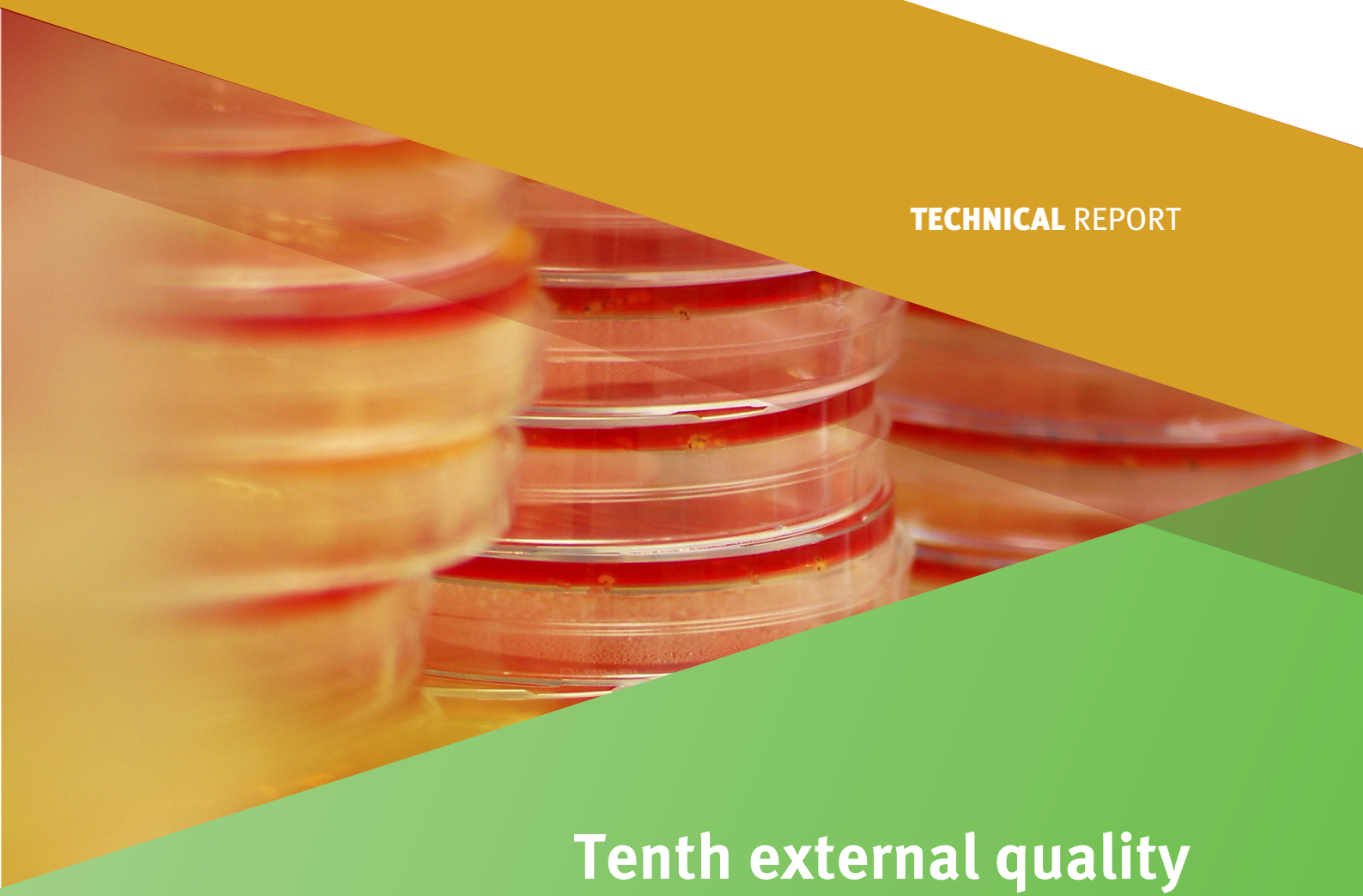
ecdc
EUROPEAN CENTRE FOR DISEASE PREVENTION AND CONTROL



TECHNICAL REPORT

Tenth external quality assessment scheme for *Salmonella* typing

www.ecdc.europa.eu

ECDC TECHNICAL REPORT

Tenth external quality assessment scheme for *Salmonella* typing

This report was commissioned by the European Centre for Disease Prevention and Control (ECDC), coordinated by Taina Niskanen (ECDC Emerging, Food- and Vector-borne Diseases Programme), and produced by Susanne Schjørring, Gitte Sørensen, Louise Gade Dahl, Kristoffer Kiil, and Eva Møller Nielsen (Section for Foodborne Infections, Statens Serum Institut, Artillerivej 5, DK-2300 Copenhagen S, Denmark).

Suggested citation: European Centre for Disease Prevention and Control. Tenth external quality assessment scheme for *Salmonella* typing. Stockholm: ECDC; 2021.

Stockholm, May 2021

ISBN: 978-92-9498-536-1
doi: 10.2900/258986
Catalogue number: TQ-03-21-228-EN-N

Contents

Abbreviations 2
Executive summary 3
1 Introduction 5
1.1 Background 5
1.2 Surveillance of non-typhoidal salmonellosis 5
1.3 Objectives 6
2 Study design 7
2.1 Organisation 7
2.2 Selection of test isolates/genomes 7
2.3 Distribution of isolates/genomes 8
2.4 Testing 8
2.5 Data analysis 8
3 Results 10
3.1 Participation 10
3.2 Multiple locus variable number of tandem repeats analysis 10
3.3 Molecular typing-based cluster analysis 14
4 Discussion 26
4.1 Multiple-locus variable number of tandem repeats analysis 26
4.2 Molecular typing-based cluster analysis 26
5 Conclusions 30
6 Recommendations 31
6.1 Laboratories 31
6.2 ECDC and FWD-Net 31
6.3 EQA provider 31
References 32

Figures

Figure 1. Participant scores for MLVA typing of the 10 *S.* Typhimurium test isolates 11
Figure 2. Average percentage scores of the 10 *S.* Typhimurium test isolates 11
Figure 3. Correct MLVA typing of two repeat *S.* Typhimurium isolates from EQA-4 to 10 (for laboratories participating in EQA-9 and/or EQA-10) 15
Figure 4. Participant scores for MLVA typing of the 10 *S.* Enteritidis test isolates 12
Figure 5. Average percentage score of the 10 MLVA *S.* Enteritidis test isolates 13
Figure 6. Correct MLVA typing of two repeat *S.* Enteritidis isolates from EQA-8 to 10 (for laboratories participating in EQA-9 and/or EQA-10) 14
Figure 7. Reported MLVA results of each test isolate 16
Figure 8. Reported SNP distances or allele differences for each test isolate to selected cluster representative isolate 20
Figure 9. Minimum spanning tree of core genome multilocus sequence typing, participant FASTQ files 21
Figure 10. Participant allele difference from reference result (EQA provider) for each test isolate 22

Tables

Table 1. Serovars of test isolates 7
Table 2. Number and percentage of laboratories submitting results for each method 10
Table 3. Results of cluster identification based on PFGE-derived data 15
Table 4. Results of cluster identification based on MLVA-derived data 15
Table 5. Results of cluster identification based on WGS-derived data 17
Table 6. Reported results of SNP-based cluster analysis 18
Table 7. Reported results of allele-based cluster analysis 19
Table 8. Summary of selected QC parameters reported by participants 23
Table 9. Results of participants' raw sequence data evaluated by EQA provider's QC pipeline 24
Table 10. Results of the participants' assessment of the EQA provided genomes 25

Annexes

Annex 1. List of participants....33
Annex 2. Participation overview EQA-9 and 10....34
Annex 3. Reason(s) for participating in EQA....35
Annex 4. Scores of MLVA results *S.* Typhimurium....36
Annex 5. Scores of MLVA results *S.* Enteritidis....37
Annex 6. EQA provider cluster analysis based on WGS-derived data....38
Annex 7. Reported cluster of closely related isolates based on PFGE-derived data....39
Annex 8. Reported sequencing details....40
Annex 9. Reported cluster of closely related isolates based on MLVA-derived data....41
Annex 10 Reported MLVA profile data....42
Annex 11. Reported cluster of closely related isolates based on WGS-derived data....43
Annex 12. Reported SNP distance and allelic differences....44
Annex 13. Additional reported QC parameters....45
Annex 14. Calculated qualitative/quantitative parameters....46
Annex 15. Accessing Genomes....52
Annex 16. EQA-9 laboratory questionnaire....55

Abbreviations

BN	BioNumerics
cgMLST	Core genome multilocus sequence typing
EQA	External Quality Assessment
FWD	Food- and waterborne diseases and zoonoses
FWD-Net	Food- and Waterborne Diseases and Zoonoses Network
MLVA	Multiple locus variable number of tandem repeats analysis
PFGE	Pulsed field gel electrophoresis
NPHRL	National public health reference laboratories
QC	Qualitative control
SNP	Single nucleotide polymorphism
SNV	Single-nucleotide variant based on cgMLST
SSI	Statens Serum Institut
ST	Sequence type
TESSy	The European Surveillance System
wgMLST	Whole genome multilocus sequence typing
WGS	Whole genome sequencing

Executive summary

This report presents the results of the 10th round of the external quality assessment (EQA-10) scheme for typing of *Salmonella enterica* subsp. *enterica* organised for the national public health reference laboratories (NPHRLs) in ECDC's Food- and Waterborne Diseases and Zoonoses network (FWD-Net), managed by the European Centre for Disease Prevention and Control (ECDC). The EQA-10 scheme was arranged by the Section for Foodborne Infections at the Statens Serum Institut (SSI) in Denmark.

Salmonellosis was the second most commonly reported zoonotic disease in the European Union (EU) in 2018, with a notification rate of 20.1 cases per 100 000 population. The most commonly reported *Salmonella* serovars in 2018 were *S.* Enteritidis, *S.* Typhimurium, monophasic *S.* Typhimurium, and *S.* Infantis. The total number of reported cases was 91 857 [3]. Since 2007, ECDC has been responsible for the EU-wide surveillance of salmonellosis, including facilitating the detection and investigation of food-borne outbreaks. Surveillance data, including certain basic typing parameters, are reported by Member States to The European Surveillance System (TESSy). Since 2012, the EQA scheme has covered molecular typing methods used for EU-wide surveillance.

The effective molecular typing-enhanced surveillance relies on the capacity of NPHRLs in the FWD-Net to produce comparable typing results. ECDC has opened the possibility to Member States to submit WGS data for *Salmonella* and *Listeria monocytogenes* to TESSy to be used for EU-wide surveillance and cross-sector comparison. The previous EQA schemes from EQA-4 to EQA-8 included assessment of the PFGE typing methods for all *Salmonella* serovars and multiple locus variable number of tandem repeats analysis (MLVA) for *Salmonella* Typhimurium (STm). Since EQA-9, the PFGE part was modified into a part where the ability of identifying a cluster based on molecular typing by either PFGE, MLVA, and/or whole genome sequencing (WGS) derived data was assessed. Since EQA-9, participants have also been able to participate in MLVA for *S.* Enteritidis (SE).

The objectives of the EQA-10 scheme were to assess the quality of data and comparability of molecular typing analysis results produced by NPHRLsin FWD-Net. Test isolates for the EQA were selected to cover isolates currently relevant for public health in Europe. Three sets of 10 isolates were selected, including *S.* Typhimurium and *S.* Enteritidis isolates for the two MLVA methods and a mixture of different sequence types (ST) in the cluster analysis.

Twenty-two laboratories signed up, and 19 completed the exercise. This is a decrease from EQA-9 (N=23) with 17%. It is unknown if the removal of the PFGE part (gel quality and analysis) was the cause of this. All 19 laboratories participated in the molecular typing-based cluster analysis. Out of the 19 laboratories participating in EQA-10, 15 (79%) performed molecular typing-based cluster analysis, which was 25% (12 laboratories participated) increase compared to EQA-9.

In total, eight laboratories participated in the *S.* Typhimurium MLVA and nine participated in the *S.* Enteritidis MLVA. For the *S.* Typhimurium and *S.* Enteritidis MLVA schemes, a lower number of participants was observed compared with previous EQA-9, when 10 laboratories participated. The performance level was high for both analyses (93%) and within the range of the previous years. One laboratory was responsible for 50% of the errors in *S.* Typhimurium and 66% in the *S.* Enteritidis.

The aim of the cluster analysis part of the EQA was to assess the NPHRL's ability to identify a cluster of genetically closely related isolates i.e. correctly categorise cluster test isolates regardless of the method used, instead of the ability to follow a specific procedure.

The cluster of closely related monophasic *S.* Typhimurium ST34 isolates could be identified by PFGE, MLVA, and WGS-derived data. The expected cluster was based on a predefined categorisation by the organiser and contained four isolates based on WGS-derived data.

Seven laboratories used PFGE for cluster analysis, and for two participants PFGE was the only cluster identification method. Six laboratories were able to identify the correct cluster using PFGE. Five laboratories used MLVA for cluster analysis in combination with either PFGE or WGS and all identified the correct cluster. One laboratory performed cluster analysis using all three methods.

The performance among the 15 participants using WGS derived data was very high, 14 (93%) correctly identified the cluster of closely related isolates. In this EQA cluster analysis, the focus has been on one test isolate (REF6), where challenges and discrepancies among the participants interpretation of whether REF6 should be included in the cluster of closely related isolates. Both inclusion and exclusion of REF6 was accepted as a correct result. Sixty percent of the participants did not include REF6 in the cluster, which is in accordance with "original" pre-defined cluster by the provider.

The results were comparable between single nucleotide polymorphism (SNP) and allele-based analysis, but the allele-based analysis gave a less clear discrimination and resulted in different interpretation by the participants (REF6).

In this EQA, the EQA provider introduced an additional part to the molecular typing-based cluster analysis: an assessment of four EQA provided genomes. In an urgent outbreak situation, the sequence data available is not always of high quality, therefore this EQA-part was designed to mimic this situation. The participants should assess additional

genomes, some were modified by the EQA provider in order to give a realistic view of different quality issues. The majority of participants (87%-100%; 13-15/15) successfully identified the genomes of high quality as a cluster isolate and 14 of 15 the non-cluster isolate. The inclusion of *E. coli* species in one of the genomes were only described by five participants, but 13 of 15 did identify the genome as a non-cluster isolate. The genome with the poor quality in one of the genomes were observed by 13 of 15 participants.

1 Introduction

1.1 Background

ECDC is an EU agency with a mandate to operate dedicated surveillance networks. The agency's mission is to identify, assess, and communicate current and emerging threats to human health from communicable diseases. ECDC's founding regulation outlines its mandate as fostering the development of sufficient capacity within the EU/EEA network for diagnosis, detection, identification, and characterisation of infectious agents that may threaten public health. ECDC maintains and extend such cooperation and support the implementation of quality assurance schemes [1].

External quality assessments (EQA) are an essential part of laboratory quality management, and use an external organiser to assess the performance of laboratories on test samples supplied specifically for the purpose of quality assessment.

ECDC has outsourced the organisation of EQA schemes for EU/EEA countries in the disease networks. EQAs aim to identify areas for improvement in laboratory diagnostic capacities relevant to epidemiological surveillance of communicable diseases as set forth in Decision No 1082/2013/EU [2] and ensure the reliability and comparability of results generated by laboratories across all EU/EEA countries.

The main objectives of the EQA schemes are to:

- assess the general standard of performance ("state-of–the-art");
- assess the effects of analytical procedures (method principle, instruments, reagents, calibration);
- support method development;
- evaluate individual laboratory performance;
- identify problem areas;
- provide continuing education; and
- identify needs for training activities.

Since 2012, the Section for Foodborne Infections at the Statens Serum Institut (SSI) in Denmark has been the EQA provider for the typing of *S. enterica* subsp. *enterica*, Shiga toxin/verocytotoxin-producing *Escherichia coli* (STEC/VTEC) and *Listeria monocytogenes*. In 2016, SSI was also granted the new round of tenders (2017 to 2020) for all three pathogens. The contracted Lot1 scheme for *Salmonella* covers MLVA typing of both *S.* Typhimurium and *S.* Enteritidis and molecular typing-based cluster analysis. This report presents the results of the *Salmonella* EQA-10.

1.2 Surveillance of non-typhoidal salmonellosis

In 2018, non-typhoidal salmonellosis (later "salmonellosis") was the second most commonly reported zoonotic disease in the EU, with 91 857 cases reported by 27 of the EU Member States (EU notification rate of 20.1 cases per 100 000 population), a small increase from 2017. As in previous years, the most commonly reported *Salmonella* serovars were *S.* Enteritidis, *S.* Typhimurium, monophasic *S.* Typhimurium, and *S.* Infantis. *S.* Newport was reported on the fifth place with an increase of 18.0% and 43.3% compared with the two previous years [3].

One of the key objectives of ECDC is to improve and harmonise the surveillance system in the EU and increase scientific knowledge of aetiology, risk factors, and the burden of food- and waterborne diseases and zoonoses (FWD). The surveillance data, including some basic typing parameters for the isolated pathogen, are reported by Member States to The European Surveillance System (TESSy). In addition to the basic characterisation of the pathogens isolated from human infections, there is a public health value in using more discriminatory typing techniques in the surveillance of foodborne infections. In 2012, ECDC initiated enhanced EU-level surveillance by incorporating molecular typing data into the reporting of foodborne pathogens. Since 2019, countries have been able to report WGS data to TESSy for *Salmonella*. The overall aims of integrating molecular typing data into EU-level surveillance are to:

- foster the rapid detection of dispersed international clusters/outbreaks;
- facilitate the detection and investigation of transmission chains and relatedness of strains across EU/EEA countries and contribute to global outbreak investigations;
- detect the emergence of new evolving pathogenic strains;
- support investigations to trace the source of an outbreak and identify new risk factors; and
- aid the study of particular pathogens' characteristics and behaviour in community of hosts.

Molecular typing-enhanced surveillance gives Member State users access to EU-wide molecular typing data for the pathogens included. It also provides users with the opportunity to perform cluster searches and cross-sector comparability of EU-level data to determine whether isolates characterised by molecular typing at the national level(s) are part of a multinational cluster that may require cross-border response collaboration.

EQA schemes offer quality support at NPHRLs performing molecular typing-enhanced surveillance and those implementing it in their surveillance system at national level.

1.3 Objectives

1.3.1 Multiple locus variable number of tandem repeats analysis typing of *S.* Typhimurium and *S.* Enteritidis

The *Salmonella* EQA-10 for MLVA aimed to determine and support the assessment of analytical results' quality (reproducibility) and comparability of *S. enterica* subsp. *enterica* serovar Typhimurium and serovar Enteritidis MLVA results in the participating laboratories. The MLVA part covered both the laboratory procedure and subsequent data analysis (calibration of raw data into correct MLVA alleles according to the nomenclature [4–5]).

1.3.2 Molecular typing-based cluster analysis

The objective of the molecular typing-based cluster analysis of *Salmonella* EQA-10 was to assess the ability of laboratories to detect a cluster of genetically closely related isolates. Laboratories could perform the analyses using PFGE, MLVA and/or derived data from WGS. In addition, the participant was offered to assess extra genomes, and determine whether the genomes were part of the defined cluster and describe their observations and considerations leading to the decision. The EQA provider had manipulated some of the genomes. In the individual reports, this analysis was not evaluated and therefore not directly commented on but will be summarised in this report.

2 Study design

2.1 Organisation

The *Salmonella* EQA-10 was funded by ECDC and arranged by SSI following the requirements in ISO/IEC 17043:2010 [6]. The EQA-10 included MLVA of *S.* Typhimurium and *S.* Enteritidis and molecular typing-based cluster analysis using either PFGE, MLVA, and/or WGS-derived data. From EQA-8 to EQA-10, a change was made excluding the quality assessment part with PFGE. EQA-10 was conducted between June 2019 and November 2019.

Invitations were emailed to ECDC contact points in FWD-Net (26 countries nominated laboratories to participate in the EQA rounds from 2017–2020) by 24 May 2019, with a deadline to respond by 2 June 2019. In addition, invitations were sent to EU enlargement countries Serbia, Turkey, and Kosovo[i], which signed up to the *Salmonella* EQA rounds from 2017-2020. Each laboratory was asked to give their reasons for participating or not participating.

Twenty-two NPHRLs in the EU/EEA and EU enlargement countries accepted the invitation to participate, but only 19 submitted results (Annex 1). This was a decrease from last year (EQA-9), when 23 countries participated. In Annex 2, details of participation in EQA-9 and EQA-10 are listed to give an overview of the trend in the number of participants.

The EQA test isolates were sent to participants from 19 June 2019. Participants were asked to submit their results by 21 October 2019 using the online form (Annex 16).

If WGS was performed, submission of the raw reads (FASTQ files) to https://sikkerftp.ssi.dk was also requested. The EQA submission protocol was distributed by email, and Excel sheets for the MLVA reference isolates and MLVA allele calling were available on the online site.

2.2 Selection of test isolates/genomes

Forty-six *Salmonella* test isolates were selected to fulfil the following criteria:

- represent commonly reported isolates in Europe;
- remain stable during the preliminary test period at the organising laboratory;
- include repeat isolates from EQA-4 through 10; and
- include genetically closely related isolates.

The 46 selected isolates were analysed using the methods in the EQA (MLVA and WGS) before and after having been re-cultured. All candidate isolates remained stable using these methods and the final test isolates and additional genomes were selected (Table 1). For the MLVA part, 10 *S.* Typhimurium and 10 *S.* Enteritidis were selected to cover common and various MLVA profiles (Annex 4-5). The 10 isolates for cluster analysis were selected to include isolates with different (or varying) relatedness and comprised different 7-gene Multi Locus Sequence Types (ST) (ST34, ST4430, ST4431 and ST5296). For the additional genomes, two were altered; one with reduced coverage and one mixed with 10% *Escherichia coli.*

Table 1. Serovars of test isolates/genomes

Method	Number of test isolates	Serovars	Annex
MLVA *S.* Typhimurium	10 isolates STm1-10	Typhimurium/monophasic Typhimurium *STm9 (3-12-9-NA-211), *STm10 (3-13-NA-NA-211)	4
MLVA *S.* Enteritidis	10 isolates SE1-10	Enteritidis **SE10 (1-10-7-3-2), **SE9 (3-11-4-4-1)	5
Cluster analysis	10 isolates REF1-10	monophasic *S.* Typhimurium (ST34, ST4430, ST4431 and ST5296)	6-7+9–12
	4 genomes REF11-REF14	monophasic *S.* Typhimurium 3 x ST34 (one with reduced coverage), 1 x ST4431 (contaminated with 10% *E. coli)*	15

**: repeat isolates included in EQA-4 to 10.*
***: repeat isolates included in EQA-8 to -10.*
NA: designates a locus not present (-2 by submission, Annex 4 and 5).

[i] This designation is without prejudice to positions on status, and in line with UNSCR 1244/99 and the ICJ Opinion on the Kosovo Declaration of Independence.

2.3 Distribution of isolates/genomes

All 30 test isolates were blinded and shipped on 19 June 2019. The protocol for the EQA exercise and a letter stating the unique isolate IDs were included in the packages and distributed individually to participants by email on 20 June 2019 as an extra precaution. Fourteen participants received their dispatched isolates within one day, four within two days, three within five days, and only one participant received the isolates after six days. The packages were shipped from SSI labelled as UN 3373 Biological Substance. No participants reported damage to the shipment or errors in the unique specific isolate IDs.

On 3 July 2019, instructions on the submission of results procedure were emailed to the participants. This included the links to the online site for uploading sequences and downloading the additional genomes and the empty submission form.

At the site, participants could download four Excel sheets; a compensatory table for both MLVA reference isolates and a sheet for the subsequent calculation of MLVA alleles for both *S.* Typhimurium and *S.* Enteritidis (MLVA part).

2.4 Testing

In the MLVA part, the 10 *S.* Typhimurium and 10 *S.* Enteritidis test isolates were tested to assess the participants' ability to obtain the true number of repeats in each of the five MLVA loci for each scheme. The participants were instructed to use ECDC's laboratory standard operating procedure for MLVA of *Salmonella enterica* serotype Typhimurium [4] and MLVA of *Salmonella enterica* serotype Enteritidis [5]. The distributed Excel sheets could be used to convert the measured fragment sizes to true allele numbers based on the results obtained for the 33 *S.* Typhimurium and 16 *S.* Enteritidis reference isolates. The allelic profiles should be submitted using the online submission form, -2 was used instead of NA when a locus was missing [4–5].

In the cluster analysis part, the participants could choose to perform the laboratory part using PFGE, MLVA and/or WGS derived data, however the cluster categorisation was based on WGS data and therefore the correct cluster delineation might be difficult to obtain by the use of less discriminatory methods, e.g. PFGE and/or MLVA. The participants were instructed to report the IDs of the isolates included in the cluster of closely related isolates by method. A PDF version of the online form was also available for the participants. (Annex 16). If MLVA was performed, the participants were instructed to report the MLVA scheme used and the number of repeats in each of the loci per isolate.

Laboratories performing WGS could use their own analysis pipeline for the cluster analysis, e.g. SNP-based or allele-based and were asked to submit the isolates, identified as cluster of closely related isolates, based on the analysis used. The laboratories could report results from up to three analyses (one main and two additional), but the detected cluster had to be based on results from the main analysis. The laboratories reported SNP distance or allelic differences between a selected cluster isolate and each test isolate, and uploaded the raw reads (FASTQ files). In this EQA, the laboratories had the possibility to submit the ST of isolates in the cluster analysis and were also asked to report the number of loci in the used allelic scheme and the name of the used SNP pipeline.

In addition, each participant could assess extra genomes (manipulated by the EQA provider) and determine whether the genomes were part of the defined cluster (Yes/No) and describe their observations and considerations leading to the decision.

2.5 Data analysis

As the participating laboratories submitted their results, the MLVA and cluster analysis results as well as raw reads, these were imported to a dedicated *Salmonella* EQA-10 BN database.

The MLVA results were evaluated according to the percentage of correctly assigned allelic profiles generating a score from 0 to 100% correct profiles.

The cluster analysis part was evaluated according to correct or incorrect identification of the cluster of closely related isolates based on a predefined categorisation by the EQA provider.

The expected cluster of closely related monophasic *S.* Typhimurium ST34 isolates contained four isolate based on WGS derived data in an allele-based analysis (cgMLST, [7]) and a SNP analysis [8], which showed at most 1 allele difference or 1 SNP distance between any two isolates in the cluster. REF6 was pre-defined outside the expected cluster, as it differed with six alleles or 10 SNPs to the expected cluster, however, the EQA provider accepted participants inclusion of REF6 as an acknowledgment that a definitive cut-off in *Salmonella* has not been formally established. The EQA provider did however express the concern (in the individual evaluation) that many unrelated isolates can potentially be within six allelic differences due to the clonal nature of monophasic *S.* Typhimurium. The cluster categorisation is based on WGS data and the correct cluster delineation might be difficult to obtain by using less discriminatory methods, e.g. MLVA and/or PFGE. Therefore, the evaluation consists of including at least all the WGS defined cluster isolates. For MLVA the cluster MLVA profile is 3-15-11-NA-211. The characteristics of the test isolates and reported results are listed in Annex 4-14.

The participant's description of the EQA-provider's manipulated genomes are listed in Annex 15. This analysis was not commented in the individual reports, but will be summarised in this report.

Individual evaluation reports were distributed to the participants in January 2020 and certificates of attendance in June 2020. If WGS data were used, the evaluation report included a quality assessment made by the EQA provider's in-house quality control pipeline (e.g. coverage, N50, sequence length and number of contigs). The evaluation report did not include an evaluation based on quality thresholds.

Four laboratories were contacted after deadline in order to get the raw reads uploaded to the ftp site. Additionally, laboratory 36 and 129 were contacted by the EQA provider after the submission deadline due to possible errors in renaming the isolates during the WGS cluster analysis. Subsequently, corresponding isolate IDs and obtained results were resubmitted by the two participants making it possible for the EQA provider to assess their performance. Only the resubmitted data for laboratory 36 and 129 are included in this report.

3 Results

3.1 Participation

The laboratories could participate in either the full EQA scheme or one part only (MLVA *S.* Typhimurium, MLVA *S.* Enteritidis and/or molecular typing-based cluster analysis based on PFGE, MLVA and/or WGS-derived data). Out of the 22 participants who signed up for the EQA, 19 managed to complete and submit their results. None of the laboratories who did not sign up at all gave a reason for not participating.

Eight laboratories completed MLVA (STm and SE) and molecular typing-based cluster analysis (PFGE, MLVA and/or WGS). Eight (42%) laboratories participated in the MLVA part, in both *S.* Typhimurium and *S.* Enteritidis, and one laboratory participated in the MLVA part for *S.* Enteritidis, only. Most of the laboratories not participating in the MLVA part responded that "MLVA *S.* Typhimurium is not relevant to our laboratory". All 19 laboratories (100%) participated in the cluster analysis part, and most of them (15, 79%) reported cluster analysis results based on WGS, whereas seven laboratories (37%) reported based on PFGE-derived, and five laboratories (26%) reported based on MLVA-derived results. Only two participants (11%) reported cluster identification using only PFGE, and 10 (53%) reported cluster identification using only WGS; none reported the cluster identification using only MLVA (Table 2). One laboratory reported "We introduced MLVA for *Salmonella* in our laboratory but we have problems with results interpretation" as the reason for participating in the cluster analysis (See Annex 3 for details).

Table 2. Number and percentage of laboratories submitting results for each method

	MLVA			Cluster analyses							All
	STm+ SE	SE only	Total	PFGE only	WGS only	PFGE + WGS	PFGE + MLVA + WGS	PFGE + MLVA	MLVA + WGS	Total	Total
Number of participants	8	1	9	2	10	2	1	2	2	19	19
Percentage of participants	89	11	47*	11	53	11	5	11	11	100	100

**: percentage of the total number of participating laboratories (19)*
STm: S. *Typhimurium*
SE: S. *Enteritidis*

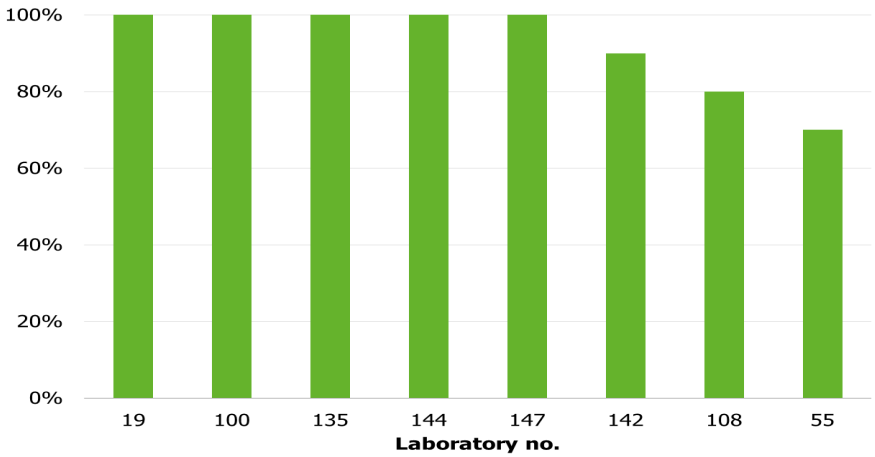
3.2 Multiple locus variable number of tandem repeats analysis

In total, nine laboratories (47%) participated in the MLVA part of the EQA, and eight of these participated in both MLVA for *S.* Typhimurium and for *S.* Enteritidis (Annex 4 and 5).

3.2.1 MLVA for *S.* Typhimurium

Eight out of the 19 participants in EQA-10 (42%) performed the MLVA typing of *S.* Typhimurium, and five of these (63%) reported the correct allelic profiles for all 10 test isolates (Figure 1). Laboratory 100 reported the correct MLVA profile for all 10 test isolates but for isolate STm1 laboratory 100 reported a single-locus variant in STTR6. Due to the fast-changing nature of this locus the EQA provider accepted the assigned number. Three participants did not assign correct MLVA profiles for some of the test isolates. Laboratory 55 had the most errors, reporting a fragment in an absent locus (STTR10) in three of the test isolates (STm5, 6 and 7). Laboratory 108 had two errors, reporting a fragment in two absent loci (STTR6 and STTR10) for isolate STm6. Furthermore, laboratory 108 also reported an incorrect allele number in STTR3 for isolate STm2. Laboratory 142 had one error, reporting an incorrect allele number in STTR3 for isolate STm8.

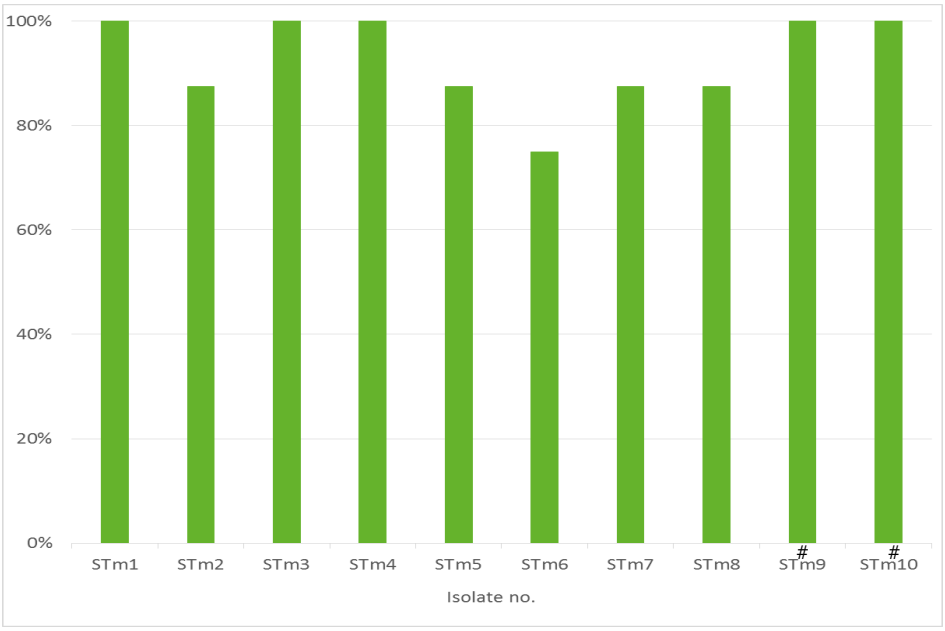
Figure 1. Participant scores for MLVA typing of the 10 *S.* Typhimurium test isolates



Arbitrary numbers represent the participating laboratories. Bars represent the number of correctly assigned MLVA profiles (including accepted profiles).

The results for each test isolate are summarised in Figure 2. The correct MLVA profile was reported for five of the 10 *S.* Typhimurium test isolates by all participants. No common isolate characteristics caused the problems (Annex 4) as the six incorrect MLVA profiles concerned five different isolates (STm2, 5, 6, 7 and 8). Only one identical error occurred as both laboratory 55 and 108 reported an incorrect fragment in STTR10 for STm6, but the two laboratories reported different fragment size. Furthermore, laboratory 108 also for STm6 reported an incorrect allele number on fragment STTR6.

Figure 2. Average percentage scores of the 10 *S.* Typhimurium test isolates

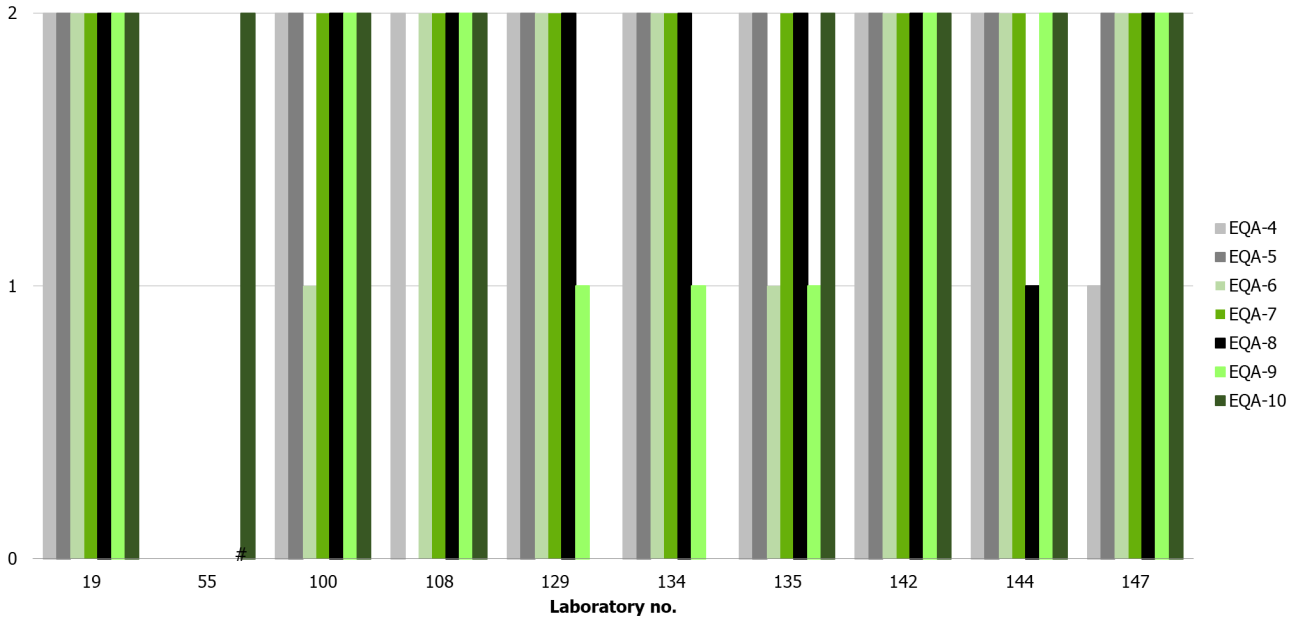


Bars represent the percentage of MLVA profiles correctly assigned by the participants.
#: repeat isolates (STm9 and STm10) in EQA-4 to 10.

To follow the development of individual laboratory performance, two repeat isolates with different allelic profiles were included in EQA-4 through 10: isolate STm9 (3-13-NA-NA-211) and STm10 (3-12-9-NA-211). Figure 3 shows the individual performance by the laboratories of these two repeat isolates during the seven EQAs (only showed for the laboratories participating at least in EQA-9 and/or EQA-10). Most participants (8/10; 80%) performed at the same or a better level than the last time they participated.

All participants in EQA-10 were able to identify the correct MLVA profile for both repeat isolates, and an increasing performance on these isolates was seen compared to EQA-9.

Figure 3. Correct MLVA typing of two repeat *S.* Typhimurium isolates from EQA-4 to 10 (for laboratories participating in EQA-9 and/or EQA-10)

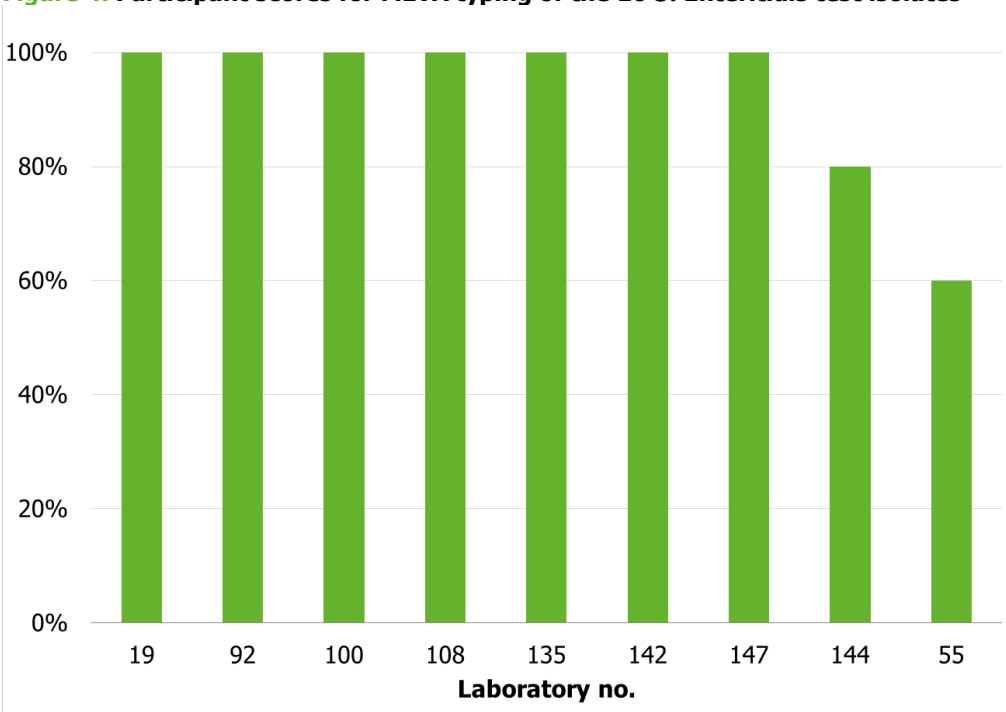


Arbitrary numbers represent the participating laboratories. Bars represent the number of correctly assigned allelic profiles for the two repeat isolates (STm9 and STm10).
#: laboratory did not correctly identify any of the two repeat isolates.

3.2.2 MLVA for *S.* Enteritidis

Nine out of the 19 participants (47%) in EQA-10 performed the MLVA typing of *S.* Enteritidis and seven (78%) of these reported the correct allelic profiles for all 10 test isolates (Figure 4).

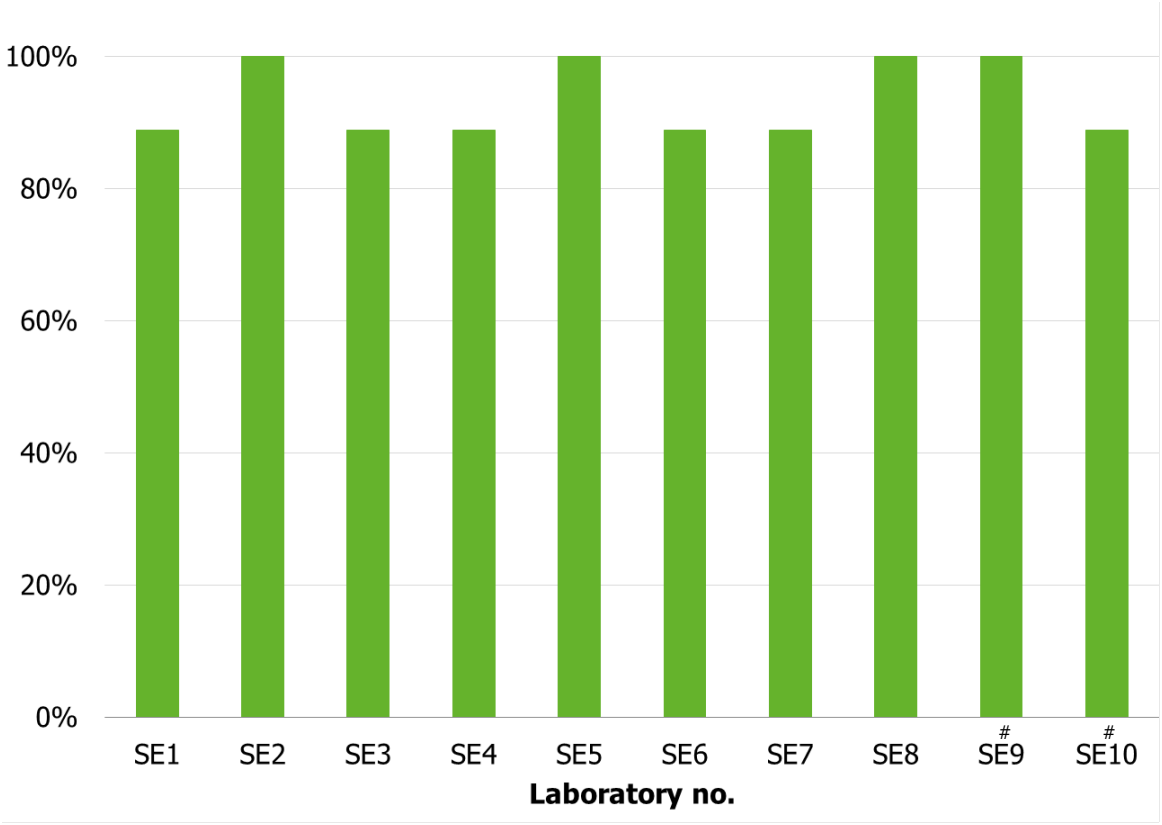
Figure 4. Participant scores for MLVA typing of the 10 *S.* Enteritidis test isolates



Arbitrary numbers represent participating laboratories. Bars represent number of correctly assigned MLVA profiles.

Laboratory 55 and 144 had four and two errors respectively in different isolates (Annex 5). Laboratory 55 reported incorrect allele numbers in SENTR4 for isolate SE1, SE3 and SE4 and in fragment SENTR5 for isolate SE7. Laboratory 144 reported an incorrect allele number in fragment SENTR7, SENTR6, SENTR4 and SE-3 for both isolate SE6 and SE10 (possible a swap of isolate ID or results).

Figure 5. Average percentage score of the 10 MLVA *S.* Enteritidis test isolates



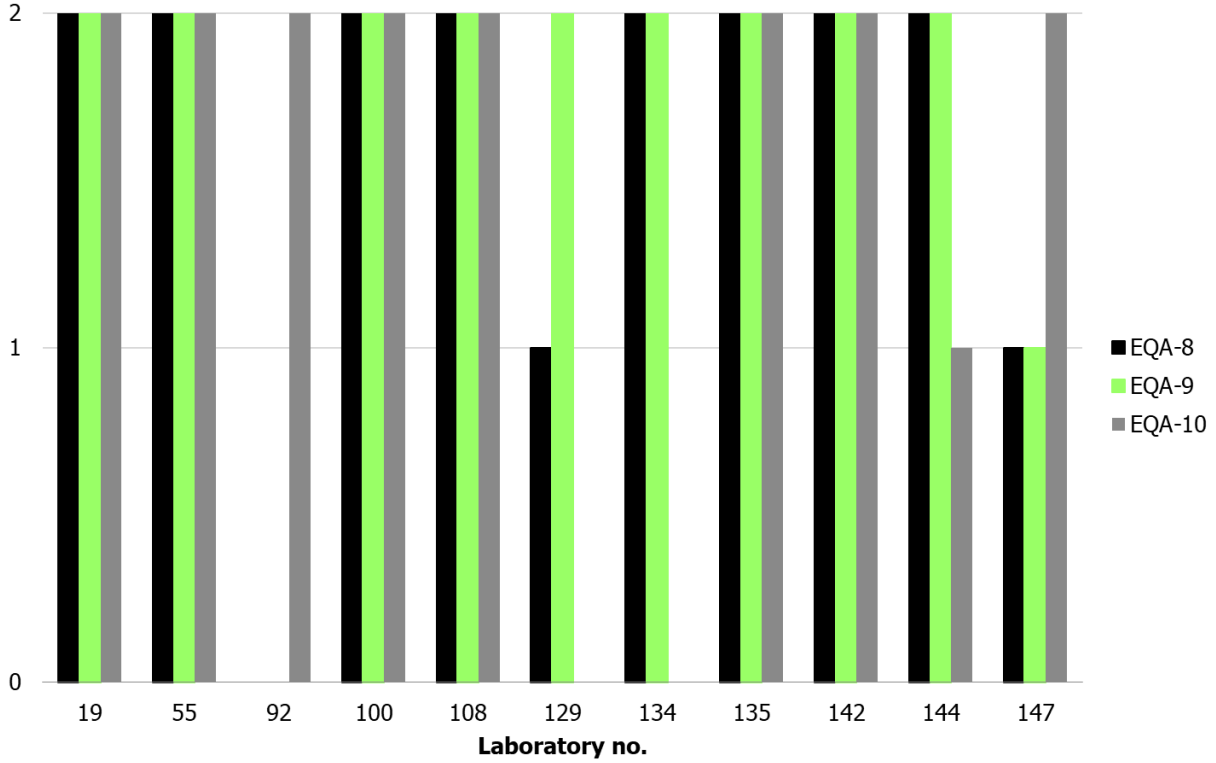
Bars represent the percentage of MLVA profiles correctly assigned by the participants.
#: repeat isolates (SE9 and SE10) in EQA-8 to 10.

The results for each test isolate are summarised in Figure 5. The correct MLVA profile was reported for four of the 10 *S.* Enteritidis test isolates by all participants. No common isolate characteristics caused the problems (Annex 5) as the six incorrect MLVA profiles concerned six different isolates.

To follow the development of individual laboratory performance, two isolates with different allelic profiles were included in EQA-8 to EQA-10: isolate SE9 (3-11-4-4-1) and SE10 (1-10-7-3-2). Figure 6 shows the individual performance by the laboratories of these two repeat isolates during the three EQAs for laboratories participating in EQA-9 and/or EQA-10. The MLVA results on the repeat isolates show stability and high performance among the participants.

All participants except one performed at the same or a better level than the last time they participated. Laboratory 144 assigned an incorrect allelic profile for repeat isolate SE10 (data might have been swapped with isolate SE6). Laboratory 92 participated for the first time in MLVA typing of *S.* Enteritidis and correctly assigned allelic profiles for the two repeat isolates. After not being able to assign correct allelic profile for one of the repeat test isolates in EQA-8 and EQA-9, laboratory 147 correctly identified both the isolates in EQA-10.

Figure 6. Correct MLVA typing of two repeat *S.* Enteritidis isolates from EQA-8 to 10 (for laboratories participating in EQA-9 and/or EQA-10)



Arbitrary numbers represent the participating laboratories. Bars represent the number of correctly assigned allelic profiles for the two repeat isolates (SE9 and SE10).

3.3 Molecular typing-based cluster analysis

In this part of the EQA, the participants should correctly identify a cluster of closely related isolates among 10 test isolates by using either PFGE, MLVA, and/or WGS-derived data. The cluster of four test isolates were pre-categorised by the EQA provider.

The expected cluster of closely related monophasic *S.* Typhimurium ST34 isolates contained four isolates based on WGS derived data. REF6 was pre-defined outside the expected cluster as it differed with six alleles or 10 SNPs to the expected cluster, but the EQA provider accepted participants inclusion of REF6 as an acknowledgment that a definitive cut-off in *Salmonella* has not been formally established. The cluster MLVA profile was 3-15-11-NA-211. The characteristics of the test isolates and reported results are listed in Annexes 4-15.

3.3.1 PFGE-derived data

Seven (7/19, 37%) participants performed cluster analysis using PFGE-derived data. The cluster categorisation was based on WGS data and therefore as expected the correct cluster delineation was difficult to obtain by the use of less discriminatory methods, e.g. PFGE. All seven participants included the four (or five if including REF6) WGS cluster isolates as a part of the correct cluster, however in addition laboratory 144 included REF7 (ST4430) and REF8 (ST34).

Table 3 shows the overview of the cluster analyses of the isolates each participant included or excluded in their cluster.

Table 3. Results of cluster analyses based on PFGE-derived data

		Laboratory						
Isolate	**ST**	**5**	**9**	**9**	**106**	**127**	**14**	**144**
REF1‡	34	Ye	Ye	Ye	Yes	Yes	Yes	Yes
REF2	4431	No	No	No	No	No	No	No
REF3‡#	34	Ye	Ye	Ye	Yes	Yes	Yes	Yes
REF4	5296	No	No	No	No	No	No	No
REF5	34	No	No	No	No	No	No	No
REF6(¤)	34	No	No	No	Yes	No	No	Yes
REF7	4430	No	No	No	No	No	No	Yes
REF8‡	34	Ye	Ye	Ye	Yes	Yes	Yes	Yes
REF9	34	No	No	No	No	No	No	No
REF10 ‡#	34	Ye	Ye	Ye	Yes	Yes	Yes	Yes
Correct cluster ‡		**Ye**	**Ye**	**Ye**	**(Yes)**	**Yes**	**Yes**	**No**
Included the five WGS cluster		**Ye**	**Ye**	**Ye**	**Yes**	**Yes**	**Yes**	**Yes**

‡: closely related isolates based on WGS (in grey)
(¤): Accepted as closely related isolate based on WGS
#: technical duplicate isolates (in bold)

3.3.2 MLVA-derived data

Five participants (26%) performed cluster analysis using MLVA-derived data, and all selected the *S.* Typhimurium scheme and reported the loci in the correct order: STTR9, STTR5, STTR6, STTR10 and STTR3.

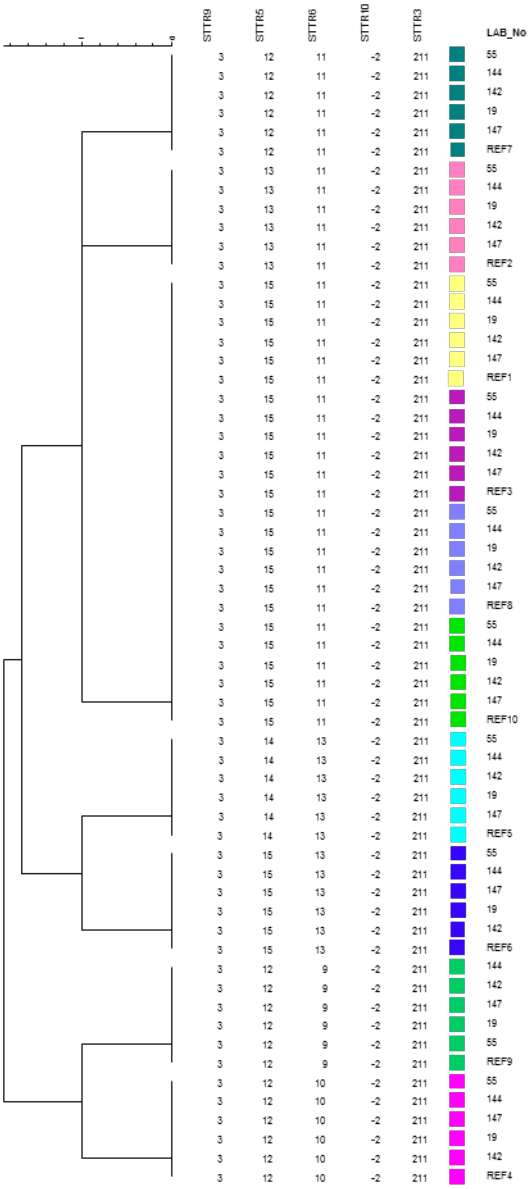
Performance was high, as all participants were able to identify the correct cluster of closely related isolates (MLVA profile: 3-15-11-NA-211) defined by a pre-categorisation based on WGS by the EQA provider among the 10 cluster test isolates. None of the laboratories included REF6 (3-15-13-NA-211) or other isolates with one locus-variation (REF2 and RE7) in their cluster of closely related isolates. Table 4 shows the overview of the isolates each participant included 'Yes' and excluded 'No' in their cluster analysis. Figure 7 shows a dendrogram of the reported MLVA results. All laboratories reported the correct MLVA profile for all 10 test isolates in the cluster analyses based on MLVA-derived data. All data are available in Annexes 9 and 10.

Table 4. Results of cluster analyses based on MLVA-derived data

										Laboratory ID				
Isolate number	**ST**	**MLVA-profile**								**19**	**55**	**142**	**144**	**147**
REF1‡	34	3	-	15	-	11	-	NA	-	211	Yes	Yes	Yes	Yes	Yes
REF2	4431	3	-	13	-	11	-	NA	-	211	No	No	No	No	No
REF3‡#	34	3	-	15	-	11	-	NA	-	211	Yes	Yes	Yes	Yes	Yes
REF4	5296	3	-	12	-	10	-	NA	-	211	No	No	No	No	No
REF5	34	3	-	14	-	13	-	NA	-	211	No	No	No	No	No
REF6	34	3	-	15	-	13	-	NA	-	211	No	No	No	No	No
REF7	4430	3	-	12	-	11	-	NA	-	211	No	No	No	No	No
REF8‡	34	3	-	15	-	11	-	NA	-	211	Yes	Yes	Yes	Yes	Yes
REF9	34	3	-	12	-	9	-	NA	-	211	No	No	No	No	No
REF10‡#	34	3	-	15	-	11	-	NA	-	211	Yes	Yes	Yes	Yes	Yes
Cluster-identified		3	-	15	-	11	-	NA	-	211	**Yes**	**Yes**	**Yes**	**Yes**	**Yes**

‡: closely related isolates derived from WGS (in grey)
#: technical duplicate isolates (in bold)
NA: designates a locus not present (-2 by submission, Annex 10).

Figure 7. Reported MLVA results of each test isolate



Dendrogram from BioNumerics of MLVA profiles reported by the laboratories. Each of the 10 test isolates has a different colour. REF1 to REF10: results from EQA provider.

3.3.3 WGS-derived data

Reported results from participants

Fifteen participants (79%) performed cluster analysis using WGS-derived data. Only one laboratory reported using external assistance for sequencing. Different sequencing platforms were listed among the participants: one MiniSeq, eight MiSeq, one HiSeq, three NextSeq, and one Ion Torrent. All reported using commercial kits for library preparation. Of the 15 participants, 12 (80%) used Illumina's Nextera kit. One laboratory reported volume changes from the manufacturer's protocol (Annex 8).

Performance was high in cluster analysis with WGS-derived data, with 14 (93%) participants correctly identifying the cluster of closely related isolates defined by a pre-categorisation by the EQA provider among the 10 test isolates, when REF6 could be included in the cluster definition. Over half (9/15, 60%) of the participants did not include REF6 in the cluster, which is in accordance with the "original" pre-defined cluster by the provider.

Thirteen laboratories correctly reported ST of all 10 isolates. Based on the results and sequences provided by Laboratory 108, a possible swapping of REF3 and REF4 must have occurred, meaning that the MLST results are also incorrect. In addition, laboratory 148 used the submission field for a number of six digits, without any explanations, but probably Enterobase [5] level cluster codes.

Table 5. Results of cluster analyses based on WGS-derived data

		Laboratory ID														
Isolate number	**ST**	**19**	**36**	**49**	**100**	**106**	**108**	**127**	**129**	**134**	**135**	**142**	**147**	**148**	**149**	**150**
REF1‡	34	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
REF2	4431	No	No	No	No	No	No	No	No	No	No	No	No	No	No	No
REF3‡#	34	Yes	Yes	Yes	Yes	Yes	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
REF4	5296	No	No	No	No	No	Yes	No	No	No	No	No	No	No	No	No
REF5	34	No	No	No	No	No	No	No	No	No	No	No	No	No	No	No
REF6(¤)	34	No	No	No	No	Yes	No	No	Yes	Yes	Yes	No	Yes	No	Yes	No
REF7	4430	No	No	No	No	No	No	No	No	No	No	No	No	No	No	No
REF8‡	34	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
REF9	34	No	No	No	No	No	No	No	No	No	No	No	No	No	No	No
REF10‡#	34	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Main analysis		Allele	Allele	Allele	SNP	Allele	SNP	Allele	Allele	Allele	Allele	Allele	Allele	Allele	Allele	SNP
Additional analysis 1		SNP	SNP	Allele	Allele			Allele								
Additional analysis 2.				SNP	SNV			SNP								
Cluster identified		**Yes**	**Yes**	**Yes**	**Yes**	**(Yes)**	**No**	**Yes**	**(Yes)**	**(Yes)**	**(Yes)**	**Yes**	**(Yes)**	**Yes**	**(Yes)**	**Yes**

‡: closely related isolates (in grey)
(¤): accepted as closely related isolate
#: technical duplicate isolates (in bold)
"cgMLST" / "wgMLST": assigned by provider based on reported information (Table 7)
ST: sequence type
Allele: allele-based analysis
SNP: single-nucleotide polymorphism analysis
SNV: single-nucleotide variant based on cgMLST.
(Yes): laboratories including REF6 in the identified cluster.

Laboratories were instructed to report the data used for cluster analyses and select a representative isolate in the cluster for reporting SNP distance or allelic differences between the selected isolate and each test isolate included in the analysis. Laboratories could report results from up to three analyses (one main and 1 to 2 additional), but the detected cluster had to be based on results from the main analysis. Laboratories 19, 36, 49, 100 and 127 reported additional analyses.

Of the seven participants using SNP analysis, three (laboratories 100, 108, and 150) used SNP as the main analysis for cluster detection, two (laboratories 19 and 36) reported SNP as an additional analysis, and laboratories 49 and 127 reported SNP-based analyses as a third analysis. All used a reference-based approach with different *S.* Typhimurium isolates as reference. Three used an in-house pipeline to the SNP analysis, and two reported use of BioNumerics one NASP and one Enterobase. As read mapper, three used Burrows-Wheeler Aligner (BWA), one Bowtie, one BioNumerics, one Enterobase, and one used CLC. Two laboratories reported the use of GATK as variant caller, one VarScan, one BioNumerics, one Enterobase, and CLC was also used. One laboratory used no variant caller.

Tables 6 and 7 show the overview of the submitted data. For laboratory-reported SNP distance/allelic differences by isolate, see Annex 12.

Table 6. Reported results of SNP-based cluster analysis

	SNP-based analysis							
Laboratory	**SNP pipeline**	**Approach**	**Reference**	**Read mapper**	**Variant caller**	**Distance within cluster**	**REF6 included (yes or no) SNP to cluster**	**Distance outside cluster**
Provider	NASP [8]	Reference-based	REF3	BWA	GATK	0-1	No / 10	46-172
Provider	NASP [8] + recombination filter [9]	Reference-based	REF3	BWA	GATK	0-1	No / 10	46-113
19*	NASP	Reference based	ST34 (ID 5072)	BWA	GATK	0-1	No / 11	41-97
36*	in-house pipeline	Reference based	NC_003197.1 (S. Typhimurium LT2)	BWA	VarScan	0-1	No / 9	36-94
49#	BioNumerics	Reference based	ST34 5349	BioNumerics	BioNumerics	0-1	No / 12	43-275
100	wgSNP BioNumerics	Reference based	ST2, NC_003198.1	Bowtie2	-	0-5	No / 21	61-145
108	Inhouse	Reference based	in-house strain resp ST	CLC assembly cell v.4.4.2	CLC assembly cell v.4.4.2	0-1	No / 11	55-109
127#	EnteroBase SNP pipeline	Reference based	5111	EnteroBase SNP pipeline (refMapper)	EnteroBase SNP pipeline (refMapper)	0-1	No / 12	39-138
150	sNAPPER DB in house	Reference based	AE006468	BWA	GATK	0-1	No / 11	40-125

**: additional analysis 1*
#: additional analysis 2
Detailed data, see Annex 12.

Of the 13 participants using an allele-based analysis, 12 selected the method as the main analysis for cluster detection. Six of 12 (50%) used only an assembly-based allele calling method and five (42%) used both assembly- and mapping-based allele calling methods. One laboratory (8%) used only a mapping-based allele calling method. As additional analysis, two participants used both assembly- and mapping-based allele calling methods and one participant reported two allele calling methods; one using only an assembly-based allele calling method and the other without any assembly- or mapping-based method i.e single-nucleotide variant based on cgMLST (SNV).

Seven of the main analysis used SPAdes as the assembler and three used Velvet. One laboratory used BWA. Of the additional analysis two used SPAdes and one Velvet. Eleven of the 12 main analysis used a cgMLST scheme for the allele-based analysis and one reported a modified scheme of 3505 loci. Of the additional analysis one used wgMLST (15874 loci), and two used cgMLST scheme and one used SNV method based on cgMLST results.

Ten of 12 laboratories (main analysis) reporting allele analyses used Enterobase (cgMLST) as the scheme for analysis. Two laboratories (106 and 129) reported the use of cgMLST in an ad hoc scheme for *Salmonella* enterica based on 3009 and 3505 core loci respectively. In the additional analysis, Laboratory 49 used wgMLST scheme of 15874 loci and two laboratories used cgMLST scheme. Furthermore, laboratory 100 also reported an additional analysis of SNV based on the cgMLST scheme.

Table 7. Reported results of allele-based cluster analysis

	Allele-based analysis							
Laboratory	Approach	Allelic calling method	Assembler	Scheme	Number of loci	Difference within cluster	REF6 included (yes or no)/ differences to cluster	Difference outside cluster
Provider	BioNumerics	Assembly- and mapping-based	SPAdes	Applied Maths (cgMLST/Enterobase)	3002	0-1	No / 6	17-52
19	BioNumerics	Assembly based and mapping based	SPAdes	Applied Math (cgMLST/Enterobase)	3002	0-1	No / 6	17-52
36	SeqPhere	Only mapping based	-	Enterobase (cgMLST)	3002	0-2	No / 6	19-52
49	BioNumerics	Assembly based and mapping based	SPAdes	Applied Math (cgMLST/Enterobase)	3002	0-1	No / 6	17-53
49*	BioNumerics	Assembly based and mapping based	SPAdes	Applied Math (wgMLST)	15874	0-1	No / 9	35-88
100*	SeqPhere	Only assembly based	Velvet	Enterobase (cgMLST)	3002	0-1	No / 6	18-69
100*	SNV	-	-	SNVs were calculated based on cgMLST scheme	3002	0-1	No / 6	19-69
106	SeqPhere	Assembly based and mapping based	BWA	Other	3009 (2977 with no missing values) "cgMLST"	0-1	Yes / 6	19-52
127	Enterobase	Assembly based and mapping based	EnteroBase QAssembly (SPAdes)	Enterobase (cgMLST)	3002	0-1	No / 7	18-67
127*	BioNumerics	Assembly based and mapping based	SPAdes	Applied Math (cgMLST/Enterobase)	3002	0-4	No / 6	20-48
129	SeqPhere	Only assembly based	Velvet	Other	§Ad hoc scheme: 3505 cgMLST + 839 accessory ("wgMLST")	0-1	Yes / 9	27-67
134	SeqPhere	Assembly based and mapping based	Velvet	Enterobase (cgMLST)	3002	0-1	Yes / 6	18-51
135	SeqPhere	Only assembly based	SPAdes	Enterobase (cgMLST)	3002	0-1	Yes / 6	18-65
142	BioNumerics	Only assembly based	SPAdes	Enterobase (cgMLST)	3002	0-2	No / 19	31-72
147	SeqPhere	Only assembly based	SPAdes 3.11.1	Enterobase (cgMLST)	3002	0-1	Yes / 6	19-52
148	Enterobase	Only assembly based	SPAdes	Enterobase (cgMLST)	3003	0-1	No / 6	19-52
149	SeqPhere	Only assembly based	Velvet 1.1.04 (optimised k)	Enterobase (cgMLST)	3002	0-1	Yes / 6	19-51

**: additional analysis*
§: modified from submitted information
For detailed data, see Annex 12.

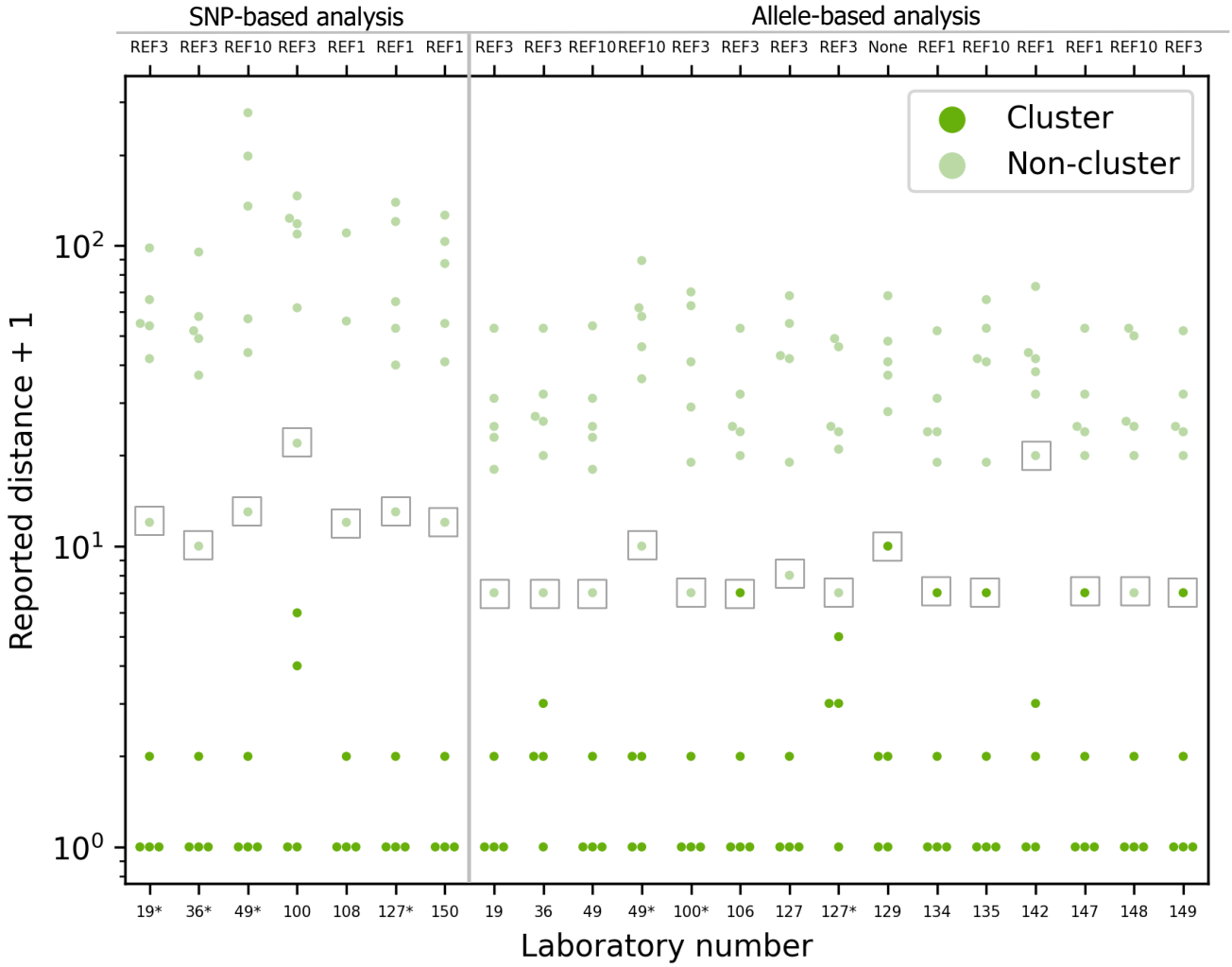
All 12 laboratories (100%) using an allele-based analysis as the main method could identify the correct cluster of the five closely related ST34 isolates (Figure 8). All these laboratories reported 0-2 allele differences in the cluster without REF6 and nine laboratories reported 6 allele differences for REF6 and three reported between 7-19 allele differences for REF6 (Table 7).

Six laboratories (6/12,50%) included REF6 in the closely related isolates in the main analysis using allele-based method. Five of these six laboratories accepted six allele differences in the cluster. One laboratory included nine allelic differences.

Furthermore, three additional analyses by two laboratories (49, 100) reported 0-1 allele difference within the cluster using wgMLST, cgMLST or SNV, and laboratory 127 reported 0-4 without REF6.

Two other test isolates (REF5 and REF9) were also ST34, and additional three were ST4430, ST4431 and ST5296 respectively. The laboratories reported allele differences to the selected cluster isolate at 17-88 for this group of isolates (difference outside cluster) by all schemes used.

Figure 8. Reported SNP distances or allele differences for each test isolate to selected cluster representative isolate



**: additional analysis*
SNP: single-nucleotide polymorphism analysis
Selected cluster representative marked as REF
Dark green: reported cluster of closely related isolates
Light green: reported not part of cluster.
SNV based results from 100 laboratory not shown
Grey box around REF6 isolate

Of the three laboratories (100, 108 and 150) performing SNP analysis as main analysis, two identified the correct cluster of closely related isolates, although laboratory 108 most likely swapped REF3 and REF4. (Table 5). Four other laboratories (19, 36, 49 and 127) that identified the correct cluster of closely related isolates by cgMLST (main analysis) also performed SNP analysis as additional first or second analysis.

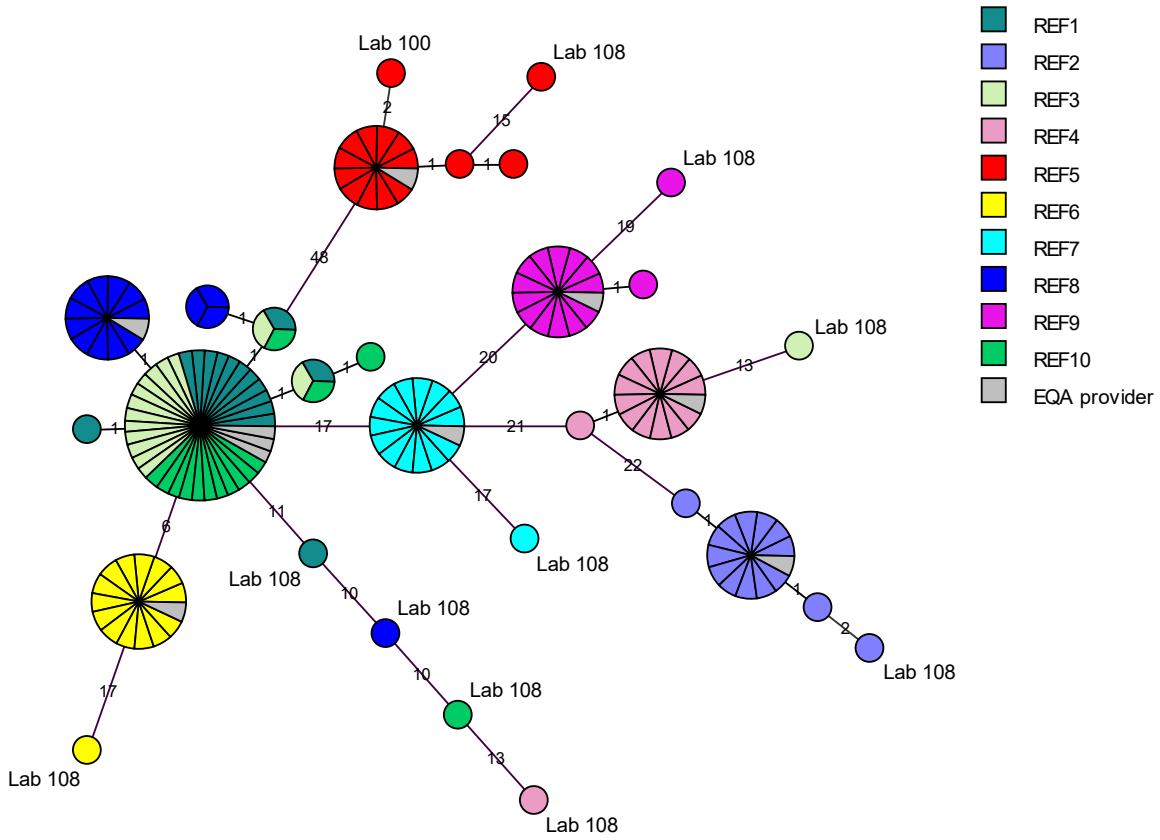
The reported SNP distances within the cluster were 0-1 for most (6/7) of the analysis. Laboratory 100 reported the SNP distances within the cluster as 0–5 using a wgSNP. Laboratory 100 reported both larger distances within the cluster and to REF6 (21 SNPs) than the other laboratories. REF6 was not included in the cluster by any of the laboratories using SNP analysis as the main or the additional analysis. The EQA provider performed two SNP analyses both with and without recombination filter (Table 6) and no areas were highly affected by the removal of recombination.

Analysis of raw reads uploaded by participants

In addition to the reported cluster analyses, participants submitted their FASTQ files to be evaluated by the EQA provider. The data were initially evaluated using the EQA provider's QC pipeline [10] and FASTQ files were uploaded to an Applied Maths calculation engine for allele calling (cgMLST/Enterobase, [7]).

The overall cgMLST analysis by the provider, shown in the minimum spanning tree (MST, Figure 9) and based on submitted raw reads from 15 laboratories, shows clear clustering of the results for each test isolate. Only data from laboratory 108 are separated (or removed) from the other results as they have apparently swapped REF3 and REF4.

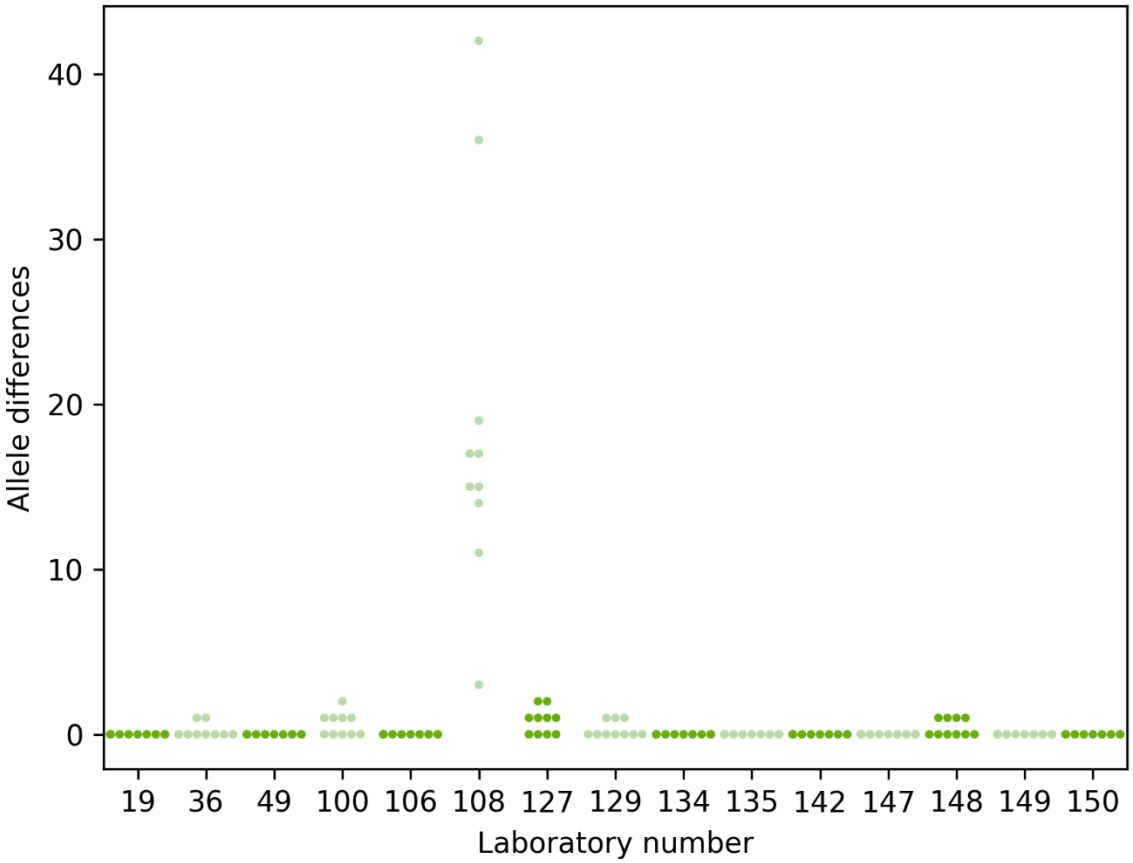
Figure 9. Minimum spanning tree of core genome multilocus sequence typing, participant FASTQ files



Minimum spanning tree (MST) in log scale of core genome multilocus sequence typing (cgMLST, [7]) based on submitted raw reads (FASTQ files).
Each REF1–REF10 test isolate has a different colour.
REF results from the EQA provider are in grey.

For each laboratory, cgMLST was performed on the submitted raw reads (FASTQ files) applying Applied Maths allele calling with the Enterobase scheme [7]. For each laboratory, a hierarchical single linkage clustering was performed on the submitted data along with the EQA provider's reference isolates. Figure 10 shows the allele differences between each submitted sequence and the corresponding reference. As seen in Figure 10, laboratory 108 is the only laboratory with many isolates with large differences from the reference and results of the other participants. Laboratory 108 also have a swap of isolates (REF3 and REF4).

Figure 10. Participant allele difference from reference result (EQA provider) for each test isolate



Allele difference of participant isolates from the corresponding REF isolates (EQA-provider) based on the submitted raw reads (FASTQ files). For 137 of 150 results (91%), 0-1 differences were identified (Figure 10). For three results, a difference of two alleles from the REF isolate was calculated. For 10 results (7%), a difference of 3-42 alleles was seen, all reported by laboratory 108. Excluding the results from the suggested mistake of swapping REF3 and REF4 the reaming difference is from 3-19 still higher than other laboratories and is the results of Ion Torrent data in analyzed in BioNumerics.

Separately, the laboratories responded to QC parameters used to evaluate their data. As seen in Table 8, both coverage and genome size were the most widely used QC parameters with 93% and 73% of the laboratories using this parameters. Different thresholds of coverage, ranging between 20-50X coverage were used. A contamination check was reported by 67% of laboratories using different programs. The number of good cgMLST loci was used by 60% of laboratories, with a threshold ranging between 90-99% reported by most. One laboratory reported using only 80% as a threshold. Q score was used by 53%. A few laboratories reported additional parameters, including laboratory 150, which in the previous three EQAs reported a larger number of QC parameters; see Annex 13 and, for the full QC evaluation of all isolates, see Annex 14.

Table 8. Summary of selected QC parameters reported by participants

Laboratory	Confirmation of genus	Coverage	Q score (Phred)	Genome size	No. of good cgMLST loci
19	Kraken and < 5% contamination with others	Min. x25	N50 value and number of contigs - No threshold	4.5 - 5.1 Mb	Min. 98% core percent and max. 15 loci with multiple consensus - No actual threshold employed on regular basis for either
36	No	30x	No	No	No
49	No	<25= Fail, 25-29 = Warn, >=30 = Pass	N50 >100000	4.5 to 5.5 MB	Core % Pass >=97% Warn 90-96%, Fail <90%
100	KmerFinder 3.1 - Center for Genomic Epidemiology	40x	FastQC	app. 5 Mb	No
106	kmerfinder	50x	FastQC	contig length 200 bp	No
108	Assembled genomes	>20x	Total match size, similarity, CDS covered	CDS covered	No
127	EnteroBase QAssembly pipeline (Kraken)	threshold >=25	No	Bionumerics de novo assebly pipeline sequence length, 3.6-6.0 Mb	Bionumerics Summary calls % of core present, > 80%
129	SeqSero 1.0	>29	No	No	No
134	Mach in SeqSphere	50 X but if it's less, the number of targets found should be > 90-95%	No	length of contigs assembled < ref genome + 10%	cgMLST found and called > 90-95%
135	No	>30	we use Quast to evaluate assambly quality	Also with QUAST	Also with QUAST
142	Kraken	30x	No	No	95% called
147	JSpecies	~ 30x coverage (min.)	No	~ 5Mbases for Salmonella	percentage good targets ~ 98%
148	Kraken (Enterobase): >70% contigs assigned	> 45x	Number of contigs < 250 (alert in house) and <500 (min. acceptance for Enterobase)	Number of bases between 4 - 5.1 Mbp (alert in house) and 4 - 5.8 Mbp (acceptance for Enterobase)	BWA back-mapping the reads, and Samtools (with BCFtools) for variant calling.
149	Kraken	Average Coverage (Assembled) > 50	N50	Consensus base count	>90 % good targets cgMLST
150	No	No	No	No	No
% of laboratories using QC parameter	**67%**	**93%**	**53%**	**73%**	**60%**

For each laboratory, the submitted raw reads (FASTQ files) were evaluated by the EQA provider's in-house quality control pipeline [10]. Table 9 shows the QC parameters and range of QC values per laboratory. For the full QC evaluation of all isolates, see Annex 14. Overall, the coverage of the raw data was sufficiently high when evaluated by the EQA provider's QC pipeline, a few isolates of different participants had a average coverage below the generally accepted threshold of 50, some as low as 26.

Table 9. Results of participants' raw sequence data evaluated by EQA provider's QC pipeline

		Laboratory No.														
Parameters	Ranges	19	36	49	100	106	108	127	129	134	135	142	147	148	149	150
Detected species	*{Se}*	Se	Se	Se	Se	Se	Se	Se	Se	Se	Se	Se	Se	Se	Se	Se
% Species 1		94.1-96.1	94.0-98.0	94.7-98.7	96.7-98.8	91.9-94.9	95.2-98.2	95.8-98.4	95.6-97.6	96.0-98.2	94.9-97.9	95.5-98.2	93.2-97.8	96.5-98.3	94.5-97.3	92.7-94.6
% Species 2	*{<5%}*	0.0-1.5	0.0-1.7	0.0-1.6	0.0-1.7	0.0-1.3	0.4-2.4	0.0-1.1	0.0-1.3	0.1-1.3	0.0-1.4	0.1-2.0	0.1-1.8	0.0-1.2	0.3-1.2	0.0-1.4
Unclassified reads (%)		3.8-5.5	1.9-5.9	1.2-5.3	1.1-3.0	5.0-7.2	1.4-2.4	1.6-3.8	2.3-3.6	1.4-3.7	2.1-5.0	1.6-3.9	2.0-6.4	1.6-2.8	2.3-4.2	5.3-7.2
Length at >25 x min. coverage (Mbp)	{>4.5 ∧ <5.3}	4.8-5.0	4.7-4.9	3.1-5.0	4.9-5.0	4.9-5.0	4.7-4.9	2.6-4.9	4.8-5.0	4.9-5.0	4.9-5.0	4.9-5.0	4.9-5.0	4.8-5.0	4.9-5.0	4.8-5.0
Length [1-25] x min. coverage (kbp)	{<0.25}	2.2-8.4	11.9-115.7	0.0-1824.4	0.0-0.0	0.0-0.5	0.0-3.0	0.0-2263.9	0.7-10.9	0.0-0.0	0.0-0.0	0.0-5.9	0.0-1.4	0.7-104.4	0.0-45.2	0.0-4.8
No. of contigs at 25 x min. coverage	{>0}	116-190	176-260	40-79	44-71	56-109	#666-2397	55-102	78-185	62-95	45-88	44-73	54-96	98-227	46-77	70-115
No. of contigs [1-25] x min. coverage	{<1 000}	3-10	11-61	0-39	0-0	0-1	#0-14	0-83	1-10	0-0	0-0	0-8	0-1	1-66	0-6	0-4
Average coverage	{>50}	102-153	44-109	26-104	51-129	202-287	85-174	26-100	78-148	42-66	107-266	45-79	67-93	79-259	78-112	53-83
No. of reads (x 1000)		3528-5325	796-1875	553-2004	1066-2682	7473-9951	1363-2718	859-3377	2670-5037	1383-2197	3565-8859	928-1728	1311-1964	2599-8421	1478-2262	2642-4125
Average read length		143-145	258-290	209-282	237-243	129-148	291-318	149-150	146-148	151-151	149-150	230-245	238-272	149-150	253-271	99
Average insert size		296-334	528-651	234-409	324-423	215-432	NA	492-608	352-417	320-364	369-433	343-468	282-382	367-415	307-373	361-440
N50 (kbp)		47-83	28-46	143-319	150-319	113-283	#3-13	63-239	47-186	144-283	135-283	178-319	105-283	37-94	144-316	107-192

**: indicative QC range*
Se: Salmonella *enterica*
NA: not analysed
#: QC values unreliable due to assembly issues for Ion Torrent data.

Assessment of the provided genomes

The four provided genomes should individually be assessed and compared with the already produced data in the cluster analysis, and the participants had to determine whether the genomes were part of the defined cluster or not.

The participants were instructed to describe their observations and considerations leading to the decision. The EQA-provider had manipulated the raw reads. The four genomes represented raw reads of two cluster isolates with high-quality raw reads or reduced coverage, and two non-cluster isolates with high-quality raw reads or contamination (table 10). Raw data can be seen in Annex 15.

For genome 1 (good quality), 100% (all 15 laboratories) accepted the quality of the genome and correctly identified genome 1 as a cluster isolate. Three laboratories performed SNP-analyses and one reported a distance of 3 SNP for genome 1 to the cluster; the two other laboratories found 0 SNP. Of the 12 laboratories that used allele-based analyses, most found 0-2 allele differences. Two laboratories (106 and 127) reported respectively less than seven and less than five allele differences of genome 1 to the cluster. Two other laboratories (129 and 142) reported a difference of eight and seven alleles respectively, but still included genome 1 in the cluster.

For genome 2 (low coverage), 87% (13/15) correctly observed poor quality in genome 2, nine used the low average coverage to disregard the genome, and other laboratories used different QC parameters like threshold of >90% good cgMLST targets or number of contigs.

Only two of the 15 laboratories did not describe the detection of low quality, however the two laboratories were among the six laboratories, which proceeded with the analysis, and concluded that the isolate was not a part of the cluster. Additional nine laboratories did not attempt to perform the cluster analysis, as the quality was too low.

For genome 3 (good quality), 93% (14/15) accepted the quality of the genome and correctly described the genome as a non-cluster isolate and not a part of the cluster of closely related isolates. One laboratory (106) incorrectly reported the genome 3 as a cluster isolate, and states they found "one allele difference with the representative genome".

For genome 4 (good quality but contaminated), 87% (13/15) correctly described the genome as a non-cluster isolate, however only 33% (5/15) correctly described contamination present in genome 4. Two of the five described the added species *E. coli* and one laboratory (19) of the five did not perform the analysis because of the contamination. One laboratory (106) incorrectly reported the genome 4 as a cluster isolate.

Table 10. Results of the participants' assessment of the EQA provided genomes

Genome	Characteristics	Characteristics identified by participants	Yes	No	Not analysed
1	A cluster isolate (REF11), good quality of reads, 0 allelic difference to the REF1, REF3 and REF10	Quality accepted	15	0	0
		Suggested to be a cluster isolate	15	0	0
2	A cluster isolate (REF12) with altered coverage (reduced to 10x)	Poor quality was observed	13	2	0
		Suggested to be a cluster isolate	0	6	9
3	A non-cluster isolate (REF13). 31 allelic difference to the cluster, good quality of reads	Quality accepted	15	0	0
		Suggested to be a cluster isolate	1	14	0
4	A non-custer isolate (REF14) mixed with a *Escherichia coli* (approx. 10%)	Contamination was observed	5	10	0
		Suggested to be a cluster isolate	1	13	1

Annex 15

4 Discussion

Overall, the total number of participants has decreased over time. From 26 in EQA-7, 24 in EQA-8, 23 in EQA-9 to 19 in EQA-10. Among the 19 participants, one laboratory (106), which was not participating in EQA-9, participated again in EQA-10. Five laboratories (128, 132 138, 140 and 145) participated in both EQA-8 and EQA-9 or only EQA-9, but four did not participate in EQA-10 and the last one accepted the invitation but did not submit any results in EQA-10. All these five laboratories used PFGE as typing method when participating, and the change in the structure of the EQA with less focus on PFGE analyses may have caused the absence of participation.

4.1 Multiple-locus variable number of tandem repeats analysis

Nine laboratories (47%) participated in the MLVA part, where eight laboratories performed the analysis for both *S.* Typhimurium and *S.* Enteritidis. One laboratory (92) only did the MLVA analysis for *S.* Enteritidis. MLVA for *S.* Enteritidis was included in the EQA for the third time, and the relatively high number of participants confirms that it was relevant to include this method, although one laboratory fewer participated in this EQA-10. In MLVA for *S.* Typhimurium, the number of participants again was lower than in previous years, decreasing from 15 participants in EQA-4 to 10 participants in EQA-9 and to eight in this EQA-10. This can reflect a trend, where more laboratories are switching to WGS-based surveillance and outbreak detection using WGS instead of MLVA.

Five of the eight laboratories (63%) obtained a total score of 100% for *S.* Typhimurium and reported the correct MLVA types for all 10 test isolates. The overall performance in this round was 93%, which was higher than last year and but lower compared with previous years. From EQA-4 to EQA-10, the overall performance in each round was 92%, 96%, 96%, 96%, 97%, 86% and 93%, respectively. As in EQA-9, one laboratory (laboratory 55) caused a large part (50%) of the incorrect results.

The MLVA results of the two repeat *S.* Typhimurium isolates from EQA-4 through EQA-10 showed good performance by the participants. When observing the laboratories either participating in EQA-9 and/or EQA-10, the majority of participants (80%; 8/10) performed better than the last time they participated and this year all participants correctly identified the two repeat isolates. Two laboratories that only correctly identified one of the two repeat isolates in EQA-9 did not participate in EQA-10.

Errors in the MLVA for *S.* Typhimurium were mainly caused by reporting alleles in a locus with no fragment present, but also by assigning an incorrect allele in a present fragment. No common characteristics of the isolates caused problems among the participants, but laboratory 55 seemed to make the same kind of errors – reporting alleles in a locus with no fragment present.

For MLVA of *S.* Enteritidis, seven laboratories (78%) obtained a total score of 100% and the overall performance was 93% which was a bit lower compared to EQA-9 (98%), however near the same as EQA-8 (92%). The mistakes in the MLVA for *S.* Enteritidis are reported by only two laboratories, one laboratory (144) has probably swapped of the isolates/results of SE6 and SE10. The remaining four errors are all reported by the same laboratory (55); incorrect fragments sizes in SENT4 (x3) and one time in SENT5.

The MLVA results of the two repeat *S.* Enteritidis isolates from EQA-8 through EQA-10 showed good performance by the participants. When observing the laboratories either participating in EQA-9 and/or EQA-10, the majority of participants (91%; 10/11) performed better than the last time they participated. Two laboratories, which correctly identified the two repeat isolates in EQA-9, did however not participate in EQA-10. Only one laboratory (144) did not identify one of the repeat isolates as they previous did in EQA-9, (based on the submitted results the EQA provider suspect the laboratory must have swapped two isolates/results for SE6 and SE10, where SE10 is a repeat isolate).

Despite a better performance for *S.* Typhimurium MLVA this year, in general errors occurred may suggest a decreased use of the MLVA method combined with reduced maintenance of quality and skills while implementing WGS as routine method.

4.2 Molecular typing-based cluster analysis

In the present EQA scheme, a molecular typing-based cluster analysis was included for the third time. Participants were again free to choose their preferred method between PFGE, MLVA and/or WGS-derived data, and the identified cluster depended on the method used; in this EQA, the cluster categorisation and the evaluation was entirely based on WGS data.

The expected cluster (of closely related monophasic *S.* Typhimurium ST34 isolates) contained four isolates in the pre-defined cluster and REF6 was categorised outside the expected cluster, as it differed with six alleles or 10 SNPs. This definition was based on the clonal nature of monophasic *S.* Typhimurium and the knowledge about outbreak investigation of this serotype ([11-15], where unrelated isolates potentially can be within six allelic differences.

However, the EQA provider accepted in the individual evaluation inclusion of REF6, as an acknowledgment that a definitive cut-off in Salmonella has not been formally established.

For the first time, all the laboratories in the EQA participated in the molecular typing-based cluster analysis using either PFGE and/or MLVA and/or WGS. Four new laboratories participated in the cluster part using WGS and only one laboratory performing WGS last year, did not report WGS based results this year. The number of participants only performing WGS-based cluster analysis increased from 7 to 8 in EQA-9 to 10 in EQA-10.

The number of laboratories only performing PFGE-based cluster analysis, which increased markedly from two to nine in EQA-9, has decreased again to only two in EQA-10. Three of the five not participating in EQA-10 did not identify the correct cluster in EQA-9 using PFGE. It is unfortunate if the absence in participation is entirely due to the lack of possibility of analysing by WGS, as PFGE can still have a value for investigating outbreaks at the national level. On the other hand, it is very inspiring that more laboratories have begun to perform the cluster analysis by WGS.

4.2.1 PFGE-derived data

Of the 19 laboratories, seven (37%) performed cluster analysis using PFGE-derived data. Six laboratories (86%) correctly identified the cluster, if the broader cluster definition of five isolates (including REF6) is used. The two laboratories only performing PFGE identified also the correct cluster and showed usefulness of the method, particularly if the laboratory have high performance and experience in the method.

One laboratory did not correctly identify the cluster using PFGE and included REF7 besides REF6. This laboratory did not perform WGS but identified the correct cluster by MLVA, demonstrating that using more than one typing method can sometimes be appropriate in an outbreak investigation.

However, difficulties in inter-laboratory comparability occurs when different methods are used. Few countries still use PFGE and the method can still have a value for investigating outbreaks at the national level, and it can support bridging the historical national databases from human and veterinary sector for case finding and hypothesis generation by WGS.

4.2.2 MLVA-derived data

In total, five laboratories performed cluster analysis using MLVA-derived data and this is an increase compared to EQA-8 and EQA-9. As in the two previous years, no laboratories were only using MLVA for the cluster analysis; two also used PFGE, another two also used WGS and one used both PFGE and WGS together with MLVA.

Performance was high, as all of the five participating laboratories correctly identified the cluster of four closely related isolates using MLVA-derived data, none of the laboratories included REF6, which had a variation of two repeats in one of the fast-changing locus (STTR6). Furthermore, no laboratories included REF2 or REF7, which both had one locus-variation in the other fast-changing locus (STTR5). Despite none of laboratories solely used MLVA for the cluster analysis, the method seems very suitable for identification of the cluster in this EQA. One laboratory (147) performed both MLVA and WGS and did not submit identically clusters by the two methods, by WGS they included REF6, but not by MLVA and as already mentioned another laboratory (144) identified the correct cluster by MLVA but not by PFGE.

4.2.3 WGS-derived data

Fifteen of the 19 laboratories (79%) performed cluster analysis using WGS-derived data. This was a higher participation compared to EQA-9, where 12 of 23 laboratories (52%) performed cluster analysis using WGS-derived data. Nevertheless, a decrease in number of participants in the EQA in total. Performance was again high, as 14 (93%) correctly identified the cluster of closely related isolates, when REF6 is accepted as a part of the cluster. Only one laboratory was not able to identify the correct cluster and probably just because the laboratory had swapped REF3 and REF4. Large error was identified in the submission by laboratory 36 and 129. All 10 results/isolates were swapped, but the two participants were allowed to re-submit the corresponding isolate IDs, and the re-arranged data is included in this report. Two other laboratories (129 and 142) who in EQA-9 did not identify the correct cluster succeed in EQA-10, as the they both identified the correct cluster using the same method/scheme as in previous EQAs.

Most laboratories (14/15) reported the use of an Illumina platform, and all reported using commercial kits for library preparation. Only one laboratory reported the use of external assistance for sequencing, which is the same compared to EQA-9.

Twelve laboratories (80%) reported using an allele-based method as the main analysis, and three (20%) reported using SNP analysis. Compared to EQA-9 this is a small percentage decrease in the use of allele-based analysis but the number of participants using WGS for cluster analysis also increased from EQA-9, where 83% (10/12) reported using an allele-based method for the main analysis and 17% (2/12) reported using SNP analysis.

In this EQA, the test isolate (REF6) showed some interesting results. None of the laboratories using SNP as the main analysis included REF6 in the cluster. Furthermore, none of the laboratories using allele-based method as the main

analysis and SNP as an additional analysis included REF6 in the cluster. By the SNP analysis, laboratories reported at relatively clear identification of the cluster. The distance reported inside the cluster was 0-1 and the closest distance outside the cluster was at 9-12 SNP to REF6. One laboratory (100) reported more SNPs (0-5 inside the cluster and 21 SNP to REF6), but still a clear separation between cluster and non-cluster isolates was identified by the SNP analysis.

For the laboratories, using allele-based method, the reported allele differences were also very comparable. Inside the cluster (without REF6) the differences were 0-2 alleles for all the main analyses, but there was a variation in the cut-off used for selecting the cluster isolates. Nine alleles were the highest cut-off accepted by one laboratory (129) including REF6 using wgMLST. Five other laboratories also included REF6 in the reported cluster and accepted a cut-off of six alleles using cgMLST. Likewise, another six laboratories reported a difference at six alleles for REF6, but did not include the isolate in the cluster. It is notable that of these six laboratories, only one (laboratory 148) made the conclusion (not including REF6) based only on allele-based method, whereas the other five laboratories also made SNP analysis.

A high degree of similarity in the reported results using cgMLST/Enterobase (3002 loci) was seen for all but one laboratory. Laboratory 142 was the only laboratory to report a much higher number of allele differences for REF6 (19 allele differences, Table 7). This could, however, not be confirmed when the EQA provider analysed the raw reads by cgMLST from laboratory 142, as it showed the expected results of six allelic differences (Figure 9 and 10). Furthermore, laboratories 49 and 129 reported only nine allelic differences to REF6 when using "wgMLST" scheme (15874 and 4342 loci). During EQAs 8 to 10, the preference for using wgMLST has varied, and only laboratory 129 continues to use the ad hoc "wgMLST" scheme as their only analysis.

Laboratory 127 performed an additional cgMLST analysis in BioNumerics and found 0-4 allele differences inside the cluster. This was a peculiar result, as both laboratories 19 and 49, using same setup, obtained 0-1 allele differences as previously described. However, differences were also observed when the EQA provider analysed the data from laboratory 127. In addition, the laboratory also performed the same analysis in Enterobase for the main analysis and found only 0-1 allele differences inside the cluster. Discrepancies between the two approaches might be due to the interpretation of missing allele data. The data from laboratory 127 also showed a higher number of allelic differences when comparing to the reference sequences produced by the EQA provider (Figure 10), so generally a higher number of sequencing mistakes.

The reported SNP results were as mentioned comparable, only one laboratory (100) reported higher SNP distance (0-5) inside the cluster, and 21 to REF6 using wgSNP, this might be caused by the selected approach for analysis. This laboratory participated for the first time in the cluster analysis and identified the correct cluster based on SNP.

The EQA provider's analysis of the submitted raw data showed that, when using a standardised cgMLST analysis, a very high concordance was obtained (Figure 10). Our analysis of the data from laboratory 108 showed a higher number of allele differences, ranging from three to 42 for all isolates (including the two isolates that were supposedly swapped). This laboratory provided Ion Torrent data for which the EQA provider's analysis is not optimised, making correct assembly difficult, as also seen in the previous EQAs. Thus, the observed allele differences may be method artefacts, but the use of Ion Torrent data can complicate the communication and investigation of multi country outbreaks if only allelic method is used.

As seen in previous EQAs the two approaches to analyse WGS-derived data (allele- and SNP-based analysis) showed comparable results. Pearch and co-workers has also showed congruent results in a comparative analysis of core genome MLST and SNP typing [16]. However this year three SNP results seems to give a clearer separation of the cluster and non-cluster isolates, whereas the allele-based analyses makes the identification of the cluster less obvious, as half of the laboratories choose to include REF6 in the cluster and half did not include REF6.

The main reported QC parameters were coverage, cgMLST allele calls and genus/species confirmation, which are all essential for the end use of the data.

In the assessment of the additional EQA-provided genomes, all or almost all participants successfully identified the cluster isolate and the non-cluster isolate with the good quality. Almost all also identified the genome with the low coverage, but only a few identified the contamination (10%) in genome 4; however, most did identify the genome as a non-cluster isolate. One laboratory had unexpected results submitted for both genome 3 and 4.

Unfortunately, when observing the low-quality issue, most participants did not proceed to the cluster analysis and did not assess whether the data was of any use at all. Most responded that they would need to rerun the sequencing, and therefore said no to the question "Is this genome a part of the cluster?". The wording of the questions in the online form was apparently not formed in such a way that the participants understood that the EQA provider wished that they would proceed in their analysis. The EQA provider would have preferred the participants to see this as an urgent situation where a rerun is not possible, and therefore that they should try to make the cluster analysis with sub-optimal data and possibly tweak their standard analysis to be able to assess if the genomes could be part of the possible outbreak. For the next EQA, the EQA provider will rephrase the text and emphasise the importance of performing the cluster analysis with the available data – of course, concluded with the utmost caution.

The variation in the cut-off used for identifying the cluster was also observed in the assessment of the additional EQA-provided genomes. Laboratories 129 and 142 used the highest cut-off, and accepted respectively 8 and 7 allele differences for the inclusion of genome 1 in the cluster. As described above, the high number of allele differences by laboratory 142 in the main analysis could not be confirmed by the EQA provider, and perhaps a similar problem occurred in the analysis of genome 1. Likewise, laboratory 129 was the participant that accepted most allele differences for the inclusion of REF6 in the cluster by the main analysis.

Three laboratories (106, 142, and 149) described the use or establishment of cluster cut-off at seven allele differences (Annex 15). This threshold for cluster definition seems to have not been accepted or used by other of the participants, and a definitive cut-off in *Salmonella* has not been formally established.

5 Conclusions

Nineteen laboratories participated in the EQA-10 scheme: Nine (47%) performed MLVA and all participants (100%) performed cluster analyses using either one or more methods.

Again, in EQA-10, participation in the MLVA part was possible for both *S.* Typhimurium and *S.* Enteritidis, but the overall number of participants decreased, yet the performance level was still high for both analysis (93%). Five out of eight participants correctly assigned the MLVA profile for all 10 isolates of *S.* Typhimurium and seven out of nine in MLVA *S.* Enteritidis. One laboratory was responsible for 50% of the errors in *S.* Typhimurium and 66% in the *S.* Enteritidis.

All 19 laboratories participated in the molecular cluster analysis using either/or PFGE, MLVA or WGS. For the first time the evaluation of the cluster analyses was entirely based on a categorisation from WGS data. When using less discriminatory methods difficulties to achieve a correct cluster delineation could be expected, but only one of the laboratories analysing by PFGE did not identify the correct cluster. It was, however, also obvious that the majority of the laboratories combined the use of PFGE or MLVA with a second method, most often WGS.

Seven laboratories participated using PFGE for cluster analysis and two participated solely using PFGE-derived data for analysis. Six laboratories (86%) correctly identified the cluster, and one laboratory incorrectly included REF7. The laboratory did not perform WGS but identified the correct cluster by MLVA. The number of participants only using PFGE had decreased, whether this is due to changing towards WGS or just not participating in the EQA is for now unclear.

Five laboratories performed cluster analysis using MLVA-derived data, which is an increase compared to EQA-8 and EQA-9. Performance was high as all participating laboratories correctly identified the cluster of four isolates, none of the laboratories included isolates with one locus-variation.

Fifteen laboratories performed cluster analysis using WGS-derived data, an increase of three laboratories compared to last year. The performance was very high: 14 (93%) of the participants correctly identified the cluster of closely related isolates. The one laboratory with an incorrect result used SNP, but the error was probably due to the swapping of isolate REF3 and REF4. Twelve of 15 laboratories (80%) preferred an allele-based method, and only 20% (3/15) used SNP as the main reported cluster analysis. This was a small increase of SNP analyses compared to EQA-9.

A focus in this EQA has been one of the test isolates (REF6), where challenges and discrepancies occurred in the interpretation of whether it should be included in the cluster of closely related isolates. However, most of the participants (60%) did not include REF6 in the cluster, which is in accordance with pre-defined cluster by the provider.

The SNP analysis and allele-based analysis showed comparable results, but in this EQA the allele-based analysis gave less clear separation and resulted in different interpretation by the participants. Laboratories performing SNP analysis as a main or additional analysis had a more uniform interpretation, as the separation of cluster and non-cluster isolates seem clearer by SNP. All six laboratories including the test isolate REF6 only performed allele-based method (five cgMLST and one wgMLST). The reported allele-difference for REF6 was at six alleles for the majority of the laboratories, but they did not agree on the cut-off used for the cluster analyses. The provider expressed a concern when using a cut-off at six alleles, because that many unrelated isolates can potentially be included due to the clonal nature of monophasic *S.* Typhimurium.

Despite variation in the interpretation of REF6's relation to the cluster, the use of a standard cgMLST scheme (e.g. Enterobase) showed a high degree of homogeneity in the results. Although different approaches for analysing and different methods were probably used (different thresholds for allele calling, including or not including missing alleles in the analysis, assembly-based and/or mapping-based allele calling etc).

A new addition to the cluster analysis was introduced in EQA-10. Sequence data of four isolates were made accessible by the EQA provider, and the participants were asked to include these in the cluster analysis and report characteristics and quality issues. All but one participant correctly concluded whether the two genomes of high quality were part of the cluster or not. Almost all participants were able to identify the low-quality issue, but not all identified the 10% inclusion of *E. coli* in the *Salmonella* genome. For the genomes with low coverage, most participants did not proceed in assessing if the data would suggest the genome to be a part of the cluster. The wording of the questions in the online form will be rephrased in the next EQA to encourage participants to perform the more challenging analysis.

The current EQA scheme for typing of *Salmonella enterica* subsp. *enterica* is the 10th organised for laboratories in FWD-Net. The molecular typing-enhanced surveillance system implemented as part of TESSy relies on the capacity of FWD-Net laboratories to produce analysable and comparable typing results into a centralised database. WGS-based typing for surveillance is increasingly used in the EU. In 2019, ECDC opened up the possibility to submit WGS data for *Salmonella* to TESSy to be used for EU-wide surveillance and cross-sector comparison, but no results have been collected to date.

6 Recommendations

6.1 Laboratories

When laboratories use re-naming of the isolates, it might be useful to introduce a control procedure.

Laboratories with repeated or several errors in the MLVA part could use the possibility of repeating the MLVA analysis and submit the results for troubleshooting.

S. Enteritidis and *S.* Typhimurium are the two most common serovars in Europe and MLVA typing provides high discrimination within isolates of both serovars. Some of the laboratories that are not moving towards the use of WGS at this stage could benefit from implementing MLVA because of its low cost, easy analysis, and interpretation compared to WGS.

We encourage laboratories to submit their high-quality typing data to TESSy in as close to real time as possible.

We also recommend that laboratories use the EQA-provided data and isolates to validate their analysis methods when incorrect results (e.g. EQA) are obtained or when implementing new methods and procedures.

6.2 ECDC and FWD-Net

ECDC is working actively with the FWD-Net to improve the quality of sequence data generation and analysis through appropriate means like EQA schemes, expert exchange visits and workshops.

ECDC is encouraging more participants to take part in the new molecular typing-based cluster analysis, as well as participants who have not previously participated in the PFGE gel analysis or MLVA part.

ECDC will conduct an EQA feedback survey among participants.

6.3 EQA provider

The evaluation of the provided genome sequences was a success: almost all participant performed the analysis and identified the modifications introduced by the EQA provider. For the following EQA round, the EQA provider will continue and expand this part of the EQA in order to challenge the participants to learn to handle poor quality genomes as well as contaminated ones in their analysis when it is important to use the data available – concluded, of course, with the utmost caution.

As the interpretation of cluster cut-off was a main topic of this report, for the next FWD-Network meeting the EQA provider will suggest an open "cut-off" discussion.

References

1. Regulation (EC) No 851/2004 of the European Parliament and of the Council of 21 April 2004 establishing a European centre for disease prevention and control, Article 5.3. Available at: http://ecdc.europa.eu/en/aboutus/Key%20Documents/0404_KD_Regulation_establishing_ECDC.pdf
2. Decision No 1082/2013/EU of the European Parliament and the Council 22 October 2013 on serious cross-border threats to health and repealing Decision No 2119/98/EC (Text with EEA relevance). Available at: http://ec.europa.eu/health/preparedness_response/docs/decision_serious_crossborder_threats_22102013_en.pdf
3. European Food Safety Authority and European Centre for Disease Prevention and Control. The European Union One Health 2018 Zoonoses Report. EFSA Journal 2019;17(12):5926. Available at: https://efsa.onlinelibrary.wiley.com/doi/pdf/10.2903/j.efsa.2019.5926
4. European Centre for Disease Prevention and Control. Laboratory standard operating procedure for MLVA of *Salmonella enterica* serotype Typhimurium. Stockholm: ECDC; 2011. Available at: http://ecdc.europa.eu/publications-data/laboratory-standard-operating-procedure-mlva-salmonella-enterica-serotype
5. European Centre for Disease Prevention and Control. Laboratory standard operating procedure for multiple-locus variable-number tandem repeat analysis of *Salmonella enterica* serotype Enteritidis. Stockholm: ECDC; 2016. Available at: http://ecdc.europa.eu/publications-data/laboratory-standard-operating-procedure-multiple-locus-variable-number-tandem
6. International Organisation for Standardization. ISO/IEC 17043:2010: Conformity assessment -- General requirements for proficiency testing. ISO: Geneva, 2010. Available at: http://www.iso.org/iso/catalogue_detail.htm?csnumber=29366
7. Warwick Medical School. EnteroBase. Coventry: University of Warwick; 2018 [cited 21 August 2018]. Available at: http://enterobase.warwick.ac.uk
8. Sahl JW, Lemmer D, Travis J, Schupp JM, Gillece JD, Aziz M, et al. NASP: an accurate, rapid method for the identification of SNPs in WGS datasets that supports flexible input and output formats. Microb Genom. 2016 Aug 25;2(8):e000074.
9. Østerlund M and Kiil K. CleanRecomb, a quick tool for recombination detection in SNP based cluster analysis. bioRxiv. 2018; Available from https://github.com/ssi-dk/CleanRecomb
10. Statens Serum Institut. SerumQC. Copenhagen: SSI; 2018 [cited 21 August 2018]. Available at: http://www.github.com/ssi-dk/SerumQC
11. Gymoese P, Sørensen G, Litrup E, Olsen JE, Nielsen EM and Torpdahl M., 2017, Emerging Infectious Diseases, Vol. 23, No. 10. Available at: https://wwwnc.cdc.gov/eid/article/23/10/pdfs/16-1248.pdf
12. Colombe S, Jernberg C, Löf E, Angervall A L, Mellström-Dahlgren H et al., Outbreak of unusual H2S-negative monophasic Salmonella Typhimurium strain likely associated with small tomatoes, Sweden, August to October 2019. Euro Surveill. 2019;24(47):pii=1900643. doi: 10.2807/1560-7917.ES.2019.24.47.1900643
13. Mair-Jenkins J, Borges-Stewart R, Harbour C, Cox-Rogers J, Dallman T, et al. Investigation using whole genome sequencing of a prolonged restaurant outbreak of Salmonella Typhimurium linked to the building drainage system, England, February 2015 to March 2016. Euro Surveill. 2017;22(49):pii=17-00037. doi: 10.2807/1560-7917.ES.2017.22.49.17-00037
14. Siira L, MacDonald E, Holmbakken G M, Sundar T, Meyer-Myklestad L, et al. Increasing incubation periods during a prolonged monophasic Salmonella Typhimurium outbreak with environmental contamination of a commercial kitchen at Oslo Airport, Norway, 2017. Euro Surveill. 2019;24(34):pii=1900207. doi: 10.2807/1560-7917.ES.2019.24.34.1900207
15. Janezic S, Storman A, Frelih T, Grilc E, Trkov M and Rupnik M. Use of MLVA and WGS based approaches for investigation of outbreak caused by monophasic Salmonella Typhimurium. (Poster 162, 12th International Meeting on Microbial Epidemiological Markers IMMEMXIII, 2019).
16. Pearce M E, Alikhan N, Dallman T, Zhou Z, Grant K, et al. Comparative analysis of core genome MLST and SNP typing within a European Salmonella serovar Enteritidis outbreak. International Journal of Food Microbiology 274 (2018) 1–11.

Annex 1. List of participants

Country	Laboratory	National institute
Austria	National Reference Centre for Salmonella Austria	AGES / Institute for Medical Microbiology and Hyiene Graz
Belgium	National Reference Center for Salmonella & Shigella	Sciensano
Czechia	National Reference Laboratory for Salmonella	National Institute of Public Health
Denmark	Diagnostics and Typing of Gastrointestinal Bacteria	Statens Serum Institut
Finland	Expert microbiology	Finnish Institute for Health and Welfare (THL)
France	National Reference Centre for E. coli, Shigella & Salmonella	Institut Pasteur
Germany	National Reference Center for Salmonella an other bacterial enteric pathogens	Robert Koch Institute
Hungary	Food and waterborne diseases national reference laboratory	National Public Health Center
Ireland	NSSLRL	University Hospital Galway
Italy	Malattie Infettive	Istituto Superiore Di Sanità
Luxembourg	Epidemiology and Microbial Genomics	Laboratoire National de Sante
Norway	National Reference Laboratory for Enteropathogenic Bacteria	Norwegian Institute of Public Health
Romania	Molecular Epidemiology	Cantacuzino National Medico-Military Institute for Research and Development
Slovak Republic	NRC for Salmonelloses Laboratory of Molecular Diagnostics	Public Health Authority of the Slovak Republic
Slovenia	Department for Public Health Microbiology	National Laboratory of Health, Environment and Food
Spain	Unidad de Enterobacterias	Instituto de Salud Carlos III
Sweden	Mikrobiologen	Folkhälsomyndigheten
The Netherlands	IDS-BSR	RIVM
United Kingdom	Salmonella Reference Laboratory	Public Health England

Annex 2. Participation overview EQA-9 and 10

	2018 to 2019 (EQA-9)						2019 to 2020 (EQA-10)					
		MLVA		Cluster				MLVA		Cluster		
Laboratory	Participation (min. 1 part)	STm	SE	PFGE	MLVA	WGS	Participation (min. 1 part)	STm	SE	PFGE	MLVA	WGS
19	X	X	X	X	X	X	X	X	X		X	X
36	X					X	X					X
49	X					X	X					X
55	X	X	X	X			X	X	X	X	X	
92	X			X			X		X	X		
96	X			X			X			X		
100	X	X	X				X	X	X			X
106							X			X		X
108	X	X	X			X	X	X	X			X
127	X			X			X			X		X
128	X			X								
129	X	X	X			X	X					X
132	X			X								
134	X	X	X			X	X					X
135*	X	X	X				X	X	X			X
138	X			X								
140	X			X								
142	X	X	X	X		X	X	X	X	X	X	X
144	X	X	X	X		X	X	X	X	X	X	
145	X			X								
147	X	X	X	X	X	X	X	X	X		X	X
148	X					X	X					X
149	X					X	X					X
150	X					X	X					X
Total number of participants	**23**	**10**	**10**	**13**	**2**	**12**	**19**	**8**	**9**	**7**	**5**	**15**

**: previously laboratory 77*

Annex 3. Reason(s) for participating in EQA

	MLVA (STm and/or SE)				Cluster			
LAB ID	**Accreditation needs**	**Institute policy**	**National policy**	**Enhance typing quality**	**Accreditation needs**	**Institute policy**	**National policy**	**Enhance typing quality**
19				X				X
36	*MLVA S. Typhimurium is not relevant to our laboratory				X	X	X	
49	*MLVA S. Typhimurium is not relevant to our laboratory				X	X		X
55				X				X
92				X		X		X
96	*MLVA S. Typhimurium is not relevant to our laboratory							X
100		X	X	X		X	X	X
106	*MLVA S. Typhimurium is not relevant to our laboratory				X	X		X
108		X		X		X		X
114	X		X	X	X		X	X
127	*Lack of laboratory capacity				X		X	X
129	*MLVA S. Typhimurium is not relevant to our laboratory					X		
130	*Lack of laboratory capacity and Lack of financial means				X	X	X	X
134	*MLVA S. Typhimurium is not relevant to our laboratory							X
135**	X	X			X	X		
138			X		*We introduced MLVA for Salmonella in our laboratory but we have problems with results interpretation.			
142**	X				X			
144	X			X	X			X
147	X	X		X	X	X		X
148	*MLVA S. Typhimurium is not relevant to our laboratory							X
149	*In our laboratory MLVA is only a supplementary method to NGS				X	X		X
150	*Lack of laboratory capacity, we have replaced MLVA typing with WGS				X	X		
Number of participants	**5**	**4**	**3**	**7**	**12**	**12**	**5**	**16**

**: Reasons given when not participating.*
*** replied copied from 2018-2019*

Future WGS planes were reported for laboratories:
92: Until now we have performed WGS in our laboratory for research purposes only
127: We are planning to implement WGS with a period of three years.
144: We are already using WGS for typing. At the moment our institution is under reorganisation and some reagents are missing.

Annex 4. Scores of MLVA results *S*. Typhimurium

	Test isolates no. /allel																								
	STm1					STm2					STm3					STm4					STm5				
Lab. no.	STTR9	STTR5	STTR6	STTR10	STTR3	STTR9	STTR5	STTR6	STTR10	STTR3	STTR9	STTR5	STTR6	STTR10	STTR3	STTR9	STTR5	STTR6	STTR10	STTR3	STTR9	STTR5	STTR6	STTR10	STTR3
Provider	2	9	12	12	212	2	13	3	-2	212	4	10	13	11	211	3	12	9	20	311	9	14	-2	-2	211
19	2	9	12	12	212	2	13	3	-2	212	4	10	13	11	211	3	12	9	20	311	9	14	-2	-2	211
55	2	9	12	12	212	2	13	3	-2	212	4	10	13	11	211	3	12	9	20	311	9	14	-2	12	211
100	2	9	13	12	212	2	13	3	-2	212	4	10	13	11	211	3	12	9	20	311	9	14	-2	-2	211
108	2	9	12	12	212	2	13	3	-2	211	4	10	13	11	211	3	12	9	20	311	9	14	-2	-2	211
135	2	9	12	12	212	2	13	3	-2	212	4	10	13	11	211	3	12	9	20	311	9	14	-2	-2	211
142	2	9	12	12	212	2	13	3	-2	212	4	10	13	11	211	3	12	9	20	311	9	14	-2	-2	211
144	2	9	12	12	212	2	13	3	-2	212	4	10	13	11	211	3	12	9	20	311	9	14	-2	-2	211
147	2	9	12	12	212	2	13	3	-2	212	4	10	13	11	211	3	12	9	20	311	9	14	-2	-2	211

	Test isolates no. /allel																								
	STm6					STm7					STm8					STm9					STm10				
Lab. no.	STTR9	STTR5	STTR6	STTR10	STTR3	STTR9	STTR5	STTR6	STTR10	STTR3	STTR9	STTR5	STTR6	STTR10	STTR3	STTR9	STTR5	STTR6	STTR10	STTR3	STTR9	STTR5	STTR6	STTR10	STTR3
Provider	2	17	-2	-2	211	3	12	10	-2	211	3	12	13	21	311	3	13	-2	-2	211	3	12	9	-2	211
19	2	17	-2	-2	211	3	12	10	-2	211	3	12	13	21	311	3	13	-2	-2	211	3	12	9	-2	211
55	2	17	-2	13	211	3	12	10	20	211	3	12	13	21	311	3	13	-2	-2	211	3	12	9	-2	211
100	2	17	-2	-2	211	3	12	10	-2	211	3	12	13	21	311	3	13	-2	-2	211	3	12	9	-2	211
108	2	17	13	21	211	3	12	10	-2	211	3	12	13	21	311	3	13	-2	-2	211	3	12	9	-2	211
135	2	17	-2	-2	211	3	12	10	-2	211	3	12	13	21	311	3	13	-2	-2	211	3	12	9	-2	211
142	2	17	-2	-2	211	3	12	10	-2	211	3	12	13	21	211	3	13	-2	-2	211	3	12	9	-2	211
144	2	17	-2	-2	211	3	12	10	-2	211	3	12	13	21	311	3	13	-2	-2	211	3	12	9	-2	211
147	2	17	-2	-2	211	3	12	10	-2	211	3	12	13	21	311	3	13	-2	-2	211	3	12	9	-2	211

Purple: repeat isolates in EQA-4 to -10
Pink: incorrect results.
Green: accepted results

Annex 5. Scores of MLVA results *S.* Enteritidis

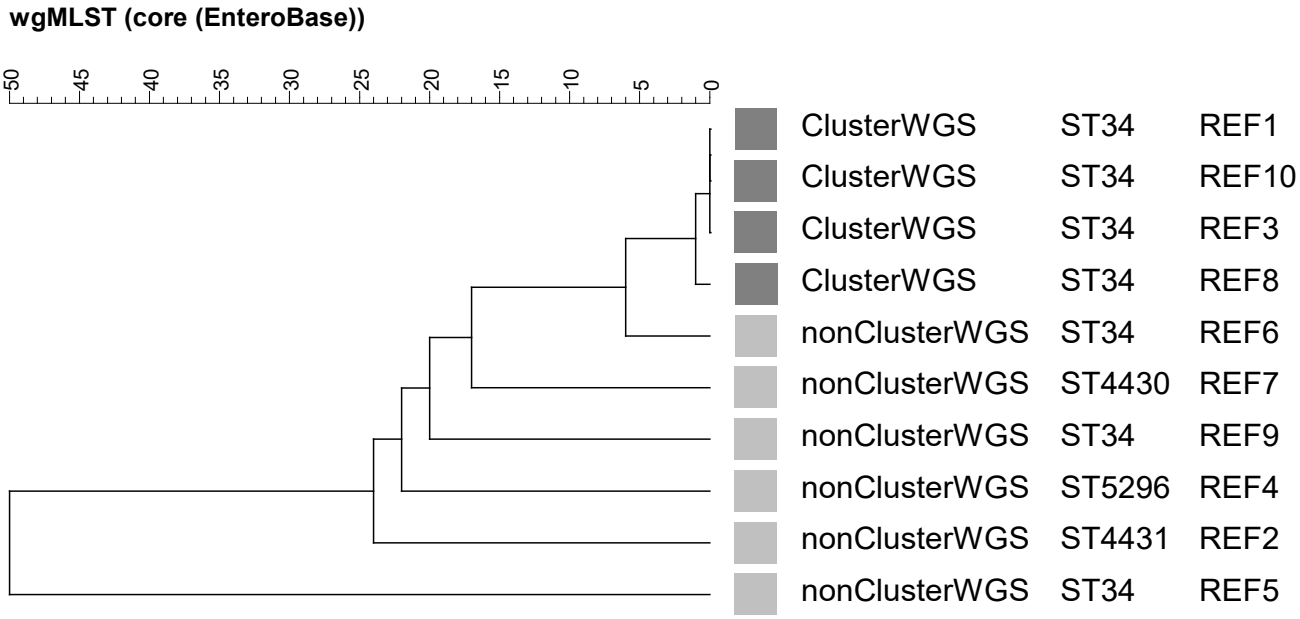
Test isolates no. /allel

Lab. no.	SE1					SE2					SE3					SE4					SE5				
	SENTR7	SENTR5	SENTR6	SENTR4	SE-3	SENTR7	SENTR5	SENTR6	SENTR4	SE-3	SENTR7	SENTR5	SENTR6	SENTR4	SE-3	SENTR7	SENTR5	SENTR6	SENTR4	SE-3	SENTR7	SENTR5	SENTR6	SENTR4	SE-3
Provider	2	12	7	3	2	2	12	3	3	2	2	13	9	3	2	3	11	6	5	1	3	12	5	4	1
19	2	12	7	3	2	2	12	3	3	2	2	13	9	3	2	3	11	6	5	1	3	12	5	4	1
55	2	12	7	5	2	2	12	3	3	2	2	13	9	5	2	3	11	6	6	1	3	12	5	4	1
92	2	12	7	3	2	2	12	3	3	2	2	13	9	3	2	3	11	6	5	1	3	12	5	4	1
100	2	12	7	3	2	2	12	3	3	2	2	13	9	3	2	3	11	6	5	1	3	12	5	4	1
108	2	12	7	3	2	2	12	3	3	2	2	13	9	3	2	3	11	6	5	1	3	12	5	4	1
135	2	12	7	3	2	2	12	3	3	2	2	13	9	3	2	3	11	6	5	1	3	12	5	4	1
142	2	12	7	3	2	2	12	3	3	2	2	13	9	3	2	3	11	6	5	1	3	12	5	4	1
144	2	12	7	3	2	2	12	3	3	2	2	13	9	3	2	3	11	6	5	1	3	12	5	4	1
147	2	12	7	3	2	2	12	3	3	2	2	13	9	3	2	3	11	6	5	1	3	12	5	4	1

Lab. no.	SE6					SE7					SE8					SE9					SE10				
	SENTR7	SENTR5	SENTR6	SENTR4	SE-3	SENTR7	SENTR5	SENTR6	SENTR4	SE-3	SENTR7	SENTR5	SENTR6	SENTR4	SE-3	SENTR7	SENTR5	SENTR6	SENTR4	SE-3	SENTR7	SENTR5	SENTR6	SENTR4	SE-3
Provider	3	10	5	5	1	3	16	5	5	2	2	4	4	4	1	3	11	4	4	1	1	10	7	3	2
19	3	10	5	5	1	3	16	5	5	2	2	4	4	4	1	3	11	4	4	1	1	10	7	3	2
55	3	10	5	5	1	3	17	5	5	2	2	4	4	4	1	3	11	4	4	1	1	10	7	3	2
92	3	10	5	5	1	3	16	5	5	2	2	4	4	4	1	3	11	4	4	1	1	10	7	3	2
100	3	10	5	5	1	3	16	5	5	2	2	4	4	4	1	3	11	4	4	1	1	10	7	3	2
108	3	10	5	5	1	3	16	5	5	2	2	4	4	4	1	3	11	4	4	1	1	10	7	3	2
135	3	10	5	5	1	3	16	5	5	2	2	4	4	4	1	3	11	4	4	1	1	10	7	3	2
142	3	10	5	5	1	3	16	5	5	2	2	4	4	4	1	3	11	4	4	1	1	10	7	3	2
144	1	10	7	3	2	3	16	5	5	2	2	4	4	4	1	3	11	4	4	1	3	10	5	5	1
147	3	10	5	5	1	3	16	5	5	2	2	4	4	4	1	3	11	4	4	1	1	10	7	3	2

Purple: repeat isolates in EQA-8 and -10
Pink: incorrect results.

Annex 6. EQA provider cluster analysis based on WGS-derived data



Single linked dendrogram of core genome multi-locus sequence typing (cgMLST) profiles of Salmonella EQA-10 isolates (cgMLST, EnteroBase, https://enterobase.warwick.ac.uk).
Analysed in BioNumerics: maximum distance of 200 exceeded; results clipped.
Dark grey: cluster isolates.
Light grey: outside cluster isolates.

Annex 7. Reported cluster of closely related isolates based on PFGE-derived data

Laboratory	Reported cluster	Corresponding REF isolates	Correct
		REF1, REF3, REF8, REF10	
55	5311 5598 5657 5679	REF10, REF3, REF1, REF8	Yes
92	5232 5807 5811 5901	REF8, REF1, REF3, REF10	Yes
96	5040 5161 5821 5999	REF3, REF8, REF1, REF10	Yes
106	5125 5214 5336 5389 5991	REF6, REF8, REF10, REF3, REF1	(Yes)
127	5111 5326 5785 5826	REF1, REF10, REF8, REF3	Yes
142	5086 5109 5775 5846	REF1, REF3, REF8, REF10	Yes
144	5045 5346 5518 5579 5569 5843	REF8, REF3, REF1, REF10, REF6, REF7	No

(Yes): laboratories including REF6 in the identified cluster

Annex 8. Reported sequencing details

Sequencing performed	Protocol (library preparation)	Commercial kit	Sequencing platform
In own laboratory	Commercial kits	Nextera XT DNA Sample Preparation Kit (Illumina)	HiSeq 2500
Externally	Commercial kits	Nextera XT DNA Library Preparation Kit	MiSeq
In own laboratory	Commercial kits	illumina Nextera DNA Flex Library Prep kit Ref: 20018705	MiSeq
In own laboratory	Commercial kits	Nextera	NextSeq
In own laboratory	Commercial kits	Nextera XT DNA Library Kit Illumina	MiSeq
In own laboratory	Commercial kits	Nextera XT	NextSeq
In own laboratory	Commercial kits	Nextera (Illumina)	MiSeq
In own laboratory	Commercial kits	NexteraXT (Illumina)	NextSeq
In own laboratory	Commercial kits	Nextera Flex Illumina*	MiniSeq
In own laboratory	Commercial kits	Ion Xpress TM Plus Fragment Library Kit for AB Library Builder TM System	Ion Torrent S5XL
In own laboratory	Commercial kits	Illumina Nextera XT	MiSeq
In own laboratory	Commercial kits	KAPA HyperPlus (Kapa Biosystems, Wilmington, US)	MiSeq
In own laboratory	Commercial kits	Nextera XT	NextSeq
In own laboratory	Commercial kits	MiSeq reagent kit v2	MiSeq
In own laboratory	Commercial kits	Nextera XT	MiSeq

**: We use half the volume of reagents for each step of the protocol*

Annex 9. Reported cluster of closely related isolates based on MLVA-derived data

Laboratory	Reported cluster	Corresponding to REF isolates	Correct
Provider		REF1, REF3, REF8, REF10	
19	5072, 5127, 5367, 5352	REF3, REF1, REF10, REF8	Yes
55	5311, 5598, 5657, 5679	REF10, REF3, REF1, REF8	Yes
142	5086, 5109, 5775, 5846	REF1, REF3, REF8, REF10	Yes
144	5045, 5346, 5518, 5579	REF8, REF3, REF1, REF10	Yes
147	5064, 5146, 5681, 5689	REF1, REF8, REF3, REF10	Yes

Annex 10 Reported MLVA profile data

				Laboratory ID				
Isolate number	**ST**	**MLVA scheme**	**Provider**	**19**	**55**	**142**	**144**	**147**
REF1‡	34	*S.* Typhimurium	3-15-11-NA-211	3-15-11-NA-211	3-15-11-NA-211	3-15-11-NA-211	3-15-11-NA-211	3-15-11-NA-211
REF2	4431	*S.* Typhimurium	3-13-11-NA-211	3-13-11-NA-211	3-13-11-NA-211	3-13-11-NA-211	3-13-11-NA-211	3-13-11-NA-211
REF3#‡	34	*S.* Typhimurium	3-15-11-NA-211	3-15-11-NA-211	3-15-11-NA-211	3-15-11-NA-211	3-15-11-NA-211	3-15-11-NA-211
REF4	5296	*S.* Typhimurium	3-12-10-NA-211	3-12-10-NA-211	3-12-10-NA-211	3-12-10-NA-211	3-12-10-NA-211	3-12-10-NA-211
REF5	34	*S.* Typhimurium	3-14-13-NA-211	3-14-13-NA-211	3-14-13-NA-211	3-14-13-NA-211	3-14-13-NA-211	3-14-13-NA-211
REF6	34	*S.* Typhimurium	3-15-13-NA-211	3-15-13-NA-211	3-15-13-NA-211	3-15-13-NA-211	3-15-13-NA-211	3-15-13-NA-211
REF7	4430	*S.* Typhimurium	3-12-11-NA-211	3-12-11-NA-211	3-12-11-NA-211	3-12-11-NA-211	3-12-11-NA-211	3-12-11-NA-211
REF8‡	34	*S.* Typhimurium	3-15-11-NA-211	3-15-11-NA-211	3-15-11-NA-211	3-15-11-NA-211	3-15-11-NA-211	3-15-11-NA-211
REF9	34	*S.* Typhimurium	3-12-9-NA-211	3-12-9-NA-211	3-12-9-NA-211	3-12-9-NA-211	3-12-9-NA-211	3-12-9-NA-211
REF10#‡	34	*S.* Typhimurium	3-15-11-NA-211	3-15-11-NA-211	3-15-11-NA-211	3-15-11-NA-211	3-15-11-NA-211	3-15-11-NA-211

‡: cluster analyses of closely related isolates (based on MLVA-derived data)
#: technical duplet
ST: 7-multi locus sequence type
NA: designates a locus not present (-2 submitted by participants).

Annex 11. Reported cluster of closely related isolates based on WGS-derived data

Laboratory	Reported cluster	Corresponding to REF isolates	Correct
Provider		REF1, REF3, REF8, REF10	
19	5072, 5127, 5367, 5352	REF3, REF1, REF10, REF8	Yes
36	5122, 5581, 5912, 5950	REF2, REF8, REF4, REF3	Yes
49	5233, 5349, 5596, 5599	REF8, REF10, REF3, REF1	Yes
100	5138, 5210, 5784, 5864	REF8, REF3, REF10, REF1	Yes
106	5389, 5336, 5991, 5125, 5214	REF3, REF10, REF1, REF6, REF8	(Yes)
108	5961, 5651, 5071, 5457	REF1, REF4, REF10, REF8	No
127	5826, 5785, 5326, 5111	REF3, REF8, REF10, REF1	Yes
129	5871, 5401, 5188, 5153, 5974	REF2, REF9, REF5, REF10, REF3	(Yes)
134	5207, 5668, 5847, 5378, 5840	REF10, REF1, REF3, REF8, REF6	(Yes)
135	5278, 5416, 5564, 5896, 5503	REF10, REF3, REF1, REF8, REF6	(Yes)
142	5086, 5109, 5775, 5846	REF1, REF3, REF8, REF10	Yes
147	5064, 5146, 5681, 5689, 5810	REF1, REF8, REF3, REF10, REF6	(Yes)
148	5175, 5212, 5462, 5856	REF8, REF10, REF1, REF3	Yes
149	5241, 5578, 5594, 5631, 5783	REF6, REF3, REF10, REF8, REF1	(Yes)
150	5466, 5557, 5933, 5835	REF1, REF3, REF10, REF8	Yes

(Yes): laboratories including REF6 in the identified cluster

Annex 12. Reported SNP distance and allelic differences

SNP distances

Isolate number	ST	Provider	Provider + Recombination filter	19*	36*	49**	100	108	127**	150
REF1‡	34	0	0	0	0	0	3	0¤	0¤	0¤
REF2	4431	152	87	65	57	197	122	9999	119	125
REF3‡#	34	0¤	0¤	0¤	0¤	0	0¤	9999*	0	0
REF4	5296	63	67	54	51	275	117	0*	64	86
REF5	34	172	113	97	94	134	145	109	138	102
REF6	34	10	10	11	9	12	21	11	12	11
REF7	4430	46	46	41	36	43	61	9999	39	40
REF8‡	34	1	1	1	1	1	5	1	1	1
REF9	34	58	58	53	48	56	108	55	52	54
REF10‡#	34	0	0	0	0	0¤	0	0	0	0

* Laboratory 108, a possible swapping of REF3 and REF4.

Allelic differences

			Laboratory ID															
Isolate number	ST	Provider	19	36	49	49*	100*	100**	106	127	127*	129	134	135	142	147	148	149
REF1‡	34		0	2	0	0	0	0	0	0	2	0¤	0¤	0	0¤	0¤	0	0
REF2	4431		30	31	30	61	62	63	31	67	45	47	30	65	43	31	49	31
REF3‡#	34		0¤	0¤	0	1	0	0	0¤	0¤	0	1	0	0	1	0	0	0¤
REF4	5296		24	25	24	57	40	41	24	42	23	40	23	41	41	24	24	24
REF5	34		52	52	53	88	69	69	52	54	48	67	51	52	72	52	52	51
REF6	34		6	6	6	9	6	6	6	7	6	9	6	6	19	6	6	6
REF7	4430		17	19	17	35	18	19	19	18	20	27	18	18	31	19	19	19
REF8‡	34		1	1	1	1	1	1	1	1	2	1	1	1	2	1	1	1
REF9	34		22	26	22	45	28	28	23	41	24	36	23	40	37	23	25	23
REF10‡#	34		0	1	0¤	0	0	0	0	0	4	0	0	0¤	0	0	0¤	0

**: additional analysis*
***: 3. analysis*
‡: closely related isolates
#: technical duplicate isolates
¤: isolate used as cluster representative by the participant
ST: sequence type.

Annex 13. Additional reported QC parameters

Lab ID	1		2		3		4	
	Parameter	Threshold	Parameter	Threshold	Parameter	Threshold	Parameter	Threshold
36	percent good cgMLST-targets	0,95						
49	NrBAF perfect	>4 000	NrBAF multiple	Pass <20, Warn >20	Quality	Pass >=30		
100	N50	100 000	contig count	less than 150 contigs	contamination	BWA mapping with Listeria genome, 0 contigs assembled	SAV	cluster density, clusters passing filter and Q30 score were all according to Illumina recommendations
106	Reads QC minimum lenght	50	Reads QC: Trimming (Trimmomatic)	Q20				
127	No of contigs	>= 300	No of % N bases	< 3 %				
129	number of good cgMLST loci	>90%	Contig Count	-				
142	Phred score to evaluate quality of the sequence data	28						
147	no. of contigs	200 bases (contigs shorter than 200 bases have to be ignored)						
148	N50	>20 Kb (in house and Enterobase)	Proportion of scaffolding placeholders (N's)	< 3%	Sickle 1.33 to trim the ends of short reads (FASTQ) of base calls with low quality scores	argument: score =10		
150	Minimum Read Length	>50 after trimming with trimmomatic	Minimum Read Count	>10,000	Assessment of bacterial contamination	Kmer ID, look at similarity and reference genome, it is rejected if there is >10% unexplained similarity		

Annex 14. Calculated qualitative/quantitative parameters

Parameters	Ranges *	Laboratory 19									
		5072	5127	5211	5352	5367	5608	5632	5645	5754	5779
Detected species	{Se}	Se	Se	Se	Se	Se	Se	Se	Se	Se	Se
% Species 1		96.1	95.6	94.5	95.8	95.7	94.1	95.9	96.1	95.2	96.1
% Species 2	{<5%}	0.0	0.1	0.0	0.0	0.1	1.5	0.0	0.0	0.3	0.0
% unclassified		3.8	4.3	5.5	4.1	4.1	4.0	4.0	3.9	4.3	3.8
Length at >25 x min. coverage (Mbp)	{>4.5 ∧ <5.3}	4.9	4.9	4.8	4.9	4.9	5.0	4.9	4.9	4.9	4.9
Length at [1-25] x min. coverage (Kbp)	{<0.25}	6.4	5.1	6.2	3.3	3.9	8.4	2.2	8.2	4.2	4.5
No. of contigs at 25 x min. coverage	{>0}	123	138	144	119	125	190	134	125	140	116
No. of contigs at [1,25] x min. coverage	{<1 000}	6	8	8	4	5	7	3	10	6	4
Average coverage	{>50}	130	135	128	102	134	112	153	115	119	123
No. of reads (x 1000)		4442	4656	4389	3528	4649	3955	5325	3931	4125	4267
Average read length		145	144	143	143	143	144	143	145	144	144
Average insert size		318	296	306	318	299	308	306	334	316	321
N50 (Kbp)		83	78	56	83	75	47	83	72	69	83

Parameters	Ranges *	Laboratory 36									
		5120	5122	5414	5482	5489	5565	5581	5637	5912	5950
Detected species	{Se}	Se	Se	Se	Se	Se	Se	Se	Se	Se	Se
% Species 1		94.7	94.7	96.3	96.3	98.0	96.0	94.6	94.0	97.1	96.6
% Species 2	{<5%}	0.1	0.0	0.1	0.3	0.1	1.7	0.1	0.0	0.0	0.1
% unclassified		5.1	5.2	3.5	3.0	1.9	1.9	5.1	5.9	2.8	3.3
Length at >25 x min. coverage (Mbp)	{>4.5 ∧ <5.3}	4.7	4.8	4.8	4.8	4.9	4.9	4.7	4.7	4.7	4.8
Length at [1-25] x min. coverage (Kbp)	{<0.25}	115.7	69.4	34.3	31.6	11.9	15.8	91.1	47.2	61.7	54.0
No. of contigs at 25 x min. coverage	{>0}	221	223	176	210	208	242	200	212	260	245
No. of contigs at [1,25] x min. coverage	{<1 000}	61	46	22	21	11	18	49	33	47	38
Average coverage	{>50}	45	50	63	74	109	105	44	58	52	55
No. of reads (x 1000)		843	976	1104	1308	1875	1846	796	1002	898	950
Average read length		271	258	287	285	290	290	282	288	289	290
Average insert size		615	545	617	575	528	532	651	606	606	594
N50 (Kbp)		33	35	46	42	41	33	33	33	28	29

Parameters	Ranges *	Laboratory 49									
		5233	5251	5266	5349	5552	5596	5599	5641	5778	5992
Detected species	{Se}	Se	Se	Se	Se	Se	Se	Se	Se	Se	Se
% Species 1		98.0	98.0	98.7	98.4	98.3	96.0	94.7	98.0	96.2	97.6
% Species 2	{<5%}	0.0	0.0	0.0	0.0	0.0	0.1	0.0	0.1	1.6	0.2
% unclassified		1.9	2.0	1.2	1.6	1.6	3.8	5.3	1.9	1.8	2.1
Length at >25 x min. coverage (Mbp)	{>4.5 ∧ <5.3}	4.9	4.9	4.9	4.9	4.9	3.1	3.4	4.9	5.0	4.9
Length at [1-25] x min. coverage (Kbp)	{<0.25}	0.0	0.0	0.0	0.0	0.0	1824.4	1497.9	0.0	0.0	0.0
No. of contigs at 25 x min. coverage	{>0}	79	47	58	46	43	40	53	59	76	47
No. of contigs at [1,25] x min. coverage	{<1 000}	0	0	0	0	0	27	39	0	0	0
Average coverage	{>50}	86	87	100	104	72	26	26	95	103	54
No. of reads (x 1000)		2004	1615	1786	1837	1287	553	629	1804	1940	974
Average read length		215	267	279	282	279	239	209	264	274	277
Average insert size		234	331	377	409	390	366	278	326	362	370
N50 (Kbp)		283	239	226	316	319	213	143	268	166	316

Parameters	Ranges *	Laboratory 100									
		5015	5138	5210	5269	5307	5377	5573	5676	5784	5864
Detected species	{Se}	Se	Se	Se	Se	Se	Se	Se	Se	Se	Se
% Species 1		98.4	98.5	98.8	96.7	96.9	98.0	98.7	98.3	98.4	98.5
% Species 2	{<5%}	0.0	0.0	0.0	1.7	0.0	0.3	0.0	0.0	0.0	0.0
% unclassified		1.5	1.4	1.1	1.1	3.0	1.5	1.2	1.6	1.5	1.5
Length at >25 x min. coverage (Mbp)	{>4.5 ∧ <5.3}	4.9	4.9	4.9	5.0	4.9	4.9	4.9	4.9	4.9	4.9
Length at [1-25] x min. coverage (Kbp)	{<0.25}	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
No. of contigs at 25 x min. coverage	{>0}	47	46	52	71	44	54	58	49	52	49
No. of contigs at [1,25] x min. coverage	{<1 000}	0	0	0	0	0	0	0	0	0	0
Average coverage	{>50}	68	61	129	107	62	86	51	74	68	68
No. of reads (x 1000)		1405	1262	2682	2264	1276	1805	1066	1539	1401	1395
Average read length		243	243	241	243	242	237	242	242	242	242
Average insert size		388	423	361	388	384	324	380	386	364	387
N50 (Kbp)		319	316	316	150	239	301	226	280	245	279

Parameters	Ranges *	Laboratory 106									
		5125	5143	5214	5336	5389	5433	5505	5621	5919	5991
Detected species	{Se}	Se	Se	Se	Se	Se	Se	Se	Se	Se	Se
% Species 1		93.5	92.8	94.4	94.9	93.7	91.9	92.2	94.0	93.9	94.7
% Species 2	{<5%}	0.0	0.0	0.0	0.0	0.0	0.6	1.3	0.0	0.0	0.0
% unclassified		6.4	7.1	5.5	5.0	6.2	7.2	6.3	5.9	6.1	5.2
Length at >25 x min. coverage (Mbp)	{>4.5 ∧ <5.3}	4.9	4.9	4.9	4.9	4.9	4.9	5.0	4.9	4.9	4.9
Length at [1-25] x min. coverage (Kbp)	{<0.25}	0.0	0.0	0.0	0.0	0.0	0.0	0.5	0.0	0.0	0.0
No. of contigs at 25 x min. coverage	{>0}	77	68	63	65	72	70	109	56	79	67
No. of contigs at [1,25] x min. coverage	{<1 000}	0	0	0	0	0	0	1	0	0	0
Average coverage	{>50}	259	214	202	212	210	232	239	287	244	246
No. of reads (x 1000)		9951	8243	7473	7740	7935	8775	9318	9592	9278	9046
Average read length		131	129	134	136	132	132	132	148	132	135
Average insert size		216	216	245	263	215	231	229	432	223	254
N50 (Kbp)		226	261	245	226	245	283	113	176	211	271

Parameters	Ranges *	Laboratory 108									
		5071	5159	5457	5483	5651	5791	5961	5976	5443	5148
Detected species	{Se}	Se	Se	Se	Se	Se	Se	Se	Se	Se	Se
% Species 1		98.0	98.0	97.6	95.2	98.2	97.6	97.2	97.8	97.6	96.7
% Species 2	{<5%}	0.5	0.4	0.4	2.4	0.4	0.7	0.4	0.6	0.4	0.9
% unclassified		1.5	1.6	1.9	1.7	1.4	1.6	2.4	1.5	2.0	2.3
Length at >25 x min. coverage (Mbp)	{>4.5 ∧ <5.3}	4.8	4.7	4.9	4.9	4.7	4.7	4.9	4.8	4.6	4.9
Length at [1-25] x min. coverage (Kbp)	{<0.25}	0.0	1.3	0.0	3.0	2.6	0.0	0.0	0.0	0.4	0.0
No. of contigs at 25 x min. coverage	{>0}	1131	2230	666	2115	2385	2397	681	1091	2628	398
No. of contigs at [1,25] x min. coverage	{<1 000}	0	6	0	14	11	0	0	0	1	0
Average coverage	{>50}	174	86	109	85	85	97	118	134	108	117
No. of reads (x 1000)		2718	1414	1840	1488	1363	1589	2017	2180	1729	1933
Average read length		318	301	295	292	309	304	291	308	308	302
Average insert size		0	0	0	0	0	0	0	0	0	0
N50 (Kbp)		7	3	12	4	3	3	13	7	3	23

Parameters	Ranges *	Laboratory 127									
		5111	5326	5440	5458	5582	5755	5785	5826	5982	5986
Detected species	{Se}	Se	Se	Se	Se	Se	Se	Se	Se	Se	Se
% Species 1		97.8	97.3	96.2	96.9	96.6	98.4	97.9	97.8	95.8	96.5
% Species 2	{<5%}	0.1	0.0	0.0	0.0	1.1	0.0	0.0	0.0	0.4	0.0
% unclassified		2.1	2.7	3.8	3.0	2.1	1.6	2.0	2.1	3.6	3.5
Length at >25 x min. coverage (Mbp)	{>4.5 ∧ <5.3}	4.9	4.9	4.9	2.6	4.9	4.9	4.9	4.9	4.9	4.8
Length at [1-25] x min. coverage (Kbp)	{<0.25}	0.6	25.4	4.7	2263.9	91.3	0.0	3.1	6.3	2.1	1.4
No. of contigs at 25 x min. coverage	{>0}	58	102	83	79	99	63	59	95	55	58
No. of contigs at [1,25] x min. coverage	{<1 000}	1	10	3	83	3	0	3	5	1	2
Average coverage	{>50}	88	48	57	26	45	66	57	66	100	93
No. of reads (x 1000)		2941	1594	1893	859	1540	2222	1912	2202	3377	3051
Average read length		150	150	150	150	149	150	149	150	149	149
Average insert size		558	608	583	581	578	521	525	566	492	531
N50 (Kbp)		187	89	129	63	107	199	187	105	239	164

Parameters	Ranges *	Laboratory 129									
		5145	5153	5188	5401	5426	5587	5852	5871	5932	5974
Detected species	{Se}	Se	Se	Se	Se	Se	Se	Se	Se	Se	Se
% Species 1		97.2	97.5	97.6	97.4	97.5	96.7	95.6	97.1	96.0	96.4
% Species 2	{<5%}	0.0	0.0	0.0	0.0	0.1	0.0	1.3	0.0	0.2	0.1
% unclassified		2.7	2.4	2.3	2.5	2.4	3.3	2.8	2.8	3.6	3.5
Length at >25 x min. coverage (Mbp)	{>4.5 ∧ <5.3}	4.9	4.9	4.9	4.9	4.9	4.8	5.0	4.9	4.9	4.9
Length at [1-25] x min. coverage (Kbp)	{<0.25}	3.4	10.9	4.4	5.2	3.5	0.7	4.2	3.9	8.2	3.4
No. of contigs at 25 x min. coverage	{>0}	78	115	121	185	161	91	154	156	141	104
No. of contigs at [1,25] x min. coverage	{<1 000}	4	10	6	8	6	1	6	5	7	4
Average coverage	{>50}	105	99	94	148	110	102	127	148	78	130
No. of reads (x 1000)		3541	3332	3171	5008	3747	3409	4427	5037	2670	4396
Average read length		147	148	147	147	147	147	146	146	146	148
Average insert size		382	407	387	352	353	370	377	355	417	400
N50 (Kbp)		186	76	83	47	58	102	61	62	59	90

Parameters	Ranges *	Laboratory 134									
		5207	5378	5399	5470	5571	5668	5684	5840	5847	5857
Detected species	{Se}	Se	Se	Se	Se	Se	Se	Se	Se	Se	Se
% Species 1		97.8	97.8	97.0	96.3	96.0	98.2	98.1	97.6	97.5	97.9
% Species 2	{<5%}	0.3	0.3	0.3	1.3	0.1	0.2	0.2	0.3	0.3	0.3
% unclassified		1.7	1.7	2.2	1.8	3.7	1.4	1.5	1.8	2.1	1.7
Length at >25 x min. coverage (Mbp)	{>4.5 ∧ <5.3}	4.9	4.9	4.9	5.0	4.9	4.9	4.9	4.9	4.9	4.9
Length at [1-25] x min. coverage (Kbp)	{<0.25}	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
No. of contigs at 25 x min. coverage	{>0}	68	63	66	95	62	65	67	68	69	69
No. of contigs at [1,25] x min. coverage	{<1 000}	0	0	0	0	0	0	0	0	0	0
Average coverage	{>50}	52	53	50	50	58	66	43	43	42	49
No. of reads (x 1000)		1708	1760	1672	1699	1924	2197	1422	1437	1383	1636
Average read length		151	151	151	151	151	151	151	151	151	151
Average insert size		337	337	364	346	331	320	327	356	339	355
N50 (Kbp)		223	283	239	144	223	223	261	223	214	200

Parameters	Ranges *	Laboratory 135									
		5128	5192	5278	5405	5416	5437	5503	5564	5896	5955
Detected species	{Se}	Se	Se	Se	Se	Se	Se	Se	Se	Se	Se
% Species 1		94.9	97.9	97.8	96.0	97.8	97.6	97.8	97.9	97.8	97.0
% Species 2	{<5%}	0.0	0.0	0.0	1.4	0.0	0.1	0.0	0.0	0.0	0.3
% unclassified		5.0	2.1	2.1	2.2	2.2	2.2	2.1	2.1	2.1	2.5
Length at >25 x min. coverage (Mbp)	{>4.5 ∧ <5.3}	4.9	4.9	4.9	5.0	4.9	4.9	4.9	4.9	4.9	4.9
Length at [1-25] x min. coverage (Kbp)	{<0.25}	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
No. of contigs at 25 x min. coverage	{>0}	45	55	56	88	53	60	55	54	54	59
No. of contigs at [1,25] x min. coverage	{<1 000}	0	0	0	0	0	0	0	0	0	0
Average coverage	{>50}	232	263	198	248	168	107	195	266	158	224
No. of reads (x 1000)		7628	8683	6582	8470	5567	3565	6524	8859	5255	7476
Average read length		150	149	149	149	149	150	149	149	149	149
Average insert size		416	401	419	407	431	433	401	369	405	406
N50 (Kbp)		283	239	226	135	245	211	268	280	207	239

Parameters	Ranges*	Laboratory 142									
		5055	5086	5109	5237	5529	5568	5712	5775	5846	5882
Detected species	{Se}	Se	Se	Se	Se	Se	Se	Se	Se	Se	Se
% Species 1		97.8	97.9	97.5	95.7	97.7	97.4	95.5	98.2	98.0	97.3
% Species 2	{<5%}	0.2	0.1	0.3	0.2	0.2	0.1	2.0	0.1	0.2	0.2
% unclassified		1.7	1.9	2.0	3.9	2.0	2.3	1.8	1.6	1.6	2.3
Length at >25 x min. coverage (Mbp)	{>4.5 ∧ <5.3}	4.9	4.9	4.9	4.9	4.9	4.9	5.0	4.9	4.9	4.9
Length at [1-25] x min. coverage (Kbp)	{<0.25}	0.0	0.0	5.4	1.4	0.0	0.5	0.0	0.0	5.9	0.0
No. of contigs at 25 x min. coverage	{>0}	56	44	47	46	61	44	73	52	51	49
No. of contigs at [1,25] x min. coverage	{<1 000}	0	0	8	3	0	1	0	0	2	0
Average coverage	{>50}	79	67	50	46	58	58	58	59	56	45
No. of reads (x 1000)		1728	1429	1026	945	1226	1189	1232	1261	1182	928
Average read length		230	235	243	244	240	243	241	235	236	245
Average insert size		343	378	431	421	397	442	418	367	368	468
N50 (Kbp)		283	316	273	239	224	319	178	283	316	268

Parameters	Ranges *	Laboratory 147									
		5064	5141	5146	5374	5422	5681	5689	5804	5810	5862
Detected species	{Se}	Se	Se	Se	Se	Se	Se	Se	Se	Se	Se
% Species 1		97.5	93.2	97.5	93.4	97.5	97.7	97.8	93.7	96.2	97.5
% Species 2	{<5%}	0.1	1.8	0.2	0.1	0.1	0.1	0.2	1.2	0.1	0.2
% unclassified		2.3	4.5	2.3	6.4	2.3	2.1	2.0	3.7	3.6	2.2
Length at >25 x min. coverage (Mbp)	{>4.5 ∧ <5.3}	4.9	5.0	4.9	4.9	4.9	4.9	4.9	4.9	4.9	4.9
Length at [1-25] x min. coverage (Kbp)	{<0.25}	0.0	0.4	0.6	1.4	0.0	0.0	0.4	0.0	0.0	0.0
No. of contigs at 25 x min. coverage	{>0}	54	96	77	64	77	69	61	70	68	81
No. of contigs at [1,25] x min. coverage	{<1 000}	0	1	1	1	0	0	1	0	0	0
Average coverage	{>50}	84	67	74	75	70	93	79	74	84	93
No. of reads (x 1000)		1662	1335	1360	1385	1311	1797	1516	1415	1614	1964
Average read length		255	265	272	272	267	261	262	263	265	238
Average insert size		321	343	382	376	358	333	335	342	346	282
N50 (Kbp)		283	105	150	149	134	239	224	182	191	192

Parameters	Ranges *	Laboratory 148									
		5175	5205	5212	5265	5276	5462	5715	5856	5875	5913
Detected species	{Se}	Se	Se	Se	Se	Se	Se	Se	Se	Se	Se
% Species 1		98.3	97.2	98.3	98.3	96.5	98.3	97.9	97.9	97.4	98.1
% Species 2	{<5%}	0.0	0.0	0.0	0.0	1.2	0.0	0.0	0.0	0.1	0.0
% unclassified		1.7	2.8	1.6	1.6	2.0	1.7	2.1	2.0	2.3	1.8
Length at >25 x min. coverage (Mbp)	{>4.5 ∧ <5.3}	4.9	4.8	4.8	4.9	5.0	4.9	4.8	4.8	4.8	4.9
Length at [1-25] x min. coverage (Kbp)	{<0.25}	1.4	0.7	13.4	35.5	38.8	7.2	71.8	15.0	104.4	14.2
No. of contigs at 25 x min. coverage	{>0}	98	137	225	207	209	139	177	223	227	150
No. of contigs at [1,25] x min. coverage	{<1 000}	2	1	12	19	19	4	31	17	66	14
Average coverage	{>50}	154	259	153	135	102	140	105	98	79	82
No. of reads (x 1000)		5066	8421	5006	4456	3443	4596	3448	3211	2599	2711
Average read length		150	150	150	150	149	150	149	149	149	150
Average insert size		415	398	380	372	385	367	380	377	371	401
N50 (Kbp)		94	68	43	46	43	72	48	38	37	55

Parameters	Ranges *	Laboratory 149									
		5052	5241	5420	5430	5493	5578	5594	5631	5783	5931
Detected species	{Se}	Se	Se	Se	Se	Se	Se	Se	Se	Se	Se
% Species 1		94.8	96.7	96.1	95.6	94.9	97.3	96.6	96.5	96.8	94.5
% Species 2	{<5%}	1.2	0.6	0.5	0.8	0.6	0.3	0.6	0.4	0.4	0.7
% unclassified		2.8	2.6	2.8	3.2	4.2	2.3	2.6	2.8	2.7	3.4
Length at >25 x min. coverage (Mbp)	{>4.5 ∧ <5.3}	5.0	4.9	4.9	4.9	4.9	4.9	4.9	4.9	4.9	4.9
Length at [1-25] x min. coverage (Kbp)	{<0.25}	0.0	45.2	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.2
No. of contigs at 25 x min. coverage	{>0}	77	54	52	63	46	50	53	49	48	52
No. of contigs at [1,25] x min. coverage	{<1 000}	0	6	0	0	0	0	0	0	0	1
Average coverage	{>50}	110	78	89	112	87	95	103	96	95	90
No. of reads (x 1000)		2169	1478	1734	2262	1690	1834	2053	1927	1861	1704
Average read length		266	268	262	255	259	260	254	253	259	271
Average insert size		353	352	337	321	331	329	307	316	336	373
N50 (Kbp)		144	232	316	224	283	316	316	316	316	301

Parameters	Ranges *	Laboratory 150									
		5256	5382	5466	5557	5623	5728	5799	5835	5844	5933
Detected species	*{Se}*	*Se*	*Se*	*Se*	*Se*	*Se*	*Se*	*Se*	*Se*	*Se*	*Se*
% Species 1		*94.4*	*92.7*	*94.4*	*94.6*	*94.5*	*94.0*	*92.9*	*94.6*	*94.6*	*94.5*
% Species 2	*{<5%}*	*0.1*	*0.0*	*0.0*	*0.0*	*0.1*	*0.2*	*1.4*	*0.0*	*0.0*	*0.1*
% unclassified		*5.5*	*7.2*	*5.5*	*5.3*	*5.3*	*5.7*	*5.3*	*5.3*	*5.3*	*5.4*
Length at >25 x min. coverage (Mbp)	{>4.5 ∧ <5.3}	*4.9*	*4.8*	*4.9*	*4.9*	*4.9*	*4.9*	*5.0*	*4.9*	*4.9*	*4.9*
Length at [1-25] x min. coverage (Kbp)	{<0.25}	*0.0*	*0.0*	*0.0*	*0.0*	*4.8*	*0.0*	*0.0*	*0.0*	*0.0*	*0.0*
No. of contigs at 25 x min. coverage	{>0}	*86*	*70*	*79*	*78*	*81*	*75*	*115*	*73*	*78*	*84*
No. of contigs at [1,25] x min. coverage	{<1 000}	*0*	*0*	*0*	*0*	*4*	*0*	*0*	*0*	*0*	*0*
Average coverage	{>50}	*70*	*81*	*61*	*83*	*60*	*63*	*77*	*70*	*66*	*53*
No. of reads (x 1000)		*3484*	*3965*	*2999*	*4125*	*2983*	*3134*	*3919*	*3451*	*3237*	*2642*
Average read length		*99*	*99*	*99*	*99*	*99*	*99*	*99*	*99*	*99*	*99*
Average insert size		*436*	*382*	*434*	*381*	*363*	*361*	*404*	*420*	*433*	*440*
N50 (Kbp)		*181*	*172*	*172*	*150*	*192*	*192*	*107*	*192*	*182*	*184*

Se: Salmonella enterica
NA: not analysed.

Annex 15. Accessing Genomes

	Participant		EQA provider	
Lab ID	**Cluster**	**Genome 1**	**QC Accepted**	**Cluster**
EQA provider	Yes	A cluster isolate (REF11)	Yes	Yes
19	Yes	Genome 1 cluster with 0 allele differences to the outbreak cluster representative and is therefore considered a part of the outbreak	Yes	Yes
36	Yes	1-2 allelic differences to the other outbreak strains (Enterobase Scheme in SeqSphere) sequence quality ok (avg. coverage, percent good targets)	Yes	Yes
49	Yes	Isolate is indistinguishable by cgMLST from representative isolate.	Yes	Yes
100	Yes	Genome 1 has an average coverage of 112X. The SNP distance to the representative isolate (Isolate 1 - 5210) is 3; as a consequence, Genome 1 is closely related to the representative isolate.	Yes	Yes
106	Yes	We have established a cluster distance threshold of 7 alleles difference. We would include this genome as part of the outbreak but adding a note that epidemological link is essential to consider the case as part of the outbreak	Yes	Yes
108	Yes	The isolate clusters within ST34 based on the coregenome (0 snp).	Yes	Yes
127	Yes	Using the cgMLST scheme, AD is less than 5 alleles.	Yes	Yes
129	Yes	Genome 1 had 8 allele differences (0,24%) to three of the isolates included in the cluster. Percentage of good targets was 97.7 (pass).	Yes	Yes
134	Yes	First: good quality of the reads assembled with: high coverage = 113 and percentage of targets found = 98,8 % second: AD of 1 against representative isolate 6 (N°5668)	Yes	Yes
135	Yes	It falls in the cluster with a distance of only 2 alleles different	Yes	Yes
142	Yes	We detect 7 allelic differences with the representative, which is within the cut-off of 7 loci differences for cluster detection.	Yes	Yes
147	Yes	1 allele difference from selected representative isolate 5064 (ST34 / CT 3173) good quality of sequence – 99.3% good targets identical CT (ST34 / CT 3173) to cluster isolate 5064	Yes	Yes
148	Yes	Genome 1 has the same HC5|302 as the outbreak cluster, and it is placed at 2 alleles difference from the reference isolate (ref. sequence would be Isolate 3, #5212).	Yes	Yes
149	Yes	Genome 1 is part of the cluster as it within <=7 AD cluster definition (distance 0 AD)	Yes	Yes
150	Yes	Salm_genome1 is within the same cluster as the outbreak strains (5466,5557,5835,5933) and has 0 SNP differences, it would be considered to be part of the cluster	Yes	Yes

	Participant		EQA provider	
Lab ID	**Cluster**	**Genome 2**	**Quality issue**	**Cluster**
EQA provider	Yes	A cluster isolate (REF12) with altered coverage (reduced to 10x)	Yes	Yes
19	No	Genome 2 is discarded from the analysis because of poor quality sequence. The average read coverage is too low, the N50 value is low, the number of assembled contigs is very high, the number of unidentified bases is high, and the core percent is too low.	Yes	-
36	No	The data set was excluded from the analysis due to poor data quantity/quality.	Yes	-
49	No	The quality of this sequence was not satisfactory (e.g. AvgReadCoverage=11) so was not included in the analysis.	Yes	-
100	No	Genome 2 has an average coverage of 10X. Sequence quality is too poor to be properly evaluated and conclusions cannot be made. It has 55 SNP when compared with the representative isolate. Genome 2 is not closely related to the representative isolate and it is not part of the cluster.	Yes	No
106	No	Outside the cluster in both PFGE and WGS analysis.	No	No
108	No	ST 34 to low average-coverage to cluster (10,15x), under quality-limits.	Yes	-
127	No	This sequence failed quality checking. We are not able to assign it to the cluster.	Yes	-
129	No	Genome 2 did not pass our QC. The percentage of good targets was only 43.3 % and average coverage was 10. Also the contig count of 3353suggested poor quality. Resequencing would have been requested.	Yes	-
134	No	The assembly has a low coverage of 10, but using mapping-based assembly in SeqSphere (BWA) 92,2 % of alleles were called, which is enough for comparing strains and taking a decision. An AD of 21 against the representative isolate excludes this isolate from the outbreak.	Yes	No
135	No	It falls out of the cluster with a distance of 109 alleles different	No	No
142	No	Low quality of the data: Very low coverage, leading to many loci that aren't called. Within the called genes, more than seven loci differences with the representative are present.	Yes	No
147	No	65 allele difference to cluster isolate 5064 (ST34 / CT 3173) low coverage – av.cov.ass. 5x 91.7% good targets --> we would recommend repeating sequencing of this sample different CT (ST34 / CT 3267) to cluster isolate 5064.	Yes	No
148	No	We could not analyse 'genome 2' because the N50 (7095) and the number of contigs (1331) did not pass the quality control of Enterobase.	Yes	-
149	No	Genome 2 did not pass the cluster comparison threshold of >90% good cgMLST targets. This sample would need to be re-sequenced before cluster comparison would be possible to determine if it was part of the outbreak cluster.	Yes	-
150	No	Salm_genome 2 quality was poor and below the cut-off our quality metrics for SNP typing due to not enough depth of coverage, the yield is poor and around 40, whereas the other genomes have yields of between 230-490. Therefore this would not be included in the analysis and the sequencing would need to be repeated for this strain.	Yes	-

	Participant		EQA provider	
Lab ID	Cluster	Genome 3	QC Accepted	Cluster
EQA provider	No	A non-cluster isolate (REF13). 31 allelic difference to the cluster.	Yes	No
19	No	Genome 3 differ with 31 allele differences to the outbreak cluster representative and is therefore not considered a part of the outbreak cluster.	Yes	No
36	No	>30 allelic differences to the other outbreak strains (Enterobase Scheme in SeqSphere) sequence quality ok (avg. coverage, percent good targets).	Yes	No
49	No	This isolate differs by 31 cgMLST alleles from the representative isolate.	Yes	No
100	No	Genome 3 has an average coverage of 69X. It has 85 SNP when compared with the representative isolate. Genome 3 is not closely related to the representative isolate and it is not part of the cluster.	Yes	No
106	Yes	We have established a cluster distance threshold of 7 alleles difference. This genome has only 1 allele difference with the representative genome. Highly related.	Yes	Yes
108	No	ST34 isolate does not cluster with the other genomes (60snp apart).	Yes	No
127	No	Using the cgMLST scheme, AD is more than 10 alleles. Therefore, the genome is considered not to be a member of the cluster.	Yes	No
129	No	Genome 3 differed from isolate 5871 by 47 alleles. QC was ok for this genome.	Yes	No
134	No	AD of 31 against the representative isolate.	Yes	No
135	No	It falls out of the cluster with a distance of 63 alleles different.	Yes	No
142	No	More than 7 loci differences with representative.	Yes	No
147	No	31 allele difference to cluster isolate 5064 (ST34 / CT 3173) different CT (ST34 / CT 3229) to cluster isolate 5064.	Yes	No
148	No	Genome 3 only shares HC10|2 with the outbreak cluster, and it is at 54 alleles distance from the reference isolate. We would consider this distance as related, but not part of an ongoing outbreak, especially for the monophasic variant of S. Typhimurium (which is the case).	Yes	No
149	No	Genome 3 is not part of the cluster as it is >7 AD from the cluster (distance 31 AD).	Yes	No
150	Yes	Salm_genome3 is within 25 SNPs of the cluster strains and would not be included within the cluster.	Yes	No

	Participant		EQA provider	
Lab ID	Cluster	Genome 4	Contamination	Cluster
EQA provider	No	A non-cluster isolate (REF14) mixed with a *Escherichia coli* (approx. 10%).	Yes	No
19	No	Genome 2 is discarded from the analysis because of likely contamination. Many unclassified reads were identified in the species analysis and the genome is almost double the size than expected. Furthermore, the N50 value is low, the number of assembled contigs are very high, the number of loci with multiple consensus and number of unidentified bases is high.	Yes	-
36	No	>30 allelic differences to the other outbreak strains (Enterobase Scheme in SeqSphere) sequence quality ok (avg. coverage, percent good targets).	No	No
49	No	This isolate differs by 39 cgMLST alleles from the representative isolate.	No	No
100	No	Genome 4 has an average coverage of 86X. It has 167 SNP when compared with the representative isolate. Genome 4 is not closely related to the representative isolate and it is not part of the cluster.	No	No
106	Yes	With both genotyping methods identical to the "outbreak representative".	No	Yes
108	No	ST4431 isolate does not cluster with the other genomes (>60 snp from the ST34 cluster). Quality-limits contaminated with E.coli to large genome size.	Yes	No
127	No	Using the cgMLST scheme, AD is more than 10 alleles. Therefore the genome is considered not to be a member of the cluster. Contamination was recognised using EnteroBase QAssembly pipeline.	Yes	No
129	No	Genome 4 differed from isolate 5871 by 53 alleles and had different ST. QC was ok for this genome.	No	No
134	No	AD of 36 against the representative isolate.	No	No
135	No	It falls out of the cluster with a distance of 61 alleles different.	No	No
142	No	Data are contaminated with E. coli genome (11,4%). + More than 7 loci differences with representative.	Yes	No
147	No	30 allele difference to cluster isolate 5064 (ST34 / CT 3173) genome size too big ~10Mbases (Salmonella ~ 5Mbases) 92% good targets --> we would recommend to repeat sequencing of this sample different CT (ST34 / CT 3268) to cluster isolate 5064.	Yes	No
148	No	Genome 4 is only at the same HC20|2 than the outbreak cluster, and at 57 alleles from the reference isolate. We would consider that as too far for being part of a 4,5,12:i:- outbreak.	No	No
149	No	Genome 4 is not part of the cluster as it is >7 AD from the cluster (distance 36 AD).	No	No
150	No	Salm_genome 4 is within 10 SNPs of the strain 5728 which is not part of the same cluster as the outbreak strains and is within 25 SNPS of the cluster strains and would not be included within the cluster.	No	No

Annex 16. EQA-9 laboratory questionnaire

This is a preview of all the fields and questions available.

Please keep in mind that, depending on your answers in the questionnaire, you will not necessarily have to answer all the questions.

1. Salmonella EQA-10 2019-2020

Dear Participant
Welcome to the tenth External Quality Assessment (EQA-10) scheme for typing of Salmonella in 2019-2020.
Please note that most of the fields are required to be filled in before the submission can be completed.
Any comments can be written at the end of the form.
You are always welcome to contact us at Salm.eqa@ssi.dk.

Please start by filling in your country, your Laboratory name and your LAB_ID.

Available options in this submission form include:

- - Click "Options" and "Pause" to save your results and finish at a later time (using the same link)
- - Click "Options" and "Print" to print your answers. This can be done at any time, but before pressing "Submit results"
- - Click "Previous" to go back to the questions you have already answered
- - Click "Options" and "Go to.." to go back to a specific page number

Note: After pressing "Submit results" you will not be able to review your results.

2. Country

(State one answer only)

☐ Austria
☐ Belgium
☐ Czechia
☐ Denmark
☐ Estonia
☐ Finland
☐ France
☐ Germany
☐ Greece
☐ Hungary
☐ Iceland
☐ Ireland
☐ Italy
☐ Latvia
☐ Lithuania
☐ Luxembourg
☐ Norway
☐ Romania
☐ Scotland
☐ Serbia
☐ Slovak Republic
☐ Slovenia
☐ Spain
☐ Sweden
☐ The Netherlands
☐ Turkey
☐ United Kingdom

3. Institute name

4. Laboratory name

5. Laboratory ID

Consisting of country code (two letters) Lab ID on the vial e.g. DK_SSI

6. E-mail

7. Multiple-Locus Variable number of tandem repeats Analysis (MLVA)

8. Submitting results

(State one answer only)

☐ Online here (please fill in the isolate ID´s in the following section) - Go to 9
☐ Did not participate in the MLVA part - Go to 14

9. Select method

(State one answer only)

☐ S. Typhimurium and S. Enteritidis - Go to 10
☐ Only S. Typhimurium - Go to 10
☐ Only S. Enteritidis - Go to 12

10. MLVA isolate IDs

Please enter the MLVA isolate ID (4 digits)

	Isolate 1	Isolate 2	Isolate 3	Isolate 4	Isolate 5	Isolate 6	Isolate 7	Isolate 8	Isolate 9	Isolate 10
S. Typhimurium										

11. Results for MLVA S. Typhimurium - Allele profile

Please use -2 for not detected

	STTR9	STTR5	STTR6	STTR10	STTR3
Isolate 1					
Isolate 2					
Isolate 3					
Isolate 4					
Isolate 5					
Isolate 6					
Isolate 7					
Isolate 8					
Isolate 9					
Isolate 10					

12. MLVA isolate IDs

Please enter the MLVA isolate ID (4 digits)

	Isolate 1	Isolate 2	Isolate 3	Isolate 4	Isolate 5	Isolate 6	Isolate 7	Isolate 8	Isolate 9	Isolate 10
S. Enteritidis										

13. Results for MLVA S. Enteritidis - Allele profile

Please use -2 for not detected

	SENTR7	SENTR5	SENTR6	SENTR4	SE-3
Isolate 1					
Isolate 2					
Isolate 3					
Isolate 4					
Isolate 5					
Isolate 6					
Isolate 7					
Isolate 8					
Isolate 9					
Isolate 10					

14. Submitting Cluster results

(State one answer only)

☐ Cluster analyses based on PFGE / MLVA / WGS - Go to 15
☐ Did not participate in the Cluster part - Go to 134

15. Cluster isolate IDs

Please enter the cluster isolate ID (4 digits)

	Cluster isolate ID
Isolate 1	
Isolate 2	
Isolate 3	
Isolate 4	
Isolate 5	
Isolate 6	
Isolate 7	
Isolate 8	
Isolate 9	
Isolate 10	

16. Submitting Cluster results

(State one answer only)

☐ Cluster analysis based on PFGE - Go to 17
☐ Do not wish to submit any cluster results based on PFGE analysis - Go to 22

17. Cluster analysis based on PFGE data

18. Please list the ID for the isolates included in the cluster of closely related isolates detected by PFGE (bands >33kb used):

Please use semicolon (;) to separate the IDs

19. Select a representative isolate with the cluster profile detected by PFGE:

Indicate the isolate ID

20. Total number of bands (>33kb) in the selected representative cluster isolate

21. Results for cluster analysis - PFGE (XbaI)

Please use 9999 for not analysed

	Total number of bands (>33kb)	Number of bands with same/shared position as the profile of the selected cluster isolate (>33kb)
Isolate 1		
Isolate 2		
Isolate 3		
Isolate 4		
Isolate 5		
Isolate 6		
Isolate 7		
Isolate 8		
Isolate 9		
Isolate 10		

22. Submitting cluster results

(State one answer only)

☐ Cluster analysis based on MLVA – Go to 23
☐ Do not wish to submit any cluster results based on MLVA analysis – Go to 33

23. Cluster analysis based on MLVA data

24. Please list the ID for the isolates included in the cluster of closely related isolates detected by MLVA:

Please use semicolon (;) to separate the ID´s

25. MLVA scheme used:

Please indicate serovar and/or protocol

26. Please list the loci in scheme used

27. Locus 1:

28. Locus 2:

29. Locus 3:

30. Locus 4:

31. Locus 5:

32. Results for cluster analysis (MLVA) - Allele profile

Please use -2 for not detected, and 9999 for not analysed

	Locus 1	Locus 2	Locus 3	Locus 4	Locus 5
Isolate 1					
Isolate 2					
Isolate 3					
Isolate 4					
Isolate 5					
Isolate 6					
Isolate 7					
Isolate 8					
Isolate 9					
Isolate 10					

33. Submitting cluster results

(State one answer only)

☐ Cluster analysis based on WGS data – Go to 34
☐ Do not wish to submit any cluster results based on WGS data - Go to 134

34. Cluster analysis based on WGS data

35. Please select the analysis used to detect the cluster on data derived from WGS

As basis for the cluster detection only one data analysis can be reported. If more than one analysis is performed please report later in this submission

(State one answer only)

☐ SNP based – Go to 37
☐ Allele-based – Go to 44
☐ Other – Go to 36

36. If another analysis is used please describe your approach: - Go to 51

37. Please report the used SNP-pipeline

(reference if publicly available or in-house pipeline)

38. Please select the approach used for the SNP analysis

(State one answer only)

☐ Reference based – Go to 39
☐ Assembly based – Go to 42

39. Reference genome used

Please indicate Multil-locus Sequence Type (e.g. ST34) and isolate ID (e.g. one of the isolates from the current cluster, ID of a public reference strain or an in-house isolate)

40. Please indicate the read mapper used

(e.g. BWA, Bowtie2)

41. Please indicate the variant caller used

(e.g. SAMtools, GATK)

42. Please indicate the assembler used

(e.g. SPAdes, Velvet)

43. Please specify the variant caller used

(e.g. NUCMER)

44. Please select tools used for the allele analysis

(State one answer only)

☐ BioNumerics – Go to 46
☐ SeqSPhere – Go to 46
☐ Enterobase – Go to 46
☐ Other – Go to 45

45. If another tool is used please enter here:

46. Please indicate allele calling method:

(State one answer only)

☐ Assembly based and mapping based – Go to 47
☐ Only assembly based – Go to 47
☐ Only mapping based – Go to 48

47. Please indicate the assembler used

(e.g. SPAdes, Velvet)

48. Please select scheme used for the allele analysis

(State one answer only)

☐ Applied Math (wgMLST) – Go to 50
☐ Applied Math (cgMLST/Enterobase) – Go to 50
☐ Enterobase (cgMLST) – Go to 50
☐ Other – Go to 49

49. If another scheme (e.g. in-house) is used, please give a short description

50. Please report the number of loci in the used allelic scheme

51. Cluster detected by analysis on data derived from WGS

On this page you have to report the results for the cluster detected by the selected analysis (e.g. SNP based). If another additional analysis (e.g. allele-based or another SNP based analysis) is performed please report results later, but you will not be asked to submit the ID´s for isolates in the cluster detected with the additional analysis.

52. Please list the IDs for the isolates included in the cluster of closely related isolates:

Please use semicolon (;) to separate IDs

53. Select a representative isolate in the cluster

Indicate the isolate ID

54. Results for cluster analysis (e.g. SNP or allele-based)

Please use 9999 for not analysed

	ST	Distance (e.g. SNP) to the selected cluster isolate
Isolate 1		
Isolate 2		
Isolate 3		
Isolate 4		
Isolate 5		
Isolate 6		
Isolate 7		
Isolate 8		
Isolate 9		
Isolate 10		

55. Analysis of the EQA provided genomes

In this section, the results of the provided genomes (1-4) can be reported, mimicking an outbreak situation

56. In an outbreak situation, would you consider the EQA provided genome 1 a part of the cluster of closely related isolates?

(State one answer only)

☐ Yes, genome 1 is a part of the cluster
☐ No, genome 1 is NOT a part of the cluster

57. Please list the arguments behind the decision: Why genome 1 is a part of the cluster or why it is not part of the cluster.

58. In an outbreak situation, would you consider the EQA provided genome 2 a part of the cluster of closely related isolates?

(State one answer only)

☐ Yes, genome 2 is a part of the cluster
☐ No, genome 2 is NOT a part of the cluster

59. Please list the arguments behind the decision: Why genome 2 is a part of the cluster or why it is not part of the cluster.

60. In an outbreak situation, would you consider the EQA provided genome 3 a part of the cluster of closely related isolates?

(State one answer only)

☐ Yes, genome 3 is a part of the cluster
☐ No, genome 3 is NOT a part of the cluster

61. Please list the arguments behind the decision: Why genome 3 is a part of the cluster or why it is not part of the cluster.

62. In an outbreak situation, would you consider the EQA provided genome 4 a part of the cluster of closely related isolates?

(State one answer only)

☐ Yes, genome 4 is a part of the cluster
☐ No, genome 4 is NOT a part of the cluster

63. Please list the arguments behind the decision: Why genome 4 is a part of the cluster or why it is not part of the cluster.

64. Would you like to add results performed with another additional analysis on the data derived from the WGS?

e.g. if SNP based results are submitted you can also report allele-based results or results from a second SNP analysis

(State one answer only)

☐ Yes – Go to 65
☐ No – Go to 104

65. Please select the additional analysis used on data derived from WGS

(State one answer only)

☐ SNP based – Go to 67
☐ Allele-based – Go to 74
☐ Other – Go to 66

66. If another analysis is used please describe your approach: - Go to 81

67. Please report the used SNP-pipeline

(reference if publicly available or in-house pipeline)

68. Please select the approach used for the SNP analysis

□ Reference based – Go to 69
□ Assembly based – Go to 72

69. Reference genome used

Please indicate Multi-locus Sequence Type (e.g. ST34) and isolate ID (e.g. one of the isolates from the current cluster, ID of a public reference strain or an in-house isolate)

70. Please indicate the read mapper used

(e.g. BWA, Bowtie2)

71. Please indicate the variant caller used

(e.g. SAMtools, GATK)

72. Please indicate the assembler used

(e.g. SPAdes, Velvet)

73. Please specify the variant caller used

(e.g. NUCMER)

74. Please select tool used for the allele analysis

(State one answer only)

☐ BioNumerics – Go to 76
☐ SeqSPhere – Go to 76
☐ Enterobase – Go to 76
☐ Other – Go to 75

75. If another tool is used please list here:

76. Please indicate allele calling method:

(State one answer only)

☐ Assembly based and mapping based - Go to 77
☐ Only assembly based - Go to 77
☐ Only mapping based - Go to 78

77. Please indicate the assembler used

(e.g. SPAdes, Velvet)

78. Please select scheme used for the allele analysis

(State one answer only)

☐ Applied Math (wgMLST) – Go to 80
☐ Applied Math (cgMLST/Enterobase) – Go to 80
☐ Enterobase (cgMLST) – Go to 80
☐ Other – Go to 79

79. If another scheme (e.g. in-house) is used, please give a short description

80. Please report the number of loci in the used allelic scheme

81. Additional analysis on data derived from WGS

82. Select a representative isolate in the cluster detected by the additional analysis

Indicate the isolate ID

83. Results for the additional cluster analysis (e.g. SNP or allele-based)

Please use 9999 for not analysed

	ST	Distance (e.g. SNP) to the selected cluster isolate
Isolate 1		
Isolate 2		
Isolate 3		
Isolate 4		
Isolate 5		
Isolate 6		
Isolate 7		
Isolate 8		
Isolate 9		
Isolate 10		

84. Would you like to add results performed with a third analysis on the data derived from the WGS?

e.g. if SNP based results are submitted you can also report allele-based results or results from a second SNP analysis

(State one answer only)

☐ Yes – Go to 85
☐ No – Go to 104

85. Please select the third analysis used on data derived from WGS

(State one answer only)

☐ SNP based – Go to 87
☐ Allele-based – Go to 94
☐ Other – Go to 86

86. If another analysis is used please describe your approach: - Go to 101

87. Please report the used SNP-pipeline

(reference if publicly available or in-house pipeline)

88. Please select the approach used for the SNP analysis

☐ Reference based – Go to 89
☐ Assembly based – Go to 92

89. Reference genome used

Please indicate Multi-locus Sequence Type (e.g. ST34) and isolate ID (e.g. one of the isolates from the current cluster, ID of a public reference strain or an in-house isolate)

90. Please indicate the read mapper used

(e.g. BWA, Bowtie2)

91. Please indicate the variant caller used

(e.g. SAMtools, GATK)

92. Please indicate the assembler used

(e.g. SPAdes, Velvet)

93. Please specify the variant caller used

(e.g. NUCMER)

94. Please select tool used for the allele analysis

(State one answer only)

☐ BioNumerics - Go to 96
☐ SeqSPhere - Go to 96
☐ Enterobase - Go to 96
☐ Other - Go to 95

95. If another tool is used please enter here:

96. Please indicate allele calling method:

(State one answer only)

☐ Assembly based and mapping based - Go to 97
☐ Only assembly based - Go to 97
☐ Only mapping based - Go to 98

97. Please indicate the assembler used

(e.g. SPAdes, Velvet)

98. Please select scheme used for the allele analysis

(State one answer only)

☐ Applied Math (wgMLST) - Go to 100
☐ Applied Math (cgMLST/Enterobase) - Go to 100
☐ Enterobase (cgMLST) - Go to 100
☐ Other - Go to 99

99. If another scheme (e.g. in-house) is used, please give a short description

100. Please report the number of loci in the used allelic scheme

101. Third analysis on data derived from WGS

102. Select a representative isolate in the cluster detected by the third analysis

Indicate the isolate ID

103. Results for the third cluster analysis (e.g. SNP or allele-based)

Please use 9999 for not analysed

	ST	Distance (e.g. SNP) to the selected cluster isolate
Isolate 1		
Isolate 2		
Isolate 3		
Isolate 4		
Isolate 5		
Isolate 6		
Isolate 7		
Isolate 8		
Isolate 9		
Isolate 10		

104. Additional questions to the WGS part

105. Where was the sequencing performed

(State one answer only)

☐ In own laboratory
☐ Externally

106. Protocol used to prepare the library for sequencing:

(State one answer only)

☐ Commercial kits - Go to 107
☐ Non-commercial kits - Go to 109

107. Please indicate name of commercial kit:

108. If relevant please list deviation from commercial kit shortly in few bullets: - Go to 110

109. For non-commercial kit please indicate a short summary of the protocol:

110. The sequencing platform used

(State one answer only)

☐ Ion Torrent PGM - Go to 112
☐ Ion Torrent Proton - Go to 112
☐ Genome Sequencer Junior System (454) - Go to 112
☐ Genome Sequencer FLX System (454) - Go to 112
☐ Genome Sequencer FLX+ System (454) - Go to 112
☐ PacBio RS - Go to 112
☐ PacBio RS II - Go to 112
☐ HiScanSQ - Go to 112
☐ HiSeq 1000 - Go to 112
☐ HiSeq 1500 - Go to 112
☐ HiSeq 2000 - Go to 112
☐ HiSeq 2500 - Go to 112
☐ HiSeq 4000 - Go to 112
☐ Genome Analyzer Iix - Go to 112
☐ MiSeq - Go to 112
☐ MiSeq Dx - Go to 112
☐ MiSeq FGx - Go to 112
☐ ABI SOLiD - Go to 112
☐ NextSeq - Go to 112
☐ MinION (ONT) - Go to 112
☐ Other - Go to 111

111. If another platform is used please list here:

112. Criteria used to evaluate the quality of sequence data

In this section you can report criteria used to evaluate the quality of sequence data.

Please first reply on the use of 5 selected criteria, which were the most frequently reported by the participants in the *Salmonella* EQA-8 scheme, 2017.

Next you will be asked to report 5 additional criteria of your own choice.

For each criteria please also report the threshold or procedure used to evaluate the current criteria.

113. Did you use confirmation of species to evaluate the quality of sequence data?

(State one answer only)

☐ Yes
☐ No – Go to 115

114. Procedure used to evaluate confirmation of organism:

115. Did you use coverage to evaluate the quality of sequence data?

(State one answer only)

☐ Yes
☐ No – Go to 117

116. Procedure or threshold used for coverage:

117. Did you evaluate assembly quality?

(State one answer only)

☐ Yes
☐ No – Go to 119

118. Procedure used to evaluate assembly quality:

119. Did you use assembly length to evaluate the quality of sequence data?

(State one answer only)

☐ Yes
☐ No – Go to 121

120. Procedure or threshold used for assembly length:

121. Did you evaluate allele calling result?

(State one answer only)

☐ Yes
☐ No – Go to 123

122. Procedure used to evaluate allele calling:

123. Other criteria used to evaluate the quality of sequence data

Please list up to 5 additional criteria (e.g. N50, read length, contamination)

124. Other criteria used to evaluate the quality of sequence data - additional criteria 1:

125. Threshold or procedure used to evaluate the additional criteria 1:

126. Other criteria used to evaluate the quality of sequence data - additional criteria 2:

127. Threshold or procedure used to evaluate the additional criteria 2:

128. Other criteria used to evaluate the quality of sequence data - additional criteria 3:

129. Threshold or procedure used to evaluate the additional criteria 3:

130. Other criteria used to evaluate the quality of sequence data - additional criteria 4:

131. Threshold or procedure used to evaluate the additional criteria 4:

132. Other criteria used to evaluate the quality of sequence data - additional criteria 5:

133. Threshold or procedure used to evaluate the additional criteria 5:

134. Comment(s):

e.g. remarks to the submission, the data analyses or the laboratory methods

135. Thank you for your participation

Thank you for filling out the Submission form for the Salmonella EQA-10.

For questions, please contact salm.eqa@ssi.dk or phone +45 3268 8341 +45 3268 8372.

We highly recommend to document this Submission form by printing it. You will find the Print option after pressing the "Options" button.

Important: After pressing "Submit results" you will no longer be able to edit or print your information.

For final submission, remember to press "Submit results" after printing.

European Centre for Disease Prevention and Control (ECDC)

Gustav III:s Boulevard 40, 16973 Solna, Sweden

Tel. +46 858601000
Fax +46 858601001
www.ecdc.europa.eu

An agency of the European Union
www.europa.eu

Subscribe to our publications
www.ecdc.europa.eu/en/publications

Contact us
publications@ecdc.europa.eu

Follow us on Twitter
@ECDC_EU

Like our Facebook page
www.facebook.com/ECDC.EU

ECDC is committed to ensuring the transparency and independence of its work

In accordance with the *Staff Regulations for Officials and Conditions of Employment of Other Servants of the European Union* and the *ECDC Independence Policy*, ECDC staff members shall not, in the performance of their duties, deal with matters in which they may, directly or indirectly, have a personal interest that could impair their independence. Declarations of interest must be received from any prospective contractor before a contract can be awarded.
www.ecdc.europa.eu/en/aboutus/transparency

Publications Office
of the European Union