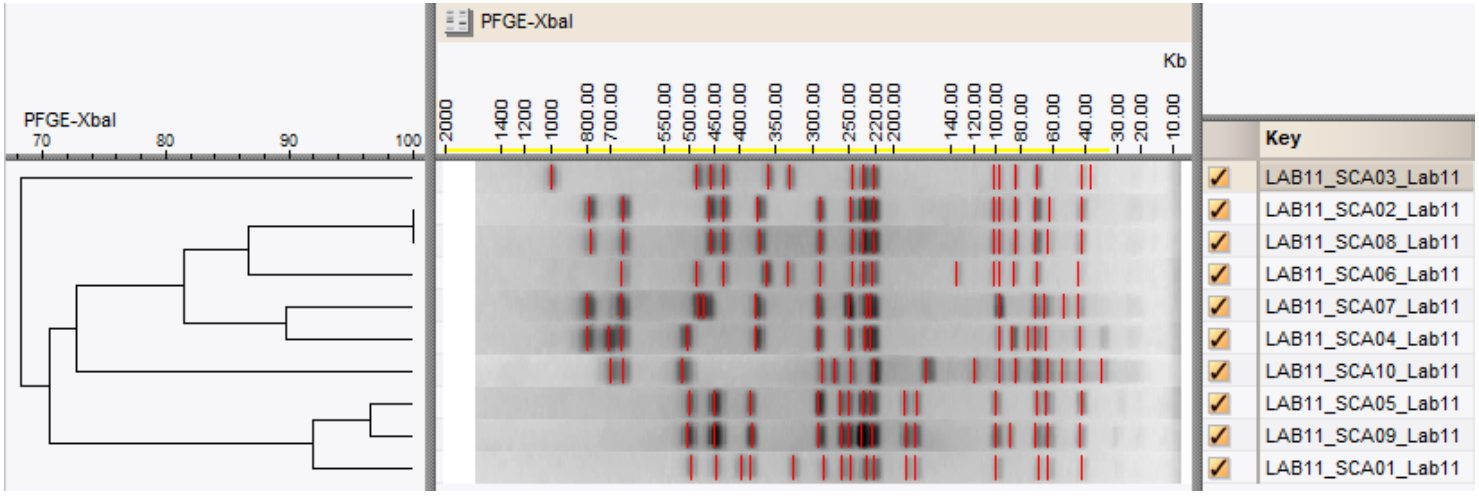
Annex 7 PFGE results cluster analysis part



	PT 2019 strain code	REF-UPDATED
✓	SCA02	REF
✓	SCA02	REF-Updated (Lab 11)
✓	SCA03/SCA06	REF
✓	SCA08	REF
✓	SCA08	REF-Updated (Lab 11)
✓	SCA07	REF
✓	SCA07	REF-Updated (Lab 11)
✓	SCA04	REF
✓	SCA04	REF-Updated (Lab 34)
✓	SCA10	REF
✓	SCA10	REF-Updated (Lab 11)
✓	SCA03	REF-Updated (Lab 34)
✓	SCA06	REF-Updated (lab 34)
✓	SCA01	REF
✓	SCA05	REF
✓	SCA05	REF-Updated (Lab 11)
✓	SCA09	REF
✓	SCA09	REF-Updated (Lab 11)
✓	SCA01	REF-Updated (Lab 11)

Figure A7.1 Original (REF) and PT2019 (REF-Updated) PFGE results of the set of strains selected for cluster analysis


PFGE-XbaI
Kb
2000
1400
1200
1000
800.00
700.00
550.00
500.00
450.00
400.00
350.00
300.00
250.00
220.00
200.00
140.00
120.00
100.00
80.00
60.00
40.00
30.00
20.00
10.00
PFGE-XbaI
70
80
90
100
Key
SCA2_Lab08
SCA8_Lab08
SCA7_Lab08
SCA4_Lab08
SCA10_Lab08
SCA3_Lab08
SCA6_Lab08
SCA5_Lab08
SCA9_Lab08
SCA1_Lab08


PFGE-XbaI
Kb
2000
1400
1200
1000
800.00
700.00
550.00
500.00
450.00
400.00
350.00
300.00
250.00
220.00
200.00
140.00
120.00
100.00
80.00
60.00
40.00
30.00
20.00
10.00
PFGE-XbaI
70
80
90
100
Key
LAB11_SCA03_Lab11
LAB11_SCA02_Lab11
LAB11_SCA08_Lab11
LAB11_SCA06_Lab11
LAB11_SCA07_Lab11
LAB11_SCA04_Lab11
LAB11_SCA10_Lab11
LAB11_SCA05_Lab11
LAB11_SCA09_Lab11
LAB11_SCA01_Lab11

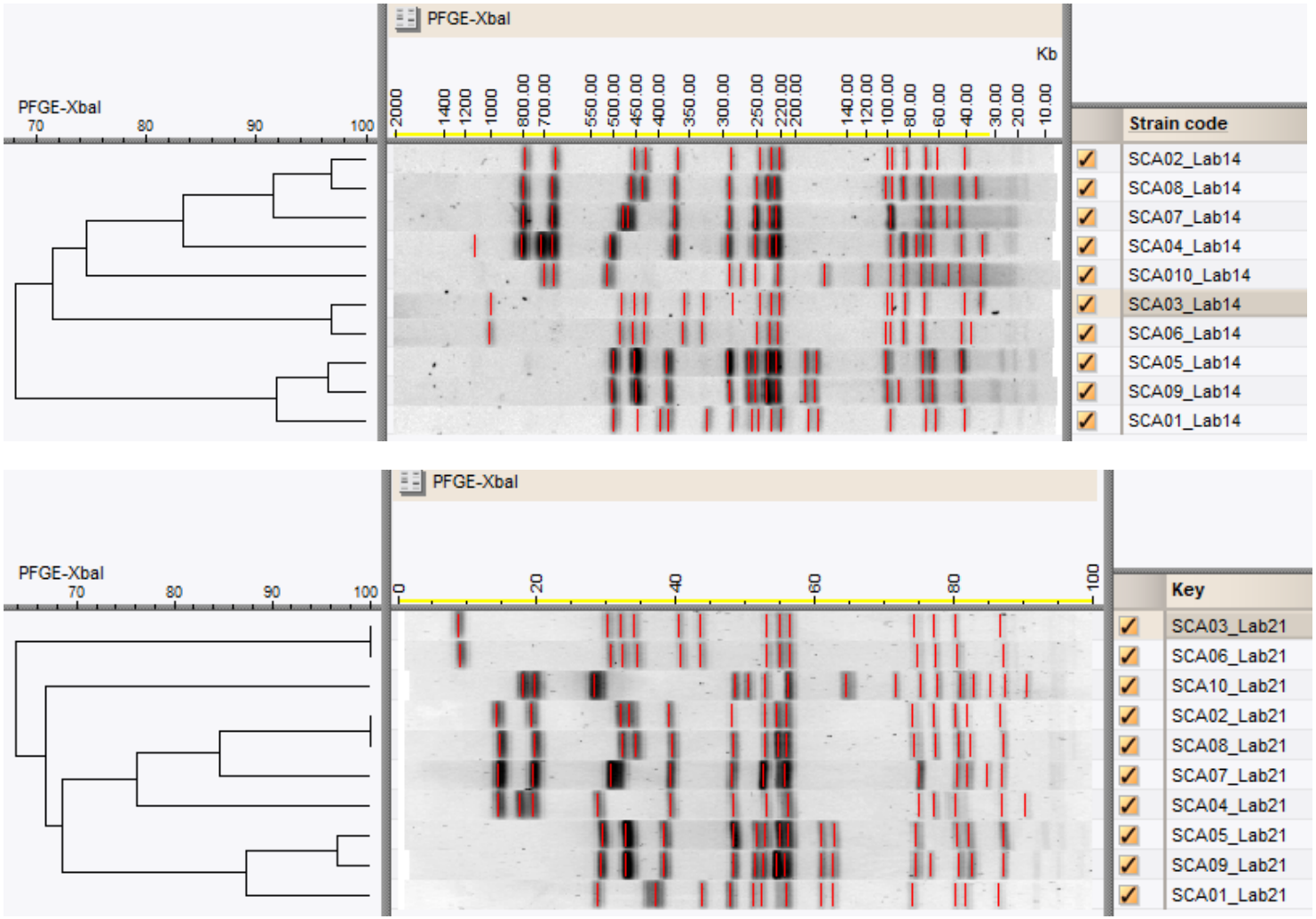

PFGE-XbaI
Kb
2000
1400
1200
1000
800.00
700.00
550.00
500.00
450.00
400.00
350.00
300.00
250.00
220.00
200.00
140.00
120.00
100.00
80.00
60.00
40.00
30.00
20.00
10.00
PFGE-XbaI
70
80
90
100
Strain code
SCA02_Lab14
SCA08_Lab14
SCA07_Lab14
SCA04_Lab14
SCA010_Lab14
SCA03_Lab14
SCA06_Lab14
SCA05_Lab14
SCA09_Lab14
SCA01_Lab14
PFGE-XbaI
0
20
40
60
80
100
PFGE-XbaI
70
80
90
100
Key
SCA03_Lab21
SCA06_Lab21
SCA10_Lab21
SCA02_Lab21
SCA08_Lab21
SCA07_Lab21
SCA04_Lab21
SCA05_Lab21
SCA09_Lab21
SCA01_Lab21

*Figure A7.2 PFGE results after analysis in Bionumerics; Laboratory codes 08, 11, 14, 21, and 34. Dendograms re-created at the EURL-*Salmonella *using the Dice coefficient with both tolerance and optimization at 1,5%*

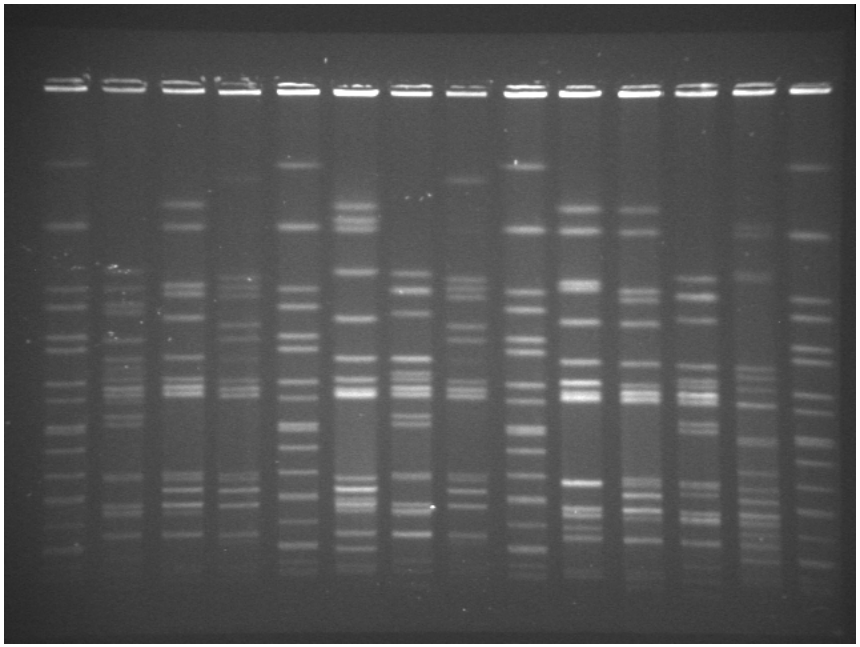


Figure A7.3 PFGE image submitted by Laboratory code 25

Annex 8 MLVA results cluster analysis part

Lab code	SCA01	SCA02	SCA03	SCA04	SCA05
REF	**3-13-9-NA-211**	**3-16-17-18-311**	**3-16-7-17-311**	**2-20-08-11-212**	**3-11-9-NA-211**
11	03-13-09-NA-211	03-16-17-18-311	03-16-07-17-311	02-20-08-11-212	03-11-09-NA-211
13	03-13-09-NA-211	03-16-17-18-311	03-16-07-17-311	02-20-08-11-212	03-11-09-NA-211
14	3-13-9-00-211	3-16-17-18-311	3-16-7-17-311	2-20-8-11-212	3-11-9-00-211
20	3 13 9 NA 211	3 16 17 18 311	3 16 7 17 311	2 20 8 11 212	3 11 9 NA 211
21	03-13-09-00-211	03-16-17-18-311	03-16-07-17-311	02-20-08-11-212	03-11-09-00-211
22	3-13-9-00-0211	3-16-17-18-0311	3-16-7-17-0311	2-20-8-11-0212	3-11-9-00-0211
26	03-13-09-00-211	03-16-17-18-311	03-16-07-17-311	02-20-08-11-212	03-11-09-00-211
34	03-13-09-NA-0211	03-16-17-18-0311	03-16-07-17-0311	02-20-08-11-0212	03-11-09-NA-0211

Lab code	SCA06	SCA07	SCA08	SCA09	SCA10
REF	**3-16-7-17-311**	**5-9-14-9-211**	**3-14-17-25-311**	**3-13-9-NA-211**	**2-12-7-9-212**
11	03-16-07-17-311	05-09-14-09-211	03-14-17-25-311	03-13-10-NA-211	02-12-07-09-212
13	03-16-07-17-311	05-09-14-09-211	03-14-17-25-311	03-13-09-NA-211	02-12-07-09-212
14	3-16-7-17-311	5-9-14-9-211	3-14-17-25-311	3-13-9-00-211	2-12-7-9-212
20	3 16 7 17 311	5 9 14 9 211	3 14 17 25 311	3 13 9 NA 211	2 12 7 9 212
21	03-16-07-17-311	05-09-14-09-211	03-14-17-25-311	03-13-09-00-00	02-12-07-09-00
22	3-16-7-17-0311	5-9-14-9-0211	3-14-17-25-0311	3-13-9-00-0211	2-12-7-9-0212
26	03-16-07-17-311	05-09-14-09-211	03-14-17-25-311	03-13-09-00-211	02-12-07-09-212
34	01-16-07-17-0311	05-09-14-09-0211	03-14-17-25-0311	03-13-09-NA-0211	02-12-07-09-0212

Loci reported in the order: STTR9, STTR5, STTR6, STTR10, STTR3.
REF: Expected MLVA results.
In blue: Deviation from the expected result.

Annex 9 WGS results cluster analysis part, methods used by the participants

Lab code	DNA extraction, library preparation and sequencing performed	WGS platform used	Data analysis used	Tool used for analysis	Method used for phylogenetic analysis
15	In-house	Illumina MiniSeq	cgMLST-based	Ridom SeqSphere	Minimum Spanning Tree (MST)
29	In-house	Illumina MiSeq	cgMLST-based	Ridom SeqSphere	Minimum Spanning Tree (MST)
22	In-house	Illumina MiSeq	cgMLST-based	Ridom SeqSphere	MST, pairwise comparison
EL	Outsourced	Illumina NovaSeq	cgMLST-based	Ridom SeqSphere	MST, pairwise comparison
17	In-house	Illumina MiSeq	cgMLST-based	Ridom SeqSphere	Neighbor joining (NJ)
26	In-house	Illumina NextSeq	cgMLST-based	Ridom SeqSphere	Neighbor joining (NJ)
28	1: in-house; 2,3: outsourced	Illumina NextSeq	cgMLST-based	Ridom SeqSphere	Neighbor joining (NJ)
11	In-house	Illumina MiSeq	cgMLST-based	in-house galaxy	Neighbor joining (NJ)
27	In-house	Illumina MiSeq	cgMLST-based	BioNumerics Center for Genomic Epidemiology	Single Linkage
34	In-house	Illumina MiSeq	cgMLST-based	inhouse automated CHEWBACCA based Pipeline	Single linkage hierarchical clustering
14	In-house	Illumina MiSeq	SNP-assembly-based	KSNP3	Neighbor joining (NJ)
32	In-house	Illumina NextSeq	SNP-reference-based	Snippy + Gubbins	Maximum likelihood (ML)
20	In-house	Illumina MiSeq	SNP-reference-based	In-house pipeline	Minimum spanning tree
9	1,2: in-house; 3: outsourced;	Illumina MiSeq	SNP-reference-based	Trimmomatic, Spades, NDtree, Seqsero	Neighbor joining (NJ)
16	In-house	Illumina MiSeq	SNP-reference-based	BWA, FreeBayes, vcflib i vcf-kit, Disty McMatrixface	Neighbor joining (NJ)

Data sorted by “Data analysis used”, and “Method used for phylogenetic analysis”.

Annex 10 WGS results cluster analysis part, QC criteria as indicated by the participants

Lab code	Listed	Criterion	Threshold (if applicable)
9	Quali1	Serotyping using Seqsero	Identification of Salmonella serotype
9	Quali2	GC content	Similar between strains
9	Quant1	Coverage	>30X
9	Quant2	N50	50000
9	Quant3	Number of contigs	<300
9	Quant4	Reference coverage	>90%
9	Quant5	Insert size	300-400 bp
11	Quali1	Contamination check (Kraken)	
11	Quali2	Serotyping (SeqSero)	
11	Quant1	Median coverage	>20x
11	Quant2	Number of contigs	<500
11	Quant3	cgMLST genes found	>95%
11	Quant4	Total length	4.5 Mb - 5.5 Mb
11	Quant5	N50	> 30 kb
14	Quant1	raw reads theoretical coverage	> 30X
14	Quant2	raw reads average	> 24 raw reads Q30
14	Quant3	Q30	> 80
14	Quant4	contigs Assembly size	> 4.5 Mases AND < 6.0 Mbases
14	Quant5	N50	> 100000
14	Quant6	Largest Contig	> 500000bp
14	Quant7	Predicted CDSs	> 4000 AND < 6000
14	Quant8	GC%	> 49% AND < 53%
15	Quali1	Contamination check	
15	Quali2	Species Match Identity	
15	Quant1	coverage	50X but if it's less, the number of targets found should be >95 %

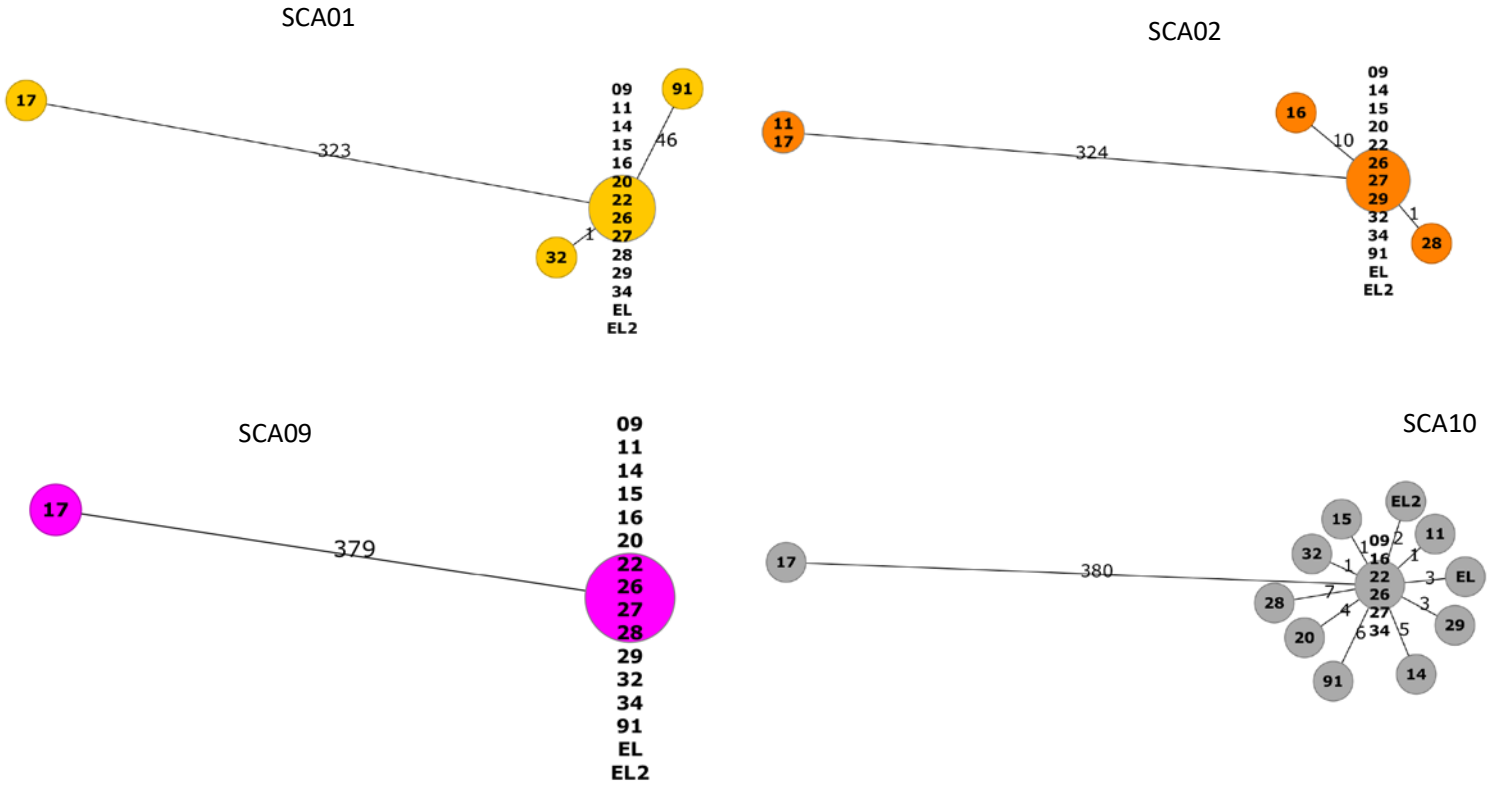
Lab code	Listed	Criterion	Threshold (if applicable)
15	Quant2	allele calling	cgMLST found and called > 95 %
15	Quant3	length of contig assembly	< reference genome + 10%
16	Quali1	contamination: kraken2/centrifuge	
16	Quant1	coverage	>30
16	Quant2	Phred score	>15
16	Quant3	Mean mapping quality	>30
16	Quant4	Min read depth	>= 10
17	Quant1	coverage	> 25x
17	Quant2	N50	> 10 000
17	Quant3	% of good cgMLST targets	>90% of 3002 targets
20	Quali1	Salmonella sp. content	80%
20	Quali2	Possibility of mixed strain sample, by looking at nr of heterogenous positions after alignment to reference	
20	Quant1	Mean base quality	20
20	Quant2	Mean coverage	10
22	Quant1	allele calling result	percentage of good targets ~ 98%
22	Quant2	No. of contigs	200 bases (contigs shorter than 200 bases have to be ignored)
22	Quant3	Assembly length	~ 5Mbases for Salmonella
22	Quant4	coverage	minimum 20-30x
26	Quali1	Percent matching targets in S. enterica cgMLST scheme	More than 90%
26	Quali2	GC percentage	51.9-52.2%
26	Quant1	phred score reads	more than 30
26	Quant2	number of contigs assembly	less than 60
26	Quant3	N50	More than 400,000
26	Quant4	Average assembly coverage	more than 10 reads
27	Quali1	Assembled Genome Length	Approx. 4.9 +- 10%
27	Quali2	Serotype (Seqsero within Bionumerics plugin)	NA

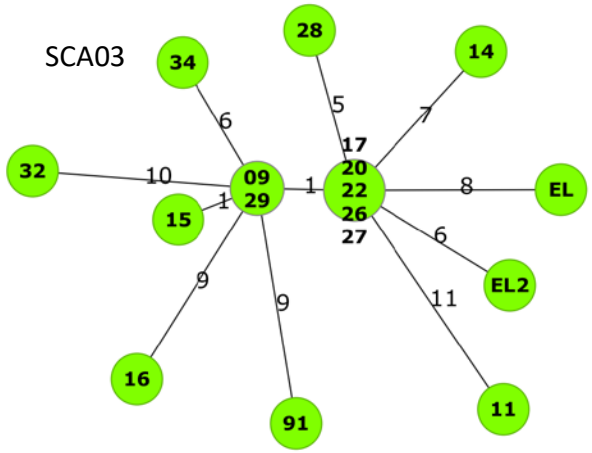
Lab code	Listed	Criterion	Threshold (if applicable)
27	Quali3	MLST (Bionumerics)	7 Loci Achtman
27	Quali4	Species Identification	K-mer finder (Center for Genomic Epidemiology)
27	Quant1	Coverage	Greater than 30X
27	Quant2	No. of Contigs	400 or Less
27	Quant3	N50	Greater than 15000
27	Quant4	Core Genome	At least 98%
28	Quali1	Contamination (Confindr)	around 10% (appreciation)
28	Quali2	Confirmation of genus (MLST/Python)	
28	Quali3	Gap Closing (GapCloser)	
28	Quali4	Genome assembly evaluation (Quast)	
28	Quali5	De novo assembly (Spades)	
28	Quant1	Coverage (BBtool)	min 30X, max 100X
28	Quant2	Trimming (Trimmomatic)	Min length : 50pb, Phred score < 20
28	Quant3	scaffolding (MeDuSa)	Delete scaffolds <200b
28	Quant4	Breath coverage (Python)	min coverage : 80%
28	Quant5	N50 (Quast)	No threshold (appreciation)
28	Quant6	Number of contigs (Quast)	No threshold (appreciation)
29	Quali1	Perc. Good cgMLST Targets	> 99 %
29	Quant1	Quality score	30
29	Quant2	Coverage	50
29	Quant3	N50	80 000
32	Quali1	Contamination	KmerID
32	Quali2	Confirmation of genus	LmerID
32	Quali3	Number of reads	NA
32	Quali4	GC content	%
32	Quant1	Coverage	10 x
32	Quant2	Number of contigs	250
32	Quant3	N50	50,000
32	Quant4	7-loci MLST mean coverage	30 x

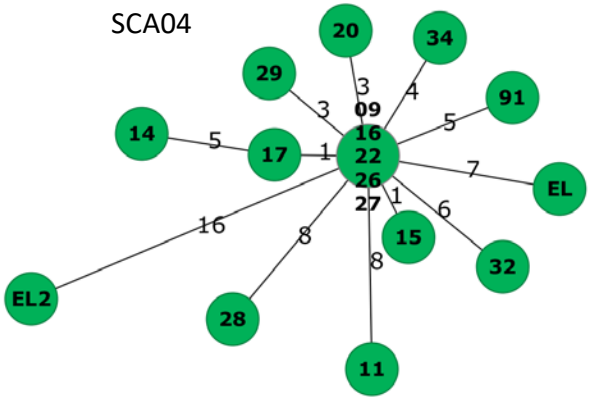
Lab code	Listed	Criterion	Threshold (if applicable)
34	Quali1	confirmation of serotyping	predicted species (mash search) and SISTR serotyping result
34	Quali2	single copy orthologs (genome completeness)	nearly all single copy orthologs present
34	Quali3	duplicated orthologs (contamination check)	almost no duplicated orthologs present
34	Quant1	number of contigs	< 200
34	Quant2	Q30 base fraction	> 0.80
34	Quant3	coverage depth	> 30
34	Quant4	fraction of reads uniquely assigned to Salmonella enterica (via KRAKEN2)	> 0.90
34	Quant5	total length	4.5-5.5 kb for Salmonella

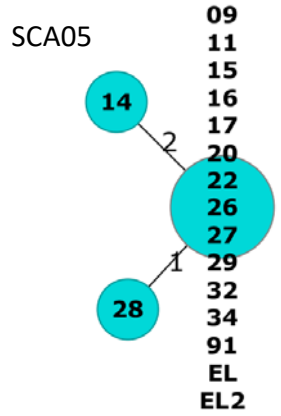
Annex 11 WGS results cluster analysis part, Minimum Spanning Tree per strain

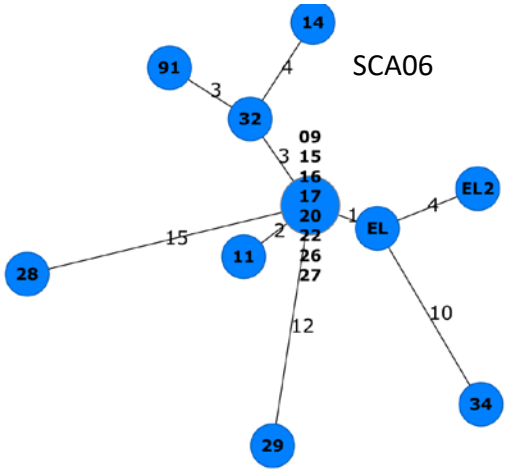
MST for each strain, using all participants' assembled raw data (Ridom SeqSphere+, cgMLST including all 3002 targets, pairwise ignoring missing values).
Note that results for Laboratory 17 indicate a swap between strains SCA01 and SCA02, as well as between SCA09 and SCA10. Results for Laboratory 16 indicate a swap between strains SCA02 and SCA03. Laboratory 11 most likely tested/analysed strain SCA01 twice: both as SCA01 and SCA02 (see Figure 10 and Figure 11).

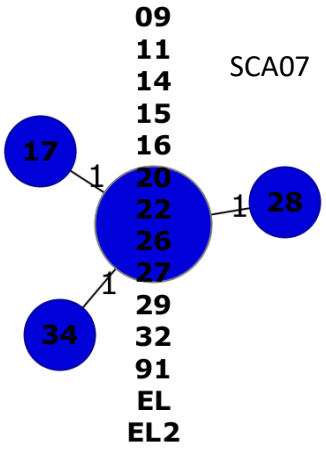


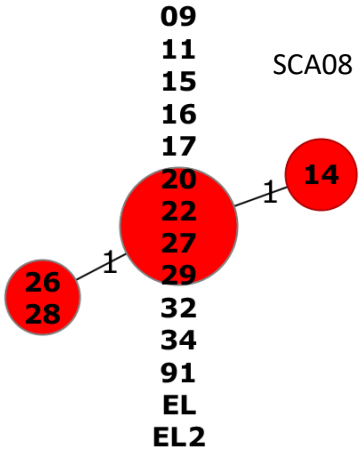

SCA03
28
34
14
6
5
7
17
20
22
26
27
32
10
09
29
1
8
EL
15
1
6
EL2
9
9
11
16
91
11


SCA04
20
34
29
3
3
4
91
09
16
22
26
27
14
5
17
1
5
EL
7
16
1
6
15
8
8
32
EL2
28
11


SCA05
09
11
15
16
17
20
22
26
27
29
32
34
91
EL
EL2
14
2
1
28


SCA06
14
91
4
3
32
09
15
16
17
20
22
26
27
3
EL2
15
1
4
EL
2
28
11
12
10
34
29


SCA07
09
11
14
15
16
20
22
26
27
29
32
91
EL
EL2
17
1
1
28
1
34


SCA08
09
11
15
16
17
20
22
27
29
32
34
91
EL
EL2
14
1
1
26
28

Annex 12 Results QC parameters on the *de novo* assembled genomes, per participant

All statistics are based on contigs of size ≥ 500 bp.

	Labcode: 09		Platform used: MiSeq							
Parameters:	**SCA01**	**SCA02**	**SCA03**	**SCA04**	**SCA05**	**SCA06**	**SCA07**	**SCA08**	**SCA09**	**SCA10**
# Contigs	75	102	87	72	85	82	61	77	75	53
Largest contig	908040	682099	682099	660595	908040	682099	907776	602604	791358	638137
Total length	4954038	5000653	4974807	4914657	5041369	4974329	4902582	4981313	4992276	4925926
N50	270591	225812	225691	257476	278830	225691	293146	223055	270584	257612
Total sequences (_R1 & _R2)	1422678	1009850	746338	809780	1402090	1094666	1147366	1068494	861882	954738
Read length	300	300	300	300	300	300	300	300	300	300
Coverage	86,2	60,6	45,0	49,4	83,4	66,0	70,2	64,4	51,8	58,1

	Labcode: 11		Platform used: MiSeq							
Parameters:	**SCA01**	**SCA02**	**SCA03**	**SCA04**	**SCA05**	**SCA06**	**SCA07**	**SCA08**	**SCA09**	**SCA10**
# Contigs	81	127	112	106	96	90	89	87	93	71
Largest contig	843389	925407	682099	660619	908040	682099	907776	682087	791358	638131
Total length	4956802	4987577	4994825	4940522	5057601	4981399	4921966	4991841	5008040	4938061
N50	261279	261279	225691	271254	316075	225691	376819	225812	270584	257612
Total sequences (_R1 & _R2)	1193740	1580108	1561494	1511436	1336668	1117050	1542100	1323918	1482664	1810236
Read length	250	250	250	250	250	250	250	250	250	250
Coverage	60,2	79,2	78,2	76,5	66,1	56,1	78,3	66,3	74,0	91,6

	Labcode: 14		Platform used: MiSeq							
Parameters:	**SCA01**	**SCA02**	**SCA03**	**SCA04**	**SCA05**	**SCA06**	**SCA07**	**SCA08**	**SCA09**	**SCA10**
# Contigs	158	136	146	165	153	105	115	91	138	111
Largest contig	172600	289257	298645	225813	251369	326644	274975	445417	277910	244464
Total length	4936582	4975780	4953659	4900519	5018072	4962268	4888261	4973163	4986188	4916894
N50	74173	99845	85628	61167	69192	138898	79936	127293	82467	96893
Total sequences (_R1 & _R2)	1486388	1380712	1285526	1150450	1531108	1266386	1194896	1008498	816734	1115088
Read length	250	250	250	250	250	250	250	250	250	250
Coverage	75,3	69,4	64,9	58,7	76,3	63,8	61,1	50,7	40,9	56,7

	Labcode: 15		Platform used: MiniSeq							
Parameters:	**SCA01**	**SCA02**	**SCA03**	**SCA04**	**SCA05**	**SCA06**	**SCA07**	**SCA08**	**SCA09**	**SCA10**
# Contigs	78	85	82	57	77	81	66	84	77	60
Largest contig	633558	500233	600535	600584	631977	604678	435961	377256	478786	550480
Total length	4932540	4968031	4951103	4893413	5017141	4952371	4884380	4963913	4977289	4910916
N50	239146	204553	204552	225831	267192	176967	213830	176966	271050	224033
Total sequences (_R1 & _R2)	3206856	853470	1196460	1331260	1452052	1767978	2390072	2724510	2670826	1550186
Read length	150	150	150	150	150	150	150	150	150	150
Coverage	97,5	25,8	36,2	40,8	43,4	53,5	73,4	82,3	80,5	47,3

	Labcode: 16		Platform used: MiSeq							
Parameters:	**SCA01**	**SCA02**	**SCA03**	**SCA04**	**SCA05**	**SCA06**	**SCA07**	**SCA08**	**SCA09**	**SCA10**
# Contigs	72	78	72	57	79	87	58	72	69	50
Largest contig	825470	602793	682099	660565	907856	601748	907776	682087	950408	638137
Total length	4951557	4969982	4982995	4907060	5037139	4974723	4901098	4978175	4993474	4924665
N50	270591	223066	225812	271254	270591	223066	293146	225812	282867	271390
Total sequences (_R1 & _R2)	1959336	1235966	1088148	1299552	1386138	1351590	1147940	1562870	1141768	1341836
Read length	300	300	300	300	300	300	300	300	300	300
Coverage	118,7	74,6	65,5	79,4	82,6	81,5	70,3	94,2	68,6	81,7

	Labcode: 17		Platform used: MiSeq							
Parameters:	**SCA01**	**SCA02**	**SCA03**	**SCA04**	**SCA05**	**SCA06**	**SCA07**	**SCA08**	**SCA09**	**SCA10**
# Contigs	79	72	76	703	67	79	66	72	62	71
Largest contig	682099	820518	602921	45024	907763	507823	793777	682087	638137	790934
Total length	4986614	4946800	4964930	4702136	5030697	4966225	4898964	4976635	4929911	4989700
N50	204653	204653	204652	12643	282782	223066	222879	225812	271390	270584
Total sequences (_R1 & _R2)	1400032	996178	1244486	1996932	1426188	1539922	1030866	2462140	1399934	2264506
Read length	300	300	300	300	300	300	300	300	300	300
Coverage	84,2	60,4	75,2	127,4	85,0	93,0	63,1	148,4	85,2	136,2

	Labcode: 20		Platform used: MiSeq							
Parameters:	**SCA01**	**SCA02**	**SCA03**	**SCA04**	**SCA05**	**SCA06**	**SCA07**	**SCA08**	**SCA09**	**SCA10**
# Contigs	75	86	155	139	70	80	82	78	94	205
Largest contig	824788	318338	168871	327271	907856	602616	815066	746688	463813	244618
Total length	4948752	4982360	4960627	4902685	5032734	4975510	4902557	4991152	4986645	4897649
N50	172971	169938	61050	85302	257784	204652	181579	204653	170901	51660
Total sequences (_R1 & _R2)	1419940	1829510	1466150	743582	1426386	5280334	1854598	4005162	1846478	1125620
Read length	250	250	250	250	250	250	250	250	250	250
Coverage	71,7	91,8	73,9	37,9	70,9	265,3	94,6	200,6	92,6	57,5

	Labcode: 22		Platform used: MiSeq							
Parameters:	**SCA01**	**SCA02**	**SCA03**	**SCA04**	**SCA05**	**SCA06**	**SCA07**	**SCA08**	**SCA09**	**SCA10**
# Contigs	111	90	93	74	78	80	63	81	82	71
Largest contig	424261	602616	361437	436620	825470	602729	622366	472244	873560	600396
Total length	4952328	4984233	4967278	4910947	5035242	4964837	4897932	4976844	5000232	4931910
N50	96518	172402	163671	193520	270591	184544	222879	185794	270584	191696
Total sequences (_R1 & _R2)	872716	1822612	1162750	1044528	2370350	1574588	1857726	1700754	1836260	1577052
Read length	300	300	300	300	300	300	300	300	300	300
Coverage	52,9	109,7	70,2	63,8	141,2	95,1	113,8	102,5	110,2	95,9

	Labcode: 26		Platform used: NextSeq							
Parameters:	**SCA01**	**SCA02**	**SCA03**	**SCA04**	**SCA05**	**SCA06**	**SCA07**	**SCA08**	**SCA09**	**SCA10**
# Contigs	76	86	83	62	79	87	63	82	75	61
Largest contig	634077	605276	601156	549980	436388	376810	435961	377066	867905	376690
Total length	4932538	4968691	4952529	4892954	5018876	4951954	4885362	4965219	4978829	4910964
N50	239915	176966	176967	196238	239317	176967	247375	178448	271050	222629
Total sequences (_R1 & _R2)	11423094	14223844	12885972	10748666	10438540	12730372	11231318	11231318	6177454	10519296
Read length	150	150	150	150	150	150	150	150	150	150
Coverage	347,4	429,4	390,3	329,5	312,0	385,6	344,8	339,3	186,1	321,3

	Labcode: 27		Platform used: MiSeq							
Parameters:	**SCA01**	**SCA02**	**SCA03**	**SCA04**	**SCA05**	**SCA06**	**SCA07**	**SCA08**	**SCA09**	**SCA10**
# Contigs	62	67	71	51	68	72	46	65	65	49
Largest contig	908040	602616	602616	660619	908040	601748	907776	602604	873560	638137
Total length	4949515	4980862	4965498	4905484	5031382	4965807	4894356	4976608	4989832	4923118
N50	270591	223065	223066	271254	282875	223066	293146	223055	270584	271297
Total sequences (_R1 & _R2)	1985088	1638620	1346878	1703378	1496484	1832364	842482	1733024	1926500	1258750
Read length	300	300	300	300	300	300	300	300	300	300
Coverage	120,3	98,7	81,4	104,2	89,2	110,7	51,6	104,5	115,8	76,7

	Labcode: 28		Platform used: NextSeq							
Parameters*:	**SCA01**	**SCA02**	**SCA03**	**SCA04**	**SCA05**	**SCA06**	**SCA07**	**SCA08**	**SCA09**	**SCA10**
# Contigs	67	65	72	49	59	74	43	66	64	61
Largest contig	528960	680789	602294	549866	526933	601816	740154	681581	681573	547921
Total length	4941268	4968016	4955718	4894971	5019832	4956355	4886663	4955262	4980375	4915617
N50	223624	222755	204921	222287	257159	222714	293015	204572	239035	225636
Total sequences (_R1 & _R2)	8066646	7244928	12734320	9291584	9952710	10520338	14176540	8825834	11643310	12902998
Read length	150	150	150	150	150	150	150	150	150	150
Coverage	244,9	218,7	385,4	284,7	297,4	318,4	435,2	267,2	350,7	393,7

*Based on filtered CLC assembly file

	Labcode: 29		Platform used: MiSeq							
Parameters:	**SCA01**	**SCA02**	**SCA03**	**SCA04**	**SCA05**	**SCA06**	**SCA07**	**SCA08**	**SCA09**	**SCA10**
# Contigs	76	73	97	56	73	77	50	75	80	64
Largest contig	8 25470	682099	682099	600683	825470	682099	907776	606236	791174	600396
Total length	4950054	4980137	4992507	4905049	5034501	4971199	4895868	4977034	4992447	4922968
N50	270591	225812	225691	271254	316036	225691	376819	223055	270584	246906
Total sequences (_R1 & _R2)	2494772	2313504	2619758	2660366	2877154	2581866	3071232	3286574	2761860	2739204
Read length	250	250	250	250	250	250	250	250	250	250
Coverage	126,0	116,1	131,2	135,6	142,9	129,8	156,8	165,1	138,3	139,1

	Labcode: 32		Platform used: NextSeq							
Parameters:	**SCA01**	**SCA02**	**SCA03**	**SCA04**	**SCA05**	**SCA06**	**SCA07**	**SCA08**	**SCA09**	**SCA10**
# Contigs	76	82	80	137	108	78	60	85	77	61
Largest contig	633837	604678	604699	549405	631410	605782	605586	604678	633715	550495
Total length	4934362	4967818	4952222	5172610	5035967	4953633	4882720	4962729	4976463	4910086
N50	239147	176966	178448	178448	267192	178448	292828	176966	267185	224033
Total sequences (_R1 & _R2)	9191630	9758850	12970380	11414492	12330584	10741440	13432400	10693688	6652048	7783554
Read length	150	150	150	150	150	150	150	150	150	150
Coverage	279,4	294,7	392,9	331,0	367,3	325,3	412,7	323,2	200,5	237,8

	Labcode: 34		Platform used: MiSeq							
Parameters:	**SCA01**	**SCA02**	**SCA03**	**SCA04**	**SCA05**	**SCA06**	**SCA07**	**SCA08**	**SCA09**	**SCA10**
# Contigs	61	67	66	51	64	68	43	68	67	50
Largest contig	907856	682099	682099	660776	908040	689437	907776	682087	873744	638137
Total length	4948374	4981351	4965974	4906363	5034645	4965751	4895280	4976721	4990054	4922924
N50	270591	225812	225691	271254	316075	225691	376816	225812	270584	257612
Total sequences (_R1 & _R2)	1219994	1383074	1217600	1708016	1885768	1626914	1381662	1347588	1206130	1526118
Read length	250	250	250	250	250	250	250	250	250	250
Coverage	61,6	69,4	61,3	87,0	93,6	81,9	70,6	67,7	60,4	77,5

	Labcode: 91		Platform used: HiSeq							
Parameters:	**SCA01**	**SCA02**	**SCA03**	**SCA04**	**SCA05**	**SCA06**	**SCA07**	**SCA08**	**SCA09**	**SCA10**
# Contigs	81	88	77	69	85	78	64	81	78	63
Largest contig	825467	605064	604976	600143	825668	604924	433733	604925	868403	600428
Total length	4986819	4959606	4949561	4890040	5010689	4945494	4879769	4956039	4970787	4905604
N50	260673	176922	182258	196111	239250	178321	223167	178321	239253	224815
Total sequences (_R1 & _R2)	5175900	4507610	5214848	3965926	3679784	4112746	3923542	4613488	4852980	3761974
Read length	100	100	100	100	100	100	100	100	100	100
Coverage	103,8	90,9	105,4	81,1	73,4	83,2	80,4	93,1	97,6	76,7

	Labcode: EL0*		Platform used: NovaSeq							
Parameters:	**SCA01**	**SCA02**	**SCA03**	**SCA04**	**SCA05**	**SCA06**	**SCA07**	**SCA08**	**SCA09**	**SCA10**
# Contigs	54	77	72	57	70	n.a.	54	63	56	54
Largest contig	825990	601876	606054	601047	908773	n.a.	908509	606054	1008986	600761
Total length	4940490	4970531	4955661	4895290	5018043	n.a.	4887094	4966422	4982234	4913202
N50	279933	180265	180263	196238	239317	n.a.	247375	204745	412724	222629
Total sequences (_R1 & _R2)	11136580	15872622	14387288	23312230	25251408	n.a.	20539004	15182152	13127508	16835394
Read length	150	150	150	150	150	n.a.	150	150	150	150
Coverage	338,1	479,0	435,5	714,3	754,8	n.a.	630,4	458,5	395,2	514,0

*EL0: EURL-*Salmonella*, 7-10-2019

	Labcode: EL*		Platform used: NovaSeq							
Parameters:	**SCA01**	**SCA02**	**SCA03**	**SCA04**	**SCA05**	**SCA06**	**SCA07**	**SCA08**	**SCA09**	**SCA10**
# contigs	73	73	86	59	73	75	58	73	72	56
Largest contig	826184	605764	605648	601047	825816	472373	825920	605822	951325	600761
Total length	4936562	4970554	4968543	4895283	5018702	4955312	4887641	4968095	4979148	4911686
N50	174625	180265	180266	196238	239140	180266	213830	180265	279925	222629
Total sequences (_R1 & _R2)	7629074	8206328	9493616	9027724	9272236	9062964	7323550	9443596	6114048	7836404
Read length	150	150	150	150	150	150	150	150	150	150
Coverage	231,8	247,6	286,6	276,6	277,1	274,3	224,8	285,1	184,2	239,3

*EL=EL1=PT2019: EURL-*Salmonella*, 25-11-2019

	Labcode: EL2*		Platform used: NovaSeq							
Parameters:	**SCA01**	**SCA02**	**SCA03**	**SCA04**	**SCA05**	**SCA06**	**SCA07**	**SCA08**	**SCA09**	**SCA10**
# contigs	71	74	73	59	72	73	47	57	75	57
Largest contig	826184	605474	605937	600760	825816	605764	825726	730235	868368	600761
Total length	4935358	4972132	4956199	4893677	5018581	4956542	4888681	4968608	4978328	4912741
N50	224019	180265	180266	196773	239140	180267	375603	335994	239141	222628
Total sequences (_R1 & _R2)	6715376	5718330	6364740	6626292	5104756	5694442	5371020	3428316	5651150	6181624
Read length	150	150	150	150	150	150	150	150	150	150
Coverage	204,1	172,5	192,6	203,1	152,6	172,3	164,8	103,5	170,3	188,7

*EL2: EURL-Salmonella, 10-2-2020

RIVM
Committed to *health and sustainability* -