TECHNICAL REPORT

Eleventh external quality assessment scheme for *Salmonella* typing

www.ecdc.europa.eu

ECDC TECHNICAL REPORT

Eleventh 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, Louise Gade Dahl, Gitte Sørensen, 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. Eleventh external quality assessment scheme for *Salmonella* typing. Stockholm: ECDC; 2021.

Stockholm, October 2021

ISBN 978-92-9498-551-4
DOI 10.2900/982538
Catalogue number TQ-07-21-017-EN-N

Contents

Figures

Tables

Abbreviations

AD	Allele differences
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 national reference laboratories
QC	Qualitative control
SNP	Single nucleotide polymorphism
SNV	Single-nucleotide variant based on cgMLST
SSI	Statens Serum Institut
ST	7-gene Multi Locus 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 11th round of the external quality assessment (EQA-11) scheme for typing of *Salmonella enterica* subsp. *enterica* organised for the national public health national 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-11 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 2019, with a notification rate of 20.0 cases per 100 000 population. The total number of reported cases was 90 105 [3]. Since 2007, ECDC has been responsible for 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 for 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 MLVA for *S.* Typhimurium (STm). Since EQA-8, the participants could participate in MLVA for *S.* Enteritidis (SE). From EQA-9, the separate PFGE part was excluded. Since then, PFGE has only been included into the cluster analyses part where the ability of identifying a cluster based on molecular typing by either using PFGE, MLVA and/or whole genome sequencing (WGS) derived data was assessed.

The objectives of the EQA-11 scheme were to assess the quality of data and comparability of molecular typing analysis results produced by NPHRLs in 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 respective MLVA methods and the cluster analysis included ten *S.* Enteritidis ST11 isolates.

Twenty-one laboratories signed up and 20 completed the exercise. This is a decrease from EQA-8 (N=23) by 13%. It is unknown if the removal of the PFGE part (gel quality and analysis) was the cause of this. Most laboratories (N=19) participated in the molecular typing-based cluster analysis. Out of the 20 laboratories participating in EQA-11, 14 (70%) performed molecular typing-based cluster analysis using WGS-derived data, which is one less compared to EQA-10.

In total, eight laboratories participated in the *S.* Typhimurium and *S.* Enteritidis MLVA analysis, which is slightly fewer compared with EQA-8, when 10 and eight laboratories participated, respectively. The performance level was high for both analyses (95% and 98% respectively) and within the range of the previous years.

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. The cluster of closely related *S.* Enteritidis ST11 isolates could be investigated by PFGE, MLVA and WGS-derived data. The expected cluster was based on a predefined categorisation by the organiser and contained five isolates based on WGS-derived data.

The number of laboratories performing WGS has stabilised as no new laboratories performed WGS this year and over time, the use of PFGE and MLVA has become less frequent. Fewer laboratories performed cluster analysis with more than one method and only one laboratory performed cluster analysis using all three methods in EQA-11.

Six laboratories used PFGE for cluster analysis and for three participants; PFGE was the only cluster analysis method. None of the six laboratories were able to identify the correct cluster using PFGE. Four laboratories used MLVA for cluster analysis and two laboratories only used MLVA for the cluster analysis. All ten test isolates had the same MLVA profile and therefore it was not possible for the participants to identify the correct cluster of closely related isolates using this less discriminatory method.

The performance among the 14 participants using WGS derived data was very high, as 13 (93%) correctly identified the cluster of closely related isolates and the use of allele-based analysing method (mainly core genome multilocus sequence typing (cgMLST)/Enterobase scheme) was dominant.

An assessment of six EQA provided genomes was an additional part to the molecular typing-based cluster analysis. 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 were asked to assess six genomes, which were modified by the EQA provider in order to give a realistic view of different quality issues. All participants (except one for one of the genomes) successfully identified the three genomes of high quality as either a cluster isolate (one genome) or a non-cluster isolate (two genomes). Two poor quality genomes with contamination of a different *Salmonella* ST34 (20%) were identified by all the participants. Ten of 14 participants identified the genome with 10% *Citrobacter* -contamination.

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 EU/EEA networks 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 (EQAs) are an essential part of laboratory quality management and use an external contractor 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 disease networks. EQAs aim to identify areas for improvement in laboratory diagnostic capacities relevant for epidemiological surveillance of communicable diseases as 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 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-11.

1.2 Surveillance of non-typhoidal salmonellosis

In 2019, non-typhoidal salmonellosis (later 'salmonellosis') was the second-most commonly reported zoonotic disease in the EU, with 90 105 cases reported and an EU notification rate of 20.0 cases per 100 000 population, which was at the same level as 2018. The trend for salmonellosis in humans has been stable over the last five years after a long period of a declining trend. *S.* Enteritidis caused the vast majority (72.4%) of the salmonellosis food-borne outbreaks [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 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 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 and contribute to global outbreak investigations
- detect the emergence of new and/or evolving pathogenic strains
- support investigations to trace the source of an outbreak and identify new risk factors
- aid the study of particular pathogen's 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.

1.3 Objectives

EQA schemes offer quality support to those NPRLs that perform molecular typing-enhanced surveillance and those implementing it into their surveillance system at national level.

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

The *Salmonella* EQA-11 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-11 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 participants were 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 the results will be summarised in this report.

2. Study design

2.1 Organisation

The *Salmonella* EQA-11 was funded by ECDC and arranged by SSI following the requirements in ISO/IEC 17043:2010 [6]. The EQA-11 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-9 onwards, the quality assessment part with PFGE was no longer a part of the EQA scheme. EQA-11 was conducted between September 2020 and December 2020.

Invitations were emailed to ECDC contact points in FWD-Net (to 26 countries, which nominated laboratories to participate in the EQA rounds from 2017–2020) by 18 May 2020, with a deadline to respond of 8 June 2020. In addition, invitations were sent to EU candidate and potential candidate countries Serbia, Turkey, and Kosovo[i], which signed up to the *Salmonella* EQA rounds in 2017-2020. As many EU countries were still struggling with the COVID-19 pandemic and laboratory staff were handling COVID-19 samples, the EQA shipment and deadlines were postponed to the fall of 2020.

Twenty-one NPHRLs in the EU/EEA and EU candidate countries accepted the invitation to participate, but only 20 submitted results (Annex 1). This was an increase from last year (EQA-10) where 19 countries participated. In Annex 2, details of participation in EQA-10 and EQA-11 are listed to give an overview of the trend in the number of participants.

The EQA test isolates were sent to participants on 2 September 2020. Participants were asked to submit their results by 22 November 2020 using the online form (Annex 15).

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 the genomes were available on the online site.

2.2 Selection of test isolates/genomes

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

- represent both commonly and rare reported isolates in Europe
- remain stable during the preliminary test period at the organising laboratory
- include repeat isolates from EQA-4/8 through 11
- include genetically closely related isolates.

The 36 selected isolates were analysed using the methods in the EQA (MLVA and/or 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 3-4). The 10 *S.* Enteritidis isolates for cluster analysis were selected to include isolates with different (or varying) relatedness and comprised identical 7-gene Multi Locus Sequence Types (ST)

[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.

(ST11). For the additional genomes, three were altered; one with reduced coverage, one mixed with 10% *Citrobacter* and one mixed with 20% *Salmonella* ST34. The last three were genomes with acceptable quality of reads and one provided as a Fasta file.

Table 1. Serovars of test isolates/genomes

Method	Number of test isolates/genomes	Serovars/ST´s	Annex
MLVA *S.* Typhimurium	10 isolates STm1-10	Typhimurium *STm2 (3-13-NA-NA-211), *STm3 (3-12-9-NA-211)	3
MLVA *S.* Enteritidis	10 isolates SE1-10	Enteritidis **SE9 (3-11-4-4-1), **SE10 (1-10-7-3-2)	4
Cluster analysis	10 isolates REF1-10 6 genomes REF1, REF3/REF10#, REF9	S. Enteritidis (ST11) *S.* Enteritidis ST11 (modified genomes: one with reduced coverage, one contaminated with 10% *Citrobacter* and one contaminated with 20% *Salmonella* ST34)	5-6; 8–11 14

**: "repeat isolates" included in EQA-4 to -11.*
***: "repeat isolates" included in EQA-8 to -11.*
#: Modified by the EQA provider in different variants
NA: designates a locus not present (-2 by submission, Annex 3 and 4).

2.3 Distribution of isolates/genomes

All 30 test isolates were blinded and shipped on 2 September 2020. 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 e-mail on 2 September 2020 as an extra precaution. Fourteen participants received their dispatched isolates within one day, six within two days and only one participant received the isolates after 16 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 the 22 September 2020, instructions for the submission of results 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.

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 allelic profiles were asked to be submitted using the online submission form. The code -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 was 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 15). 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 in the online form and uploaded the raw reads (FASTQ files) to the secure FTP site. The laboratories had the possibility to submit the 7-gene Multi Locus Sequence Types (ST) of isolates in the cluster analysis and were also asked to report the number of loci in the used allelic scheme for cluster analysis and/or the name of the used SNP pipeline.

In addition, each participant could assess extra genomes (some of them 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-11 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 *S.* Enteritidis ST11 isolates contained five isolates based on WGS derived data in an allele-based analysis (cgMLST, [7]) and a SNP analysis [8], which showed at most 2 AD or 1 SNP distance between any two isolates in the cluster. The cluster categorisation is based on WGS data and the correct cluster delineation was difficult to obtain if less discriminatory methods were used. The evaluation of the PFGE cluster analysis was based on including at least all the WGS defined cluster isolates. However, it was not possible for the participants to identify the correct cluster using MLVA, as all ten test isolates had the same profile: 3-10-5-4-1. The characteristics of the test isolates and reported results are listed in Annex: 5; 9 and 11.

The participant's descriptions of the additional genomes are listed in Annex 14. This analysis was only listed in the individual reports, but results are commented in this report.

Individual evaluation reports were distributed to the participants in January 2021 and certificates of attendance April 2021. 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, because they were either missing or the files were incomplete. Additionally, the EQA provider contacted laboratory 106 after the submission deadline to get the full data of the scheme used (number of loci).

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 21 participants who signed up for the EQA, 20 managed to complete and submit their results. The laboratory that did not submit results gave no reason for not participating, despite signing up and receiving isolates.

Eight (40%) laboratories completed MLVA (STm and SE). Nineteen laboratories (95%) participated in the cluster analysis part and most of them (14, 70%) reported cluster analysis results based on WGS, whereas six laboratories (30%) reported PFGE-derived results, and four laboratories (20%) reported MLVA-derived results. Only two participants (10%) reported the cluster using only MLVA and three participants (15%) reported the cluster using only PFGE. Ten participants (50%) reported the cluster using only WGS and one laboratory reported the cluster using all three methods.

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

	MLVA		Cluster analysis							All
	STm +SE	Total	PFGE only	MLVA only	WGS only	PFGE + WGS	PFGE + MLVA + WGS	MLVA + WGS	Total	Total
Number of participants	8	8	3	2	10	2	1	1	19	20
Percentage of participants	100	40*	16	11	53	11	5	5	95*	-

**: Percentage of the total number of participating laboratories (20)*
STm: S. Typhimurium
SE: S. Enteritidis

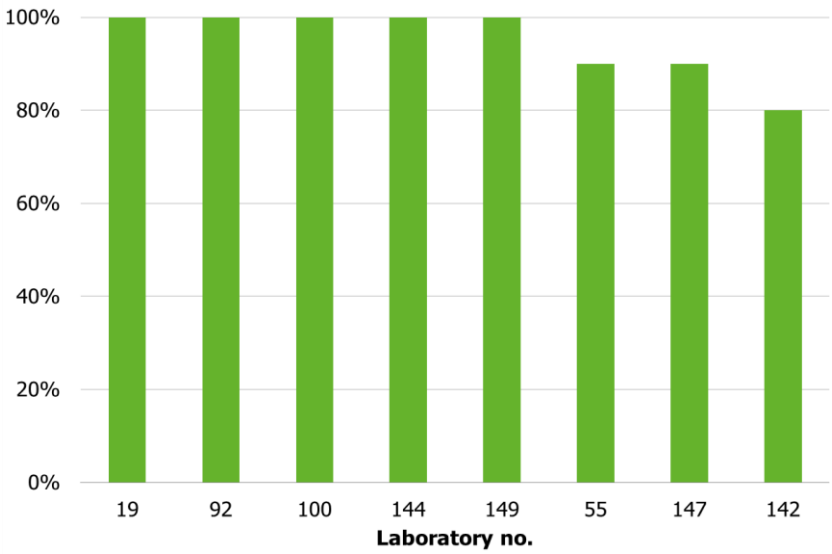
3.2 Multiple locus variable number of tandem repeats analysis

In total, eight laboratories (40%) participated in the MLVA part of the EQA and all of these participated in both the MLVA for *S.* Typhimurium and for *S.* Enteritidis (Annex 3 and 4). Two laboratories (108 and 135), which participated earlier, did not participate in this round and laboratory 149 participated again after a break of two years. One laboratory (92) participated only in the MLVA for *S.* Enteritidis last year, but in the MLVA for both *S.* Typhimurium and *S.* Enteritidis in EQA-11.

3.2.1 MLVA for *S.* Typhimurium

Eight out of the 20 participants in EQA-11 (40%) performed the MLVA typing of *S.* Typhimurium, and five of these (63%) reported the correct allelic profiles for all 10 test isolates (Figure 1). Three participants did not assign correct MLVA profiles for some of the test isolates given an average score of 95%. Laboratory 142 had the most errors, reporting a fragment in an absent locus (STTR6) in the test isolate STm8 and furthermore failed to report a fragment in STTR5 for isolate STm5. Laboratory 55 and 147 both had one error, reporting a fragment in an absent loci (STTR10) for isolates STm4 and STm1 respectively.

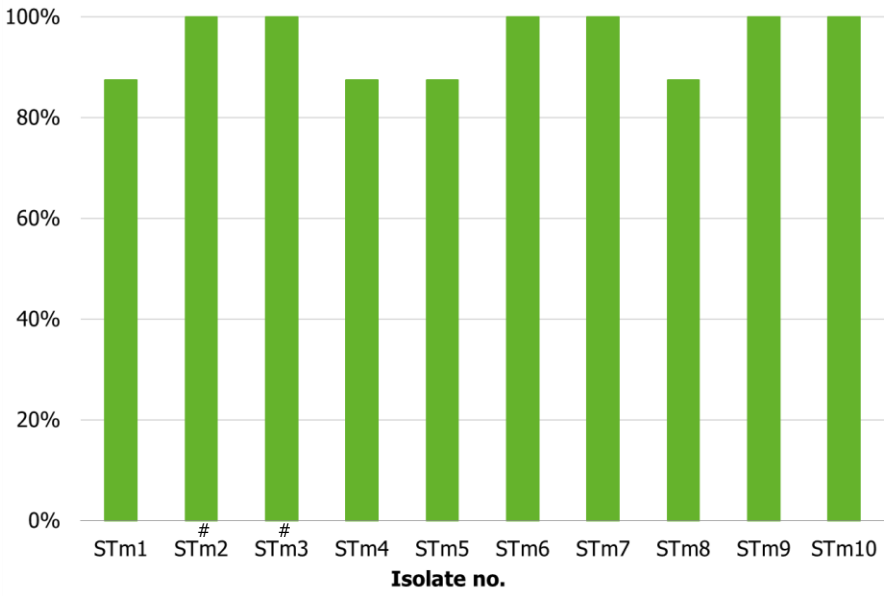
Figure 1. Participant scores for MLVA typing of the 10 *S.* Typhimurium test isolates



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. All participants reported the correct MLVA profile for six of the 10 *S.* Typhimurium test isolates. No common isolate characteristics caused the problems (Annex 3) as the four incorrect MLVA profiles concerned different isolates (STm1, 4, 5 and 8).

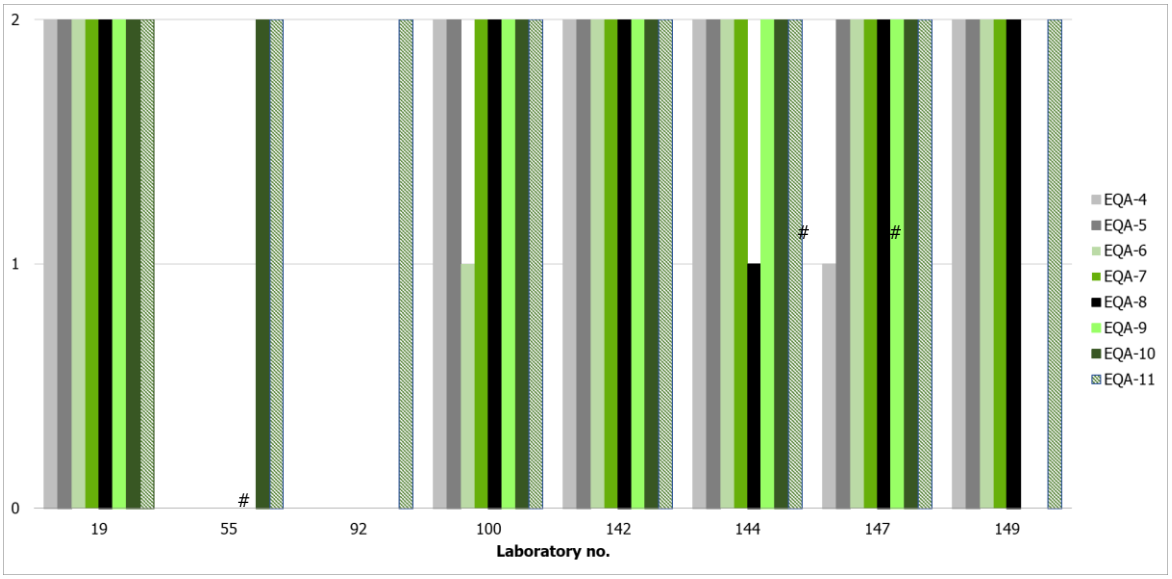
Figure 2. Average percentage scores of the 10 *S.* Typhimurium test isolates



Bars represent the percentage of MLVA profiles correctly assigned by the participants.
#: repeat isolates (STm2 and STm3) in EQA-4 to 11.

To follow the development of individual laboratory performance, two repeat isolates with different allelic profiles were included in EQA-4 through to 11: isolate STm2 (3-13-NA-NA-211) and STm3 (3-12-9-NA-211). Figure 3 shows the individual performance by the laboratories of these two repeat isolates during the eight EQAs (only showed for the laboratories participating in EQA-11). All participants were able to identify the correct MLVA profile for both repeat isolates and the performance on these isolates was increased overtime.

Figure 3. Correct MLVA typing of two repeat *S.* Typhimurium isolates from EQA-4 to 11 (for laboratories participating in EQA-11)

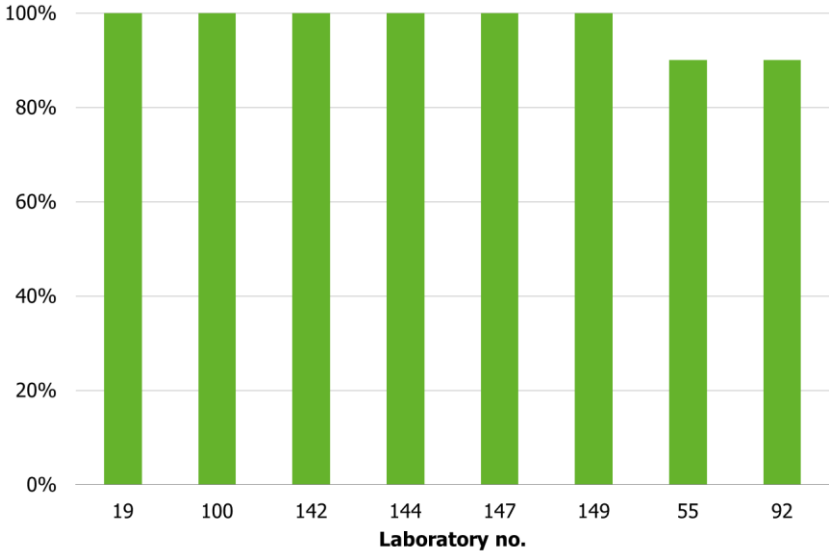


Arbitrary numbers represent the participating laboratories. Bars represent the number of correctly assigned allelic profiles for the two repeat isolates (STm2 and STm3).
#: laboratory did not correctly identify any of the two repeat isolates.

3.2.2 MLVA for *S.* Enteritidis

Eight out of the 20 participants (40%) in EQA-11 performed the MLVA typing of *S.* Enteritidis and six (75%) of these reported the correct allelic profiles for all 10 test isolates (Figure 4).

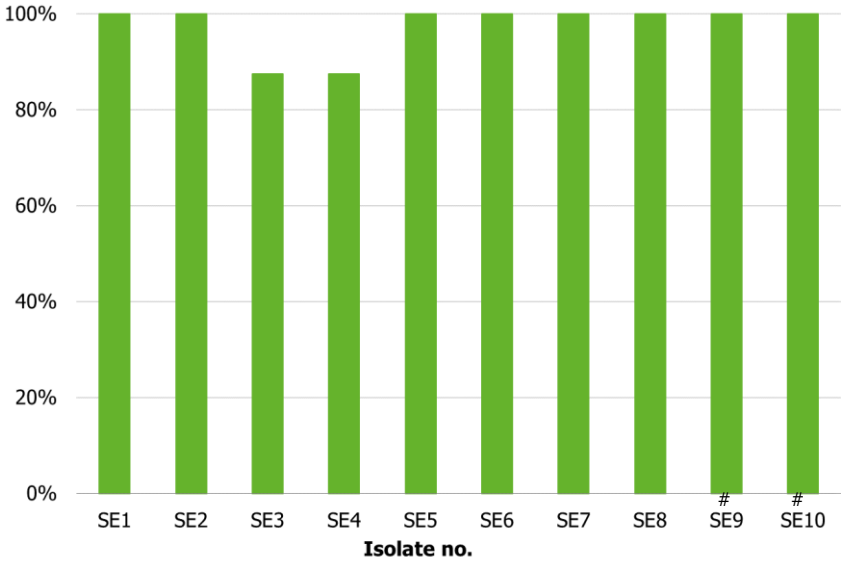
Figure 4. Participant scores for MLVA typing of the 10 *S.* Enteritidis test isolates



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

Laboratory 55 and 92 had one error respectively in different isolates (Annex 4). Laboratory 55 reported incorrect allele number 2 instead of 4 in SENTR4 for isolate SE4. Laboratory 92 reported an incorrect allele number 9 instead of 10 in fragment SENTR5 for isolate SE3.

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



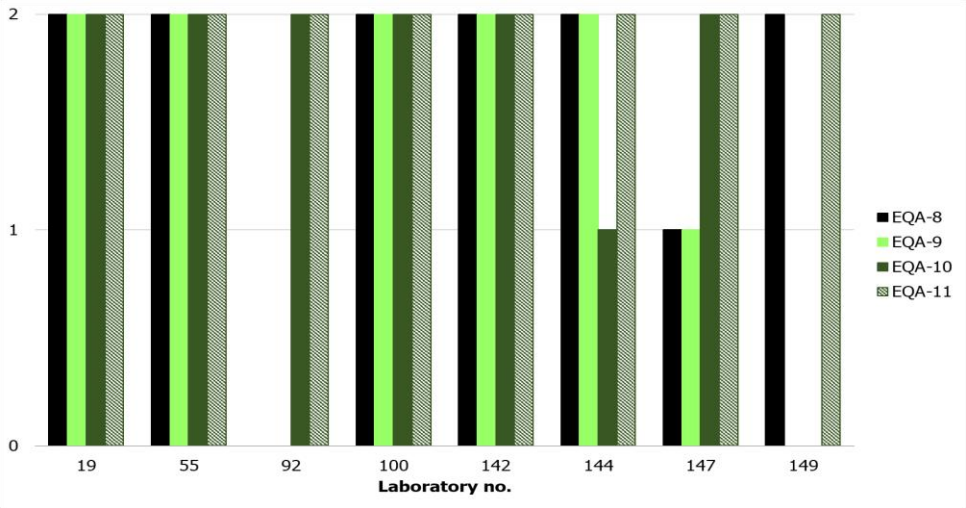
Bars represent the percentage of MLVA profiles correctly assigned by the participants.
#: repeat isolates (SE9 and SE10) in EQA-8 to 11.

The results for each test isolate are summarised in Figure 5. The correct MLVA profile was reported for eight of the 10 *S.* Enteritidis test isolates by all participants. No common isolate characteristics caused the problems (Annex 4) as the two incorrect MLVA profiles concerned two different isolates SE3 and SE4.

To follow the development of individual laboratory performance, two isolates with different allelic profiles were included in EQA-8 to EQA-11: 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 four EQAs (EQA-8 to EQA-11) for laboratories participating in this EQA-11.

The MLVA results on the repeat isolates show stability and high performance among the participants. All participants were able assign the correct MLVA profile to the repeat isolates. Thereby, all laboratories performed at the same or a higher level compared to last time they participated.

Figure 6. Correct MLVA typing of two repeat *S.* Enteritidis isolates from EQA-8 to 11 (for laboratories participating in EQA-11)



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 were asked to identify a cluster of closely related isolates among 10 test isolates by using either PFGE, MLVA and/or WGS-derived data. The cluster of five test isolates was pre-categorised by the EQA provider.

All ten test isolates had the same MLVA profile, 3-10-5-4-1, and it was therefore not possible to identify the cluster of five closely related isolates by using this method. The characteristics of the test isolates and reported results are listed in Annexes 5-6; 8-13.

3.3.1 PFGE-derived data

Six (6/20, 30%) 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 if a less discriminatory method, e.g. PFGE, was used. None of the participants identified the correct cluster of five closely related isolates, however five participants included the five WGS cluster isolates as a part of the correct cluster. Laboratory 92 did not include all the five cluster isolates, as REF6 was excluded. When using PFGE all the six participants included REF9 as part of cluster and three laboratories included REF5 as well. Table 3 shows the overview of the cluster analysis of the isolates each participant included or excluded in their cluster.

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

		Laboratory					
Isolate number	ST	92	96	106	127	130	142
REF1	11	No	No	No	No	No	No
REF2	11	No	No	No	No	No	No
REF3‡#	11	Yes	Yes	Yes	Yes	Yes	Yes
REF4‡	11	Yes	Yes	Yes	Yes	Yes	Yes
REF5	11	No	Yes	No	No	Yes	Yes
REF6‡	11	No	Yes	Yes	Yes	Yes	Yes
REF7	11	No	No	No	No	No	No
REF8‡	11	Yes	Yes	Yes	Yes	Yes	Yes
REF9	11	Yes	Yes	Yes	Yes	Yes	Yes
REF10‡#	11	Yes	Yes	Yes	Yes	Yes	Yes
Correct cluster identified ‡		**No**	**No**	**No**	**No**	**No**	**No**
Included the five WGS cluster isolates		**No**	**Yes**	**Yes**	**Yes**	**Yes**	**Yes**

‡: closely related isolates based on WGS (in grey)
#: technical duplicate isolates (in bold)
Annex 6

3.3.2 MLVA-derived data

Four participants (20%) performed cluster analysis using MLVA-derived data and all selected the *S.* Enteritidis scheme and reported the loci in the correct order: SENTR7, SENTR5, SENTR6, SENTR4 and SE-3.

As the cluster of closely related isolates was defined by a pre-categorisation based on WGS by the EQA provider and all the 10 cluster test isolates had the same MLVA profile: 3-10-5-4-1, the correct cluster delineation in this EQA was impossible to obtain by the use of MLVA. The method was not discriminatory enough to give separation between the test isolates in this EQA. Therefore, none of the four laboratories was able to identify the correct cluster of closely related isolates due to the profile. Table 4 shows the overview of the isolates each participant included 'Yes' and excluded 'No' in their cluster analysis.

Laboratories 142 and 147 reported all 10 test isolates as part of the cluster and reported the correct MLVA profile. Laboratory 55 did not include REF6 and REF10 in the cluster as they had assigned incorrectly allelic profiles (3-NA-5-4-1) reporting absent fragment in SENTR5 for REF6 and REF10. In addition, laboratory 138 did conclude that all the isolates had the same profile; however, they reported an incorrect allelic profile (2-11-7-7-2) for all 10 isolates. All data are available in Annexes 8 and 9.

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

			Laboratory ID			
Isolate number	ST	MLVA-profile	55*	138¤	142	147
REF1	11	3-10-5-4-1	Yes	Yes¤	Yes	Yes
REF2	11	3-10-5-4-1	Yes	Yes¤	Yes	Yes
REF3‡#	11	3-10-5-4-1	Yes	Yes¤	Yes	Yes
REF4‡	11	3-10-5-4-1	Yes	Yes¤	Yes	Yes
REF5	11	3-10-5-4-1	Yes	Yes¤	Yes	Yes
REF6‡	11	3-10-5-4-1	No	Yes¤	Yes	Yes
REF7	11	3-10-5-4-1	Yes	Yes¤	Yes	Yes
REF8‡	11	3-10-5-4-1	Yes	Yes¤	Yes	Yes
REF9	11	3-10-5-4-1	Yes	Yes¤	Yes	Yes
REF10‡#	11	3-10-5-4-1	No	Yes¤	Yes	Yes
Cluster-identified		3-10-5-4-1	**No**	**No**	**No**	**No**

‡: closely related isolates derived from WGS (in grey)
#: technical duplicate isolates (in bold)
**: Incorrect fragment for REF6 and REF10 (SENTR5, NA instead of 10)*
¤: Incorrect allelic profiles (2-11-7-7-2) for all cluster isolates
Annex 8 and 9.

3.3.3 WGS-derived data

Reported results from participants

Fourteen participants (70%) 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, six MiSeq, six NextSeq and one Ion Torrent. All reported using commercial kits for library preparation. Of the 14 participants, 12 (86%) used Illumina's Nextera kit (Annex 7).

Performance was high in cluster analysis with WGS-derived data, with 13 (93%) participants correctly identifying the cluster of closely related isolates defined by a pre-categorisation by the EQA provider among the 10 test isolates.

Twelve laboratories correctly reported ST of all 10 isolates. Laboratory 147 did not report the 7-gene Multi Locus Sequence Types for any of the 10 isolates and laboratory 148 used the submission field for a number of six digits, without any explanations, but probably using Enterobase [5] level cluster codes. Table 5 shows the overview of the cluster analysis of the isolates each participant included or excluded in their cluster.

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

		Laboratory ID													
Isolate number	**ST**	**19**	**36**	**49**	**100**	**106**	**108**	**127**	**129**	**134**	**135**	**142**	**147**	**148**	**149**
REF1	11	No	No	No	No	No	No	No	No	No	No	No	No	No	No
REF2	11	No	No	No	No	No	No	No	No	No	No	No	No	No	No
REF3‡#	11	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
REF4‡	11	Yes	Yes	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
REF5	11	No	No	No	No	No	No	No	No	No	No	No	No	No	No
REF6‡	11	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
REF7	11	No	No	No	No	No	No	No	No	No	No	No	No	No	No
REF8‡	11	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
REF9	11	No	No	No	No	No	No	No	No	No	No	No	No	No	No
REF10‡#	11	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Main analysis		Allele	Allele	Allele	Allele	Allele	SNP	Allele	Allele	Allele	Allele	Allele	Allele	Allele	Allele
Additional analysis 1		SNP		Allele				SNP							
Additional analysis 2.								SNP							
Cluster-identified		**Yes**	**Yes**	**No**	**Yes**	**Yes**	**Yes**	**Yes**	**Yes**	**Yes**	**Yes**	**Yes**	**Yes**	**Yes**	**Yes**

‡: closely related isolates (in grey)
#: technical duplicate isolates (in bold)
ST: sequence type
Allele: allele-based analysis
SNP: single-nucleotide polymorphism analysis
Annex 10-11.

Laboratories were instructed to report the data used for cluster analysis 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, 49 and 127 reported additional analyses.

Of the three participants using SNP analysis, one (laboratory 108) used SNP as the main analysis for cluster detection, one (laboratory 19) reported SNP as an additional analysis and laboratory 127 reported SNP-based analyses both as an additional analysis and as a third analysis. Laboratory 19 and 127 used a reference-based approach with EQA *S.* Enteritidis isolates as a reference. Laboratory 108 used SNP as the main analysis and used an in-house pipeline. Laboratory 127 used Enterobase and strict SNP filtering pipeline in BioNumerics for the two additional SNP analyses. The participants used different read mapper and variant callers.

Tables 6 and 7 show the overview of the submitted data. For laboratory reported SNP distance/allelic differences by isolate, see Annex 11.

Table 6. Reported results of SNP-based cluster analysis

	SNP-based analysis						
Laboratory	SNP pipeline	Approach	Reference	Read mapper	Variant caller	Distance within cluster	Distance outside cluster
Provider	NASP [8]	Reference-based	REF3	BWA	GATK	0-1	16-66
19*	NASP	Reference based	ST11 and ID DK-SSI-5904	BWA	GATK	0-1	17-68
108	In-house	Assembly based	-	CLC Assembly Cell v.5.2	CLC Assembly Cell v.5.2	0-1	18-71
127*	EnteroBase (SNP project)	Reference based	ST11, ID = 5945	Enterobase SNP pipeline	Enterobase SNP pipeline	0-1	50-79
127#	Bionumerics (Strict SNP filtering Closed SNP set)	Reference based	5945	Bionumerics	Bionumerics	0-1	17-116

**: additional analysis 1*
#: additional analysis 2
Detailed data, see Annex 11

Thirteen participants used an allele-based analysis; all selected the method as the main analysis for cluster detection. Nine of 13 (69%) used only an assembly-based allele calling method and four (31%) used both assembly- and mapping-based allele calling methods.

Eight of the main analysis used SPAdes as the assembler and three used Velvet. Two laboratories used SKESA. Twelve of the 13 (92%) main analysis used a cgMLST scheme (Enterobase) for the allele-based analysis and one laboratory (129) reported the use of a different scheme (Ad hoc scheme, 3696 core loci). In the additional analysis, Laboratory 49 used wgMLST scheme of 15.874 loci.

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	Difference outside cluster
Provider	BioNumerics	Assembly- and mapping-based	SPAdes	Applied Maths (cgMLST/Enterobase)	3002	0-1	10-32
19	BioNumerics	Assembly based and mapping based	SPAdes	Applied Math (cgMLST/Enterobase)	3002	0-1	10-32
36	SeqPhere	Only assembly based	SKESA 2.3.0	Enterobase (cgMLST)	3002	0-1	10-32
49	BioNumerics	Assembly based and mapping based	SPAdes	Applied Math (cgMLST/Enterobase)	3002	0	2-33
49*	BioNumerics	Assembly based and mapping based	SPAdes	Applied Math (wgMLST)	15.847	0-2	14-51
100	SeqPhere	Only assembly based	Velvet	Enterobase (cgMLST)	3002	0-1	10-42
106	SeqPhere	Only assembly based	SPAdes	Enterobase (cgMLST)	3002§	0-1	9-39
127	Enterobase	Assembly based and mapping based	SPAdes	Enterobase (cgMLST)	3002	0-1	12-51
129	SeqPhere	Only assembly based	Velvet	§Ad hoc scheme ("wgMLST")	3696	0-1	9-34
134	SeqPhere	Assembly based and mapping based	SKESA 2.3.0	Enterobase (cgMLST)	3002	0-1	10-32
135	SeqPhere	Only assembly based	SPAdes	Enterobase (cgMLST)	3002	0-2	11-32
142	Galaxy (In house)	Only assembly based	SPAdes	Enterobase (cgMLST)	3002	0-1	11-34
147	SeqPhere	Only assembly based	SPAdes 3.11.1	Enterobase (cgMLST)	3003	0-1	11-33
148	Enterobase	Only assembly based	SPAdes	Enterobase (cgMLST)	3003	0-3	13-52
149	SeqPhere	Only assembly based	Velvet	Enterobase (cgMLST)	3002	0-1	10-32

**:additional analysis*
§: modified from submitted information
For detailed data, see Annex 11.

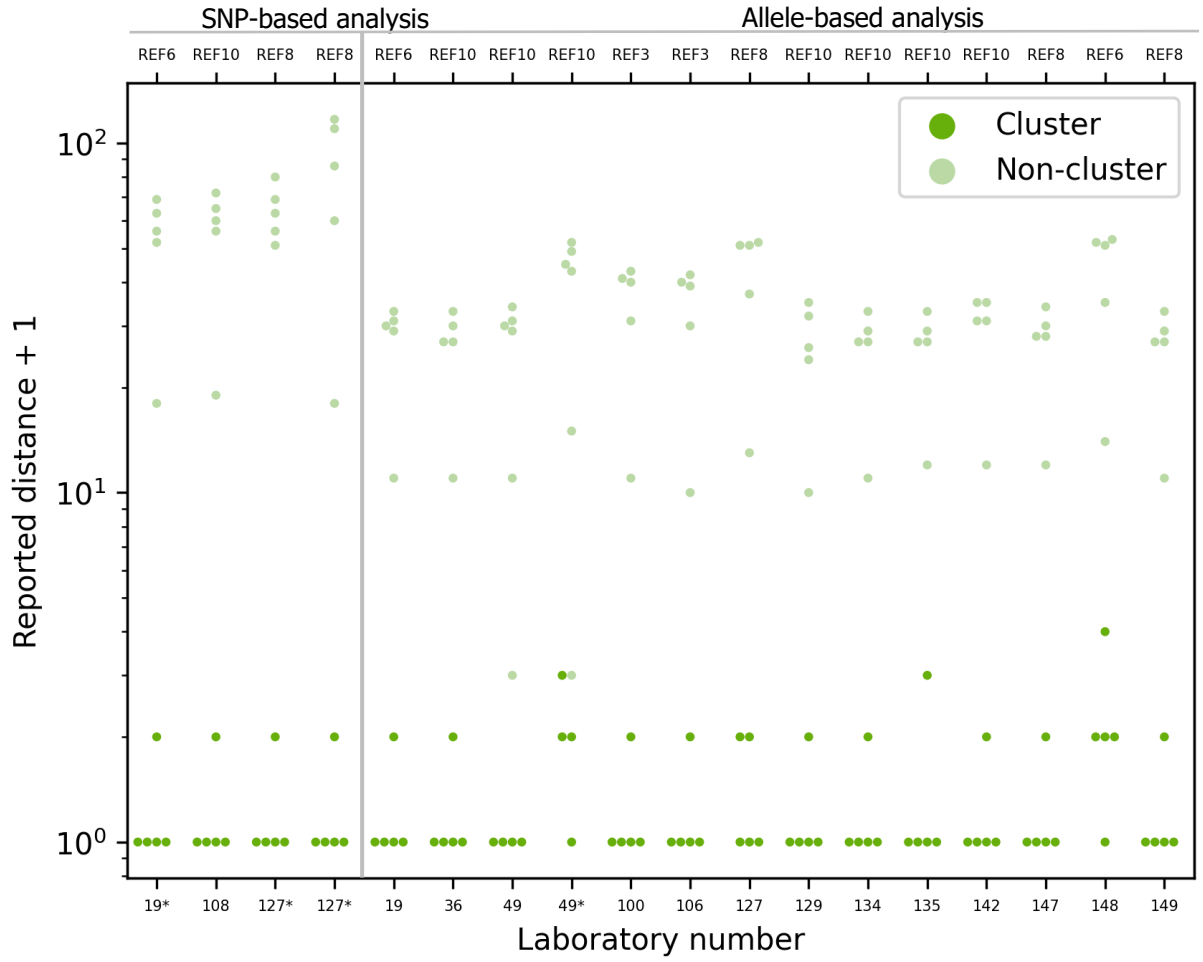
Laboratory 108 performed the only participant SNP analysis as the main analysis and identified the correct cluster of closely related isolates. Two other laboratories (19 and 127) also performed SNP analysis as additional analysis and identified the correct cluster of closely related isolates by cgMLST (main analysis). The reported SNP distances within the cluster were 0-1 for all the analysis. For the test isolates outside the cluster, the laboratories reported a SNP distance range to the selected cluster isolates at 17-116.

Twelve out of 13 laboratories (93%) used an allele-based analysis as the main method and could identify the correct cluster of the five closely related ST11 isolates (Figure 8). Ten of the laboratories reported 0-1 allele differences (AD) in the correct cluster using cgMLST, one reported 0-2 and one reported 0-3 (Table 7).

Laboratory 49 did not identify the correct cluster and had no AD inside the reported cluster. The laboratory excluded REF4 from the cluster, however only 2 AD to REF4 was reported based on the cgMLST analysis. In addition, laboratory 49 performed wgMLST, which resulted in 0-2 AD inside the reported cluster, as both REF4 and REF8 were 2 AD from the other reported cluster isolates (Figure 7, Annex 11), and based on the wgMLST analysis, laboratory 49 reported the same difference to REF4 as by cgMLST. Both laboratories 19 and 49, and the EQA provider, used the same approach (BioNumerics) and the same allele calling method, assembler, and cgMLST-scheme.

All the test isolates outside the cluster were also ST11. As the only participant, laboratory 49 used a very strict cut-off and reported differences outside the cluster down to 2 alleles by used of cgMLST. All other laboratories reported AD at 9-52 to the selected cluster isolate for this group of isolates (difference outside cluster) by any schemes used (cgMLST and wgMLST).

Figure 7. Reported SNP distances or allele differences for each test isolate to selected cluster representative isolate



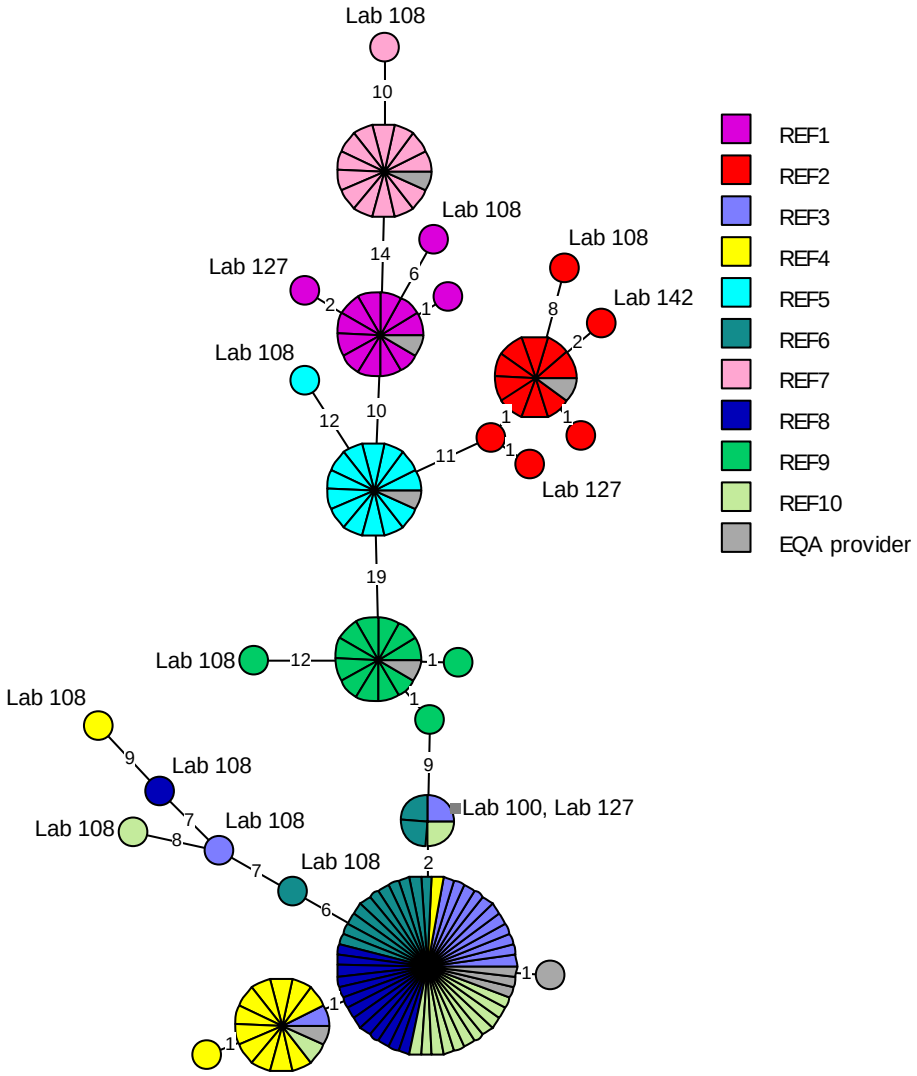
**: additional analysis*
SNP: single-nucleotide polymorphism analysis
Selected cluster representative marked as REF
Dark green: reported cluster of closely related isolates
Light green: reported not part of cluster.

Analysis of raw reads uploaded by participants

In addition to the reported cluster analysis, 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 8) and based on submitted raw reads from 14 laboratories, shows clear clustering of the results for each test isolate. Some data from laboratory 100, 108, 127 and 142 are separated withtwo2 or more AD from the other results.

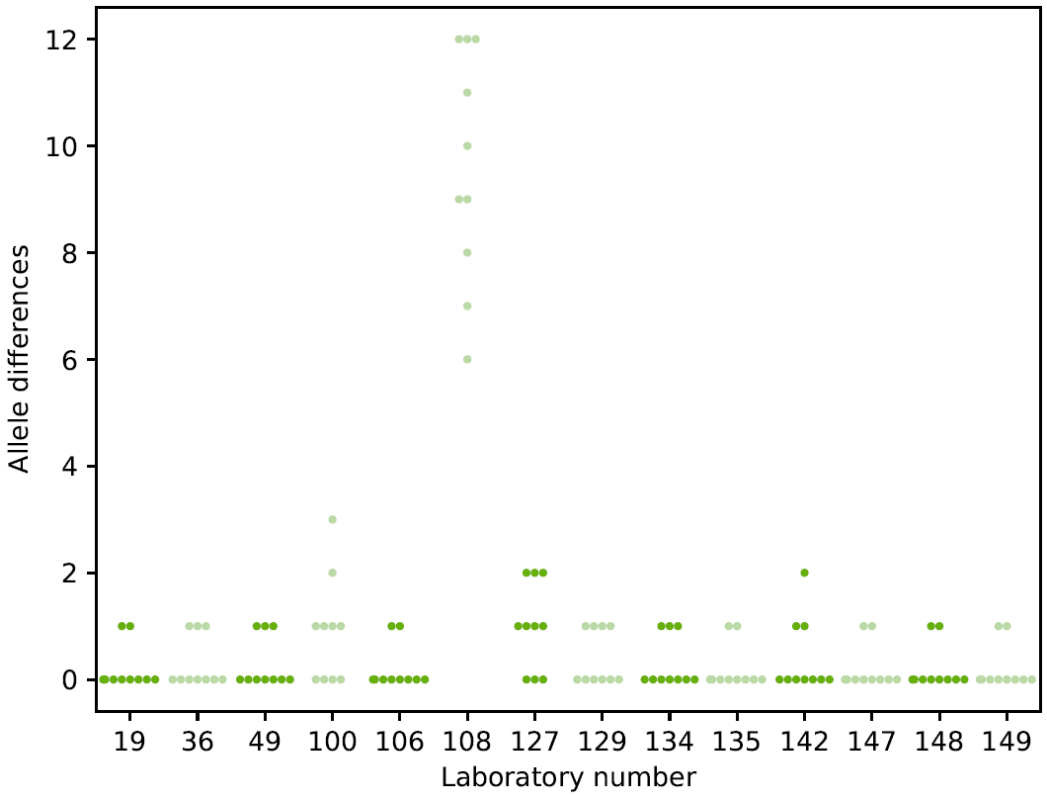
Figure 8. Minimum spanning tree of core genome multilocus sequence typing, participant FASTQ files



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 9 shows the AD between each submitted sequence and the corresponding reference. As seen in Figure 8 and 9, a few laboratories have some isolates with large differences from the reference and results of the other participants. Laboratory 108 is the one with the highest differences.

Figure 9. 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 124 of 140 results (89%), 0-1 differences were identified (Figure 9). For six results, a difference of 2-3 alleles from the REF isolate was calculated. For ten results (7%), a difference of 6-12 alleles was seen, all reported by laboratory 108. The difference is the result of Ion Torrent data when it is analysed in BioNumerics.

Separately, the laboratories responded to QC parameters used to evaluate their data. As seen in Table 8, both coverage and contamination check were the most widely used QC parameter with 100% and 86% of the laboratories using these parameters respectively. Different thresholds of coverage ranging between 10-50X were used. The genome size and the number of good cgMLST loci was used by 79% of laboratories with a threshold ranging between 4-6 Mb and 90-99% reported by most. Some laboratories did not report a specific threshold. The Q score was used by 57%. Additional QC parameters are listed in Annex 12 as reported by some of the participants.

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 other genus	Min. 50x	N50 value, number of contigs and number of unidentified bases, but no actual threshold is employed	4.5-5.3 Mb	Min. 95% core percent and max. 15 loci with multiple consensus - No actual threshold employed on regular basis for either
36	KRAKEN, MashScreen in SeqSphere	30-fold	default statistics in SeqSphere	assembly base count should resemble expected geonome size	default settings in SeqSphere
49	No	>=30 pass >=20 warn	BioNumerics >= 30= pass	4.5- 5.5 Mb. Can vary by ST/Serotpe/Subspecies	Look at multiple alleles
100	KmerFinder 3.2 CGE	40x	FastQC	app. 5 Mb	No
106	Kmer	0,95	N50, contig count, reads count, average count	No	No
108	Assembled genomes blast against reference	>20x	No	Total assembly length evaluated against predefined expected genome size	De novo assembly, remapping, length of coding sequence, sequence coverage over allele
127	No	>=40	Number of contigs <600	4-5.8 Mbp	Enterobase pipeline
129	SeqSero2 and occasionally Kraken	>29	No	No	No
134	Mach in SeqSphere	50x <50x + good targets >95%	No	length of contigs assembled < reference genome + 10%	cgMLST alleles found and called >95%
135	CheckM, 96% completeness of genus *Salmonella*	>=10	number of contigs <300	4.54-5.21 Mb	>90% assigned alleles
142	Kraken	30x	No	>10% deviation	<95% calling
147	JSpecies	Min. 30x	No	~ 5 Mb	percentage of good targets ~ 98% (min.)
148	Kraken (Enterobase) : > 70% contigs assigned	>45x	Number of contigs < 250 (alert in house) and < 500 (minimum acceptance for Enterobase)	4-5.1 Mb (alert in house) and between 4-5.8 Mbp (acceptance for Enterobase)	BWA back-mapping the reads, and Samtools (with BCFtools) for variant calling
149	Kraken	average coverage (seqsphere) + coverage per contigs in Kraken	No	No	% good targets in Seqsphere.
% of laboratories using QC parameter	**86%**	**100%**	**57%**	**79%**	**79%**

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 13. Overall, the coverage of the raw data was sufficiently high when evaluated by the EQA provider´s QC pipeline, although a few isolates of different participants had an average coverage below the generally accepted threshold of 50, some as low as 21. One laboratory (149) had a contamination with *Pseudomonas tolaasii* in seven of the ten isolates (Annex 13).

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
Detected species	{Se}	*Se*	*Se*	*Se*	*Se*	*Se*	*Se*	*Se*	*Se*	*Se*	*Se*	*Se*	*Se*	*Se*	*Se, Pt*
% Species 1		94.4-95.5	96.7-97.4	96.0-98.4	97.5-98.5	97.5-98.2	96.9-98.3	92.8-96.8	95.5-98.0	92.2-98.3	97.0-98.2	96.4-98.1	93.9-97.7	95.8-96.3	75.3-88.8
% Species 2	{<5%}	0-0.3	0-0.5	0-0.4	0-0.3	0-0.4	0.4-1.6	0-0.1	0-0.3	0-3.3	0.4-0.8	0.1-0.4	0-0.4	0-0.2	3.9-10.7
Unclassified reads (%)		4.4-5.5	2.4-2.7	1.4-3.9	1.4-2.4	1.7-2.4	1.2-1.4	3.1-7.0	1.9-4.4	1.7-3.9	1.2-2.0	1.6-3.2	2.1-5.9	3.6-4.0	6.8-12.4
Length at >25 x min. coverage (Mbp)	{>45 ∧ <53}	4.5-4.8	4.7-4.8	3.9-4.8	4.7-4.8	4.7-4.8	4.7-4.8	4.7-4.8	4.6-4.8	0.0-4.8	4.7-4.8	4.7-4.8	4.7-4.8	4.6-4.8	4.7-4.8
Length [1-25] x min. coverage (kbp)	{<250}	2.6-127.2	0-11.5	0-859.9	0	0	0	0-30	0.5-50.1	0-4711.3	0	0-4.3	0-3.9	0-10.8	0-30.0
No. of contigs at 25 x min. coverage	{>0}	85-303	29-62	18-30	22-28	22-28	372-1113#	27-109	57-179	7-50	23-31	16-24	22-70	113-201	31-38
No. of contigs [1-25] x min. coverage	{<1000}	2-112	0-5	0-3	0	0	0#	0-16	1-43	0-62	0	0-3	0-6	0-15	0-9
Average coverage	{>50}	34-96	102-163	31-97	77-127	305-405	106-120	51-160	46-123	21-98	122-291	36-103	44-164	193-257	59-114
No. of reads (x 1000)		1129-3226	3292-5258	535-1932	1570-2616	9710-12943	2000-2000	1641-5413	1480-4105	696-3302	3922-9595	711-2168	856-3393	6097-8216	2445-4278
Average read length		143-145	147-149	219-287	227-238	149-150	258-287	141-148	149-149	149-151	151-151	226-246	233-277	149-150	151-151
Average insert size		262-297	326-384	268-445	288-340	377-496	NA#	350-477	360-471	148-321	346-397	340-558	266-438	335-375	199-282
N50 (kbp)		25-133	167-406	312-490	406-493	490-490	7-20#	72-278	47-165	169-490	338-695	490-694	125-694	46-97	260-376

**: indicative QC range*
Se: Salmonella enterica
Pt: Pseudomonas tolaasii
#: QC values unreliable due to assembly issues for Ion Torrent data
Annex 13.

Assessment of the provided genomes

The six provided genomes were asked to be assessed one-by-one and compared with the already produced data in the cluster analysis. The participants had to determine whether the genomes were part of the defined cluster.

The participants were instructed to describe their observations and considerations leading to the decision. The EQA-provider had manipulated some of the raw reads. Four of the six genomes represented raw reads of a cluster isolate with either high-quality raw reads, reduced coverage, contaminated with *Citrobacter* or contaminated with a different *Salmonella* ST34. The additional two genomes were non-cluster isolates with high-quality raw reads and a Fasta file respectively (table 10). Raw data can be seen in Annex 14.

For genome 1 (a cluster isolate, good quality but contaminated with approx. 10% *Citrobacter*), 71% (10/14) correctly described contamination present in genome 1. Six of the ten participants described the added species as being *Citrobacter*. Eleven of the laboratories (11/14) performed an analysis or partially conducted an analysis despite many recommended resequencing. Based on their analysis, all eleven laboratories described as observations that the genome belonged or could belong to the cluster, despite some of them selecting 'not part of the cluster' during submission. Three did not perform the analysis because of the contamination.

For genome 2 (a cluster isolate, good quality but contaminated with approx. 20% *Salmonella* ST34), 100% (all 14 laboratories) correctly described contamination present in genome 2. Eight out of the 14 participants described contamination as a different *Salmonella enterica*. One laboratory (108) reported the genome as part of the cluster based on a SNP analysis. Four laboratories reported genome 2 as a non-cluster isolate and nine did not perform the analysis because of the contamination.

For genome 3 (a cluster isolate, low coverage), 100% (all 14 laboratories) correctly observed poor quality in genome 3, ten used either the low average coverage, N50, cgMLST targets or high number of contigs to disregard the genome. Only four laboratories proceeded with the analysis, and three concluded that the isolate was not a part of the cluster and one laboratory correctly suggested that genome 3 could be part of the cluster of closely related isolates.

For genome 4 (a non-cluster isolate of good quality), 100% 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.

For genome 5 (a non-cluster isolate of good quality, assembly FASTA), 93% (13/14) accepted the quality of the genome as the Fasta-file, one laboratory described difficulties with downloading the sequence. The remaining 13 laboratories correctly identified genome 5 as a non-cluster isolate. Only two laboratories performed SNP-analyses and reported the distances of 12 SNP/57 SNP to the cluster. Most of the additional laboratories, which used allele-based analyses, found 25-34 AD. Only one laboratory reported >50 AD to the cluster. Four laboratories observed that genome 5 was identically with the REF1.

For genome 6 (a cluster isolate, good quality), 100% all 14 laboratories accepted the quality of the genome and correctly described the genome as an isolate part of the cluster of closely related isolates. All laboratories reported 0-1 AD/SNP to the representative genomes in their 'clusters'.

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 (REF3/REF10) mixed with a *Citrobacter* (approx. 10%).	Contamination was observed	10	4	0
		Suggested to be a cluster isolate	11	0	3
2	A cluster isolate (REF3/REF10) mixed with a *Salmonella* ST34 (approx. 20%), same species contamination.	Contamination was observed	14	0	0
		Suggested to be a cluster isolate	1	4	9
3	A cluster isolate (REF3/REF10) with altered coverage (reduced to 10x).	Poor quality was observed	14	0	0
		Suggested to be a cluster isolate	1	3	10
4	A non-cluster isolate (REF9), good quality of reads. 10 allelic difference to the cluster (REF3).	Quality accepted	14	0	0
		Suggested to be a cluster isolate	0	14	0
5	A non-cluster isolate (REF1), good quality of reads assembled with SKESA to a FASTA file. 29 allelic difference to the REF3 in the cluster.	Quality accepted (Fasta)	13	0	1
		Suggested to be a cluster isolate	0	13	1
6	A cluster isolate (REF3/REF10) good quality of reads.	Quality accepted	14	0	0
		Suggested to be a cluster isolate	14	0	0

Annex 14.

4. Discussion

Overall, the total number of participants has decreased over time. From 26 participants in EQA-7 to 19 in EQA-10 and 20 in EQA-11. Among the 20 participants, two laboratories (130 and 138), which did not participate in EQA-10, participated again in EQA-11. Four laboratories (128, 132, 140 and 145) participated in both EQA-8 and EQA-9 or only EQA-9, but did not participate in EQA-10 or in EQA-11 (the last one accepted the invitation, but did not submit any results in EQA-11). All these four laboratories used PFGE as a 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

Eight laboratories (40%) participated in the MLVA part, where all eight laboratories performed the analysis for both *S.* Typhimurium and *S.* Enteritidis. In EQA-10, laboratory 92 only participated in the MLVA analysis for *S.* Enteritidis, however in EQA-11 the laboratory successfully participated in both MLVAs with 90-100 % correct results. MLVA for *S.* Enteritidis was included in the EQA for the fourth time and the number of participants has been decreasing from twelve in EQA-8 to eight in EQA-11. In MLVA for *S.* Typhimurium, the number of participants was also lower than in previous years, decreasing from 15 participants in EQA-4 to ten participants in EQA-9, to eight in both EQA-10 and EQA-11. This could reflect a trend, where more laboratories are switching to WGS-based surveillance and outbreak detection using WGS instead of MLVA. The MLVA method may be useful for screening at the national level because of its low-cost, easy analysis and interpretation compared with WGS, but it has little value at the EU level surveillance

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 95%, which was around the same level as previous years. From EQA-4 to EQA-10, the overall performance was 86-97%.

The MLVA results of the two repeat *S.* Typhimurium isolates from EQA-4 through EQA-11 showed good performance by the participants. When observing the laboratories participating in EQA-11, almost all participants (88%; 7/8) performed at the same high level as previous time they participated and 1/8 (12,5%) participated for the first time and delivered results at a high level.

The few mistakes 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.

For MLVA of *S.* Enteritidis, six laboratories (75%) obtained a total score of 100% and the overall performance was 98%, which was higher compared to EQA-10, but at the same level as EQA-9 (98%). Mistakes in the MLVA for *S.* Enteritidis were reported by only two laboratories, one error each in two different isolates.

The MLVA results of the two repeat *S.* Enteritidis isolates from EQA-8 through EQA-11 of the EQA-11 participants showed good performance by the participants. When observing the laboratories participating in EQA-11 almost all participants (88%; 7/8) performed at the same high level, as the previous time and 1/8 (12,5%) improved their performance compared with the last time they participated

The overall better performance for *S.* Typhimurium and *S.* Entertidis MLVA this year was caused by a much better performance from laboratory 55 compared with EQA-10. However, while implementing WGS as a routine method, it will probably in general be a challenge for the laboratories to maintain quality and skills of the MLVA method in the future, when the routine use of the method will decrease.

4.2 Molecular typing-based cluster analysis

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

The expected cluster of closely-related *S.* Entertidis ST11 isolates contained five isolates in the pre-defined cluster based on the WGS data. This definition was based on the clonal nature of *S.* Enteritidis and the knowledge about outbreak investigation of this serotype [11-15].

Nineteen of twenty laboratories participated in the molecular typing-based cluster analysis using either PFGE and/or MLVA and/or WGS. In the first EQA (EQA-8), with the molecular typing-based cluster analysis (EQA-8), only thirteen participated, but the overall participation in the cluster part has been stable over the last three years (21 participants in EQA-9 and 19 participants in EQA-10 and EQA-11). Compared with EQA-10, two new laboratories participated in the cluster part, one using PFGE and one using MLVA. However, two other laboratories who performed cluster analysis in EQA-10 did not participate in EQA-11 (Annex 2).

4.2.1 PFGE-derived data

Of the 19 laboratories, six (32%) performed cluster analysis using PFGE-derived data. None of the laboratories identified the correct cluster. Most of the laboratories included the five cluster isolates but all six laboratories also included additional isolate/isolates in the identified cluster. This is to be expected, as *S.* Enteritidis is normally a very genetically homogenous serovar and the discriminatory power of PFGE can be insufficient for cluster detection [16-17].

The total number of laboratories performing PFGE-based cluster analysis has varied in the different EQAs (four in EQA-8, thirteen in EQA-9, seven in EQA-10 and six in EQA-11) and the number of laboratories only performing PFGE follows the same decreasing trend (two in EQA-8, nine in EQA-9, two in EQA-10 and three in EQA-11). The increased number of participants performing PFGE-based cluster analysis from EQA-8 to EQA-9 was probably caused by the adjustment, where the PFGE part with gel quality and analysis assessment was excluded. Few countries still use PFGE for cluster analysis, and in the future, the method will probably no longer be an important or used typing method for cluster detection and outbreak investigation in EU. The method can still be useful for national purposes.

4.2.2 MLVA-derived data

In total, four laboratories performed cluster analysis using MLVA-derived data. The total number of laboratories performing MLVA-based cluster analysis has been low in all four EQAs (four in EQA-8, two in EQA-9, five in EQA-10 and four in EQA-11). However, for the first time, two laboratories were only using MLVA for the cluster analysis; one also used WGS and one used both PFGE and WGS together with MLVA.

Determination of the correct cluster of the five isolates was not possible by the use of MLVA, due to the nature of the cluster, as all the 10 test isolates had the same MLVA profile. Two of the laboratories reported the correct MLVA profile for all 10 isolates, one laboratory only reported the correct profile for 8/10 isolates and the last laboratory reported an incorrect MLVA profile for all 10 isolates.

Overall, the few laboratories performing cluster analysis by the use of MLVA and the decreasing trend in laboratory participation in the MLVA part as discussed in section 4.1, indicate that the method will continue to be a less important molecular typing method for detecting and investigating *Salmonella* outbreaks.

4.2.3 WGS-derived data

Fourteen of 19 laboratories (74%) performed cluster analysis using WGS-derived data. This was a higher participation compared with EQA-9, where 12/21 (57%) performed cluster analysis using WGS-derived data, but a bit lower than EQA-10 with 15/19 (79%). Unfortunately, no new laboratories began to use WGS this year for the cluster analysis and it showed that the method is still not implemented in some laboratories, which can be a challenge for improved inter-laboratory comparability.

Performance was again high, as 13 (93%) correctly identified the cluster of closely related isolates. Only one laboratory (49) was not able to identify the correct cluster as they used a too strict cut-off. However, laboratory 49 identified the correct cluster in the three previous EQA´s. In EQA-10 laboratory 108 did not identify the correct cluster and in EQA-9 two other laboratories (120 and 142) did not identify the correct cluster, but they all succeeded in EQA-11.

Most laboratories (13/14) reported the use of an Illumina platform and all reported using commercial kits for library preparation. In EQA-9, EQA-10 and EQA-11 only one laboratory reported the use of external assistance for sequencing, only in EQA-10 and EQA-11 it was the same laboratory. Using an external assistance for sequencing has no influence in cluster detection in any of the EQAs and no quality issues have been detected.

Thirteen laboratories (93%) reported using an allele-based method as the main analysis and for the first time only one laboratory (7%) reported using SNP analysis. Compared with EQA-9 (83%) and EQA-10 (80%) this is a small percentage increase in the use of allele-based as the main analysis. Laboratory 100 changed from SNP analysis to allele-based analysis, but all other laboratories used the same main analysis as in EQA-10.

During the EQA-8 to EQA-11 the use of the cgMLST scheme (Enterobase) has become more and more dominant, and in EQA-11, twelve laboratories used the scheme for the main analysis. The preference of using wgMLST has varied and only laboratory 129 continued to use the ad hoc 'wgMLST' scheme as their only analysis.

The reported allele differences (AD) were very comparable. Inside the cluster, the reported differences were 0-3 alleles for all the main analyses. Laboratory 49 did not accept any AD in the cluster identification and failed to include REF4 in the reported cluster.

In general, a high similarity was seen from the reported results using both cgMLST/Enterobase (3002 loci) and 'wgMLST' schemes (15874 and 3696 loci). Two laboratories (148 and 127), both using the Enterobase scheme, reported higher AD (50-52) for REF2, REF5 and REF7 outside the cluster compared with the other laboratories (Figure 7). A reason for this variation could be analysing directly in Enterobase, as all other laboratories performed the analysis by use of the Enterobase scheme included into BioNumerics, SeqPhere or a Galaxy-In house approach. However, the use of different methods for allele calling or the use of different assemblers did not seem to influence the results in this EQA.

The EQA provider's analysis of the submitted raw data showed a very high concordance, when using the standardised cgMLST/Enterobase analysis (93% below 3 AD, Figure 9). The EQA provider could not verify the differences of 50-52 alleles for REF2, REF5 and REF7 outside the cluster as were reported by laboratories 148 and 127 (Figure 8).

The laboratories performing SNP analysis reported a clear separation of the cluster and the non-cluster isolates, despite some variation in the distance outside the cluster (Figure 7). Particularly, laboratory 127 reported a high SNP distance for REF9 in one of their analysis. Laboratory 127 performed two additional SNP analyses and found AD inside the cluster for both 0-1. This was in line with all the other SNP analyses reported by the participants and the EQA provider.

On the other hand, the analysis by the EQA provider (using the standardised cgMLST/Enterobase analysis) of the data from laboratory 108 showed a higher number of ADs ranging from 6-12 for all isolates. This laboratory provided Ion Torrent data for which the EQA provider's analysis was not optimised, making correct assembly difficult (also seen in the previous EQAs). Thus, the observed higher ADs seem to be method artefacts, and subsequently the use of Ion Torrent data can complicate the comparison of data in investigation of multi country outbreaks if only allelic method is used.

As seen in the previous EQAs, the two approaches to analyse WGS-derived data (allele- and SNP-based analysis) showed correct results, albeit sequences generated with Ion Torrent cannot be compared with those generated by other sequencers. In a comparative analysis of core genome MLST and SNP typing, Pearch et al. also showed congruent results [18].

The main reported QC parameters in the EQA-11 was a threshold of coverage. However also genus/species confirmation, genome size and cgMLST allele calls were essential for assessing the end-use of the data. For one participant, the EQA provider detected 5-10% *Pseudomonas tolaasii* contamination in 70% of the isolates. However, this did not influence the identification of the correct cluster using cgMLST.

In the assessment of the additional EQA-provided genomes, all but one of the participants successfully identified the one cluster isolate (genome 6) and the two non-cluster isolates (genomes 4 and 5) with the good quality. Laboratory (106) did not analyse one of the good quality genomes due to difficulties with downloading the sequence.

For the three cluster genomes of poor quality, all laboratories identified the low coverage for genome 3 and the contamination (20%) of a different *Salmonella* ST34 for genome 2. The contamination of 10% with *Citrobacter* was apparently more difficult to detect as only 71% (10/14) reported this.

As reported by many of the participants, the three cluster genomes (genomes 1-3) of poor quality needed to be re-tested, but compared with EQA-10, more participants also tried to perform a cluster analysis and describe the observations on the sub-optimal data, and not just only respond to a need to rerun the sequencing. Most often for genome 1 with the 10% contamination many of the participants performed an analysis and despite the contamination, correctly suggested the genome to be part for the cluster.

For genome 2 (20% added *Salmonella*) and genome 3 (low coverage) a few participants performed an analysis, but many gave a thorough description of observations. It was more difficult to conclude whether genome 2 was part of the cluster as detection of a contamination with the same species is more challenging using allele-based method. Only laboratory 108 using SNP analysis correctly and suggested genome 2 to be part for the cluster. For genome 3, only one laboratory (134) suggested the genome to be part of the cluster and gave a very adequate description 'In an outbreak situation, I would not exclude this strain from the cluster but definitively need to re-isolate, re-extract and re-run this sample to conclude'.

5. Conclusions

Twenty laboratories participated in the EQA-11 scheme: Eight (40%) performed MLVA and 19 participants (95%) performed cluster analysis using one or more methods. In EQA-11, 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 analyses (95% and 98%). Five out of eight participants correctly assigned the MLVA profile for all ten isolates of *S.* Typhimurium and six out of eight in MLVA *S.* Enteritidis. No single laboratory was responsible for all detected errors.

Nineteen laboratories participated in the molecular cluster analysis using PFGE and/or MLVA and/or WGS. As in the EQA-10, the evaluation of the cluster analysis was entirely based on a categorisation from WGS data, which influenced the conclusion for the reported PFGE and MLVA results. A correct cluster delineation is in general difficult to achieve using a less discriminatory methods and in EQA-11, all the test isolates for the cluster analysis had the same MLVA profile.

Six laboratories participated using PFGE for cluster analysis; three of them only used PFGE-derived data for analysis. None the six laboratories correctly identified the cluster, however five of them included the five cluster isolates in the reported cluster. The number of participants only using PFGE was low, whether this is due to changing towards WGS or not participating in the EQA is for now unclear. Four laboratories performed cluster analysis using MLVA-derived data, which is a decrease from EQA-10. As all the test isolates had the same MLVA profile, none of the laboratories reported the correct cluster of closely-related isolates, however two of the laboratories did report the correct MLVA profile for the 10 isolates.

Fourteen laboratories performed cluster analysis using WGS-derived data, a small decrease of one laboratory compared with last year. The performance was very high, 13 (93%) of participants correctly identifying the cluster of closely-related isolates. One laboratory with an incorrect result used cgMLST and selected a very strict cut-off. Thirteen of 14 (93%) laboratories preferred an allele-based method and only 7% (1/14) used SNP as the main reported cluster analysis. This was a decrease of SNP analyses compared with the previous EQA-10 (from 3/15; 20%).

The SNP analysis and allele-based analysis showed comparable results in EQA-11. Despite different approaches for analysing and different methods probably being used (different thresholds for allele calling, including or not including missing alleles in the analysis, assembly based and/or mapping based allele calling etc.) both methods had a uniform interpretation and a clear separation of cluster and non-cluster isolates. The only difference was by one laboratory in the interpretation of REF4`s relation to the cluster.

A new addition to the cluster analysis was introduced in EQA-10 and continued in EQA-11. The EQA provider made sequence data of six isolates accessible, and the participants were asked to include these in the cluster analysis and report characteristics and quality issues. All but one participant made a correct conclusion whether the three genomes of high quality were part of the cluster. All participants were able to identify the genome with low quality issue and the genome with 20% added *Salmonella* ST34 in the *Salmonella* ST11. The presence of 10% *Citrobacter* in the genome 1 was identified by 71% of the participants.

For encouraging the participants to perform the more challenging analysis of the genomes with poor quality, the wording of the questions in the online form was rephrased. This year, an increased number of participants proceeded with this part and gave an assessment of the data and determined whether the genome was part of the defined cluster. The current EQA scheme for typing of *Salmonella enterica* subsp. *enterica* is the eleventh 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/EEA. In 2020, ECDC has opened up the possibility of submitting WGS data for *Salmonella* to TESSy to be used for EU-wide surveillance and cross-sector comparison.

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 of the most common serovars in Europe and MLVA typing provides moderate discrimination within isolates of both serovars. Some of the laboratories, who 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 with WGS, but its usefulness for EU level surveillance is limited to screening of isolates for sequencing.

The laboratories are encouraged to submit their high-quality molecular typing data to TESSy as close to real time as possible.

The laboratories are recommended to 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.

The laboratories are encouraged to move towards WGS. For laboratories who need a lot of troubleshooting in the use of PFGE or MLVA, it could be an advantage to implement WGS.

6.2 ECDC and FWD-Net

ECDC is actively working with FWD-Net to support the improvement of the quality of sequence data generation and analysis including agreed cut-off values for cluster detection through appropriate means like EQA schemes, expert exchange visits and workshops.

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

6.3 EQA provider

The evaluation of the provided genome sequences was a success, almost all the participants performed the analysis and identified the modifications introduced by the EQA provider. This part can be expanded further.

The interpretation of cluster cut-off is an important topic, and the EQA provider will suggest an open 'cut-off' discussion for the next FWD-Network meeting.

It is advised to include rare serotypes in the cluster analysis.

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 from: 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 from: 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 2019 Zoonoses Report. EFSA Journal 2021;19(2):6406. Available from: https://efsa.onlinelibrary.wiley.com/doi/epdf/10.2903/j.efsa.2021.6406
4. European Centre for Disease Prevention and Control. Laboratory standard operating procedure for MLVA of *Salmonella enterica* serotype Typhimurium. Stockholm: ECDC; 2011. Available from: 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 from: 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 from: http://www.iso.org/iso/catalogue_detail.htm?csnumber=29366.
7. Warwick Medical School. EnteroBase [Internet]. Coventry: University of Warwick; 2018 [cited 21 August 2018]. Available from: 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 [Internet]. Copenhagen: SSI; 2018 [cited 21 August 2018]. Available from: http://www.github.com/ssi-dk/SerumQC.
11. Inns T, Ashton PM, Herrera-Leon S, et al. Prospective use of whole genome sequencing (WGS) detected a multi-country outbreak of *Salmonella* Enteritidis. Epidemiol Infect 2017; 145:289–298. https://pubmed.ncbi.nlm.nih.gov/27780484/
12. Dallman, T, Inns, T, Jombart, T, et al. Phylogenetic structure of European *Salmonella* Enteritidis outbreak correlates with national and international egg distribution network. Microbial Genomics. Published online 26 May 2016. doi:10.1099/mgen.0.000070 https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5320589/
13. Lisa Di Marcantonio, Anna Janowicz, Katiuscia Zilli, Romina Romantini ,Stefano Bilei, Daniela Paganico, Tiziana Persiani, Guido Di Donato and Elisabetta Di Giannatale. Genomic Comparison of *Salmonella* Enteritidis Strains Isolated from Laying Hens and Humans in the Abruzzi Region during 2018. Pathogens 2020, 9, 349; doi:10.3390/pathogens9050349 https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7281747/
14. Besser JM, Carleton HA, Trees E, Stroika SG, Hise K et al. Interpretation of whole-genome sequencing for enteric disease surveillance and outbreak investigation. Foodborne Pathog Dis 2019;16:504–512 https://pubmed.ncbi.nlm.nih.gov/31246502/
15. Anthony M. Smith, Nomsa P. Tau, Hlengiwe M. Ngomane, Phuti Sekwadi, Ntsieni Ramalwa, Krishnee Moodley, Chetna Govind, Shameema Khan, Moherndran Archary and Juno Thomas. Whole-genome sequencing to investigate two concurrent outbreaks of *Salmonella* Enteritidis in South Africa, 2018 Smith et al., Journal of Medical Microbiology 2020;69:1303–1307DOI 10.1099/jmm.0.001255
16. Deng X, Shariat N, Driebe EM, Roe CC, Tolar B, Trees E, et al. Comparative Analysis of Subtyping Methods against a Whole-Genome-Sequencing Standard for *Salmonella enterica* Serotype Enteritidis. J Clin Microbiol. 2015 Jan;53(1):212-8.
17. Den Bakker HC, et al. Rapid whole-genome sequencing for surveillance of *Salmonella enterica* serovar Enteritidis. Emerging Infectious Diseases 2014; 20: 1306–1314. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4111163/
18 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 Hygiene Graz
Belgium	NRC *Salmonella*	Sciensano
Czechia	NRL for *Salmonella*	The 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* and other bacterial enteric pathogens	Robert Koch Institute
Greece	National Reference Centro for *Salmonella*	University of West Attica
Hungary	Division of Microbiological Reference Laboratories	National Public Health Center
Ireland	NSSLRL	University Hospital Galway
Italy	Antibiotico Resistenza e Patogeni Speciali	Istituto Superiore Di Sanità
Latvia	Infectology Centre of Latvia	Riga East University hospital
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
Slovakia	NRC for Salmonelloses Laboratory of Molecular Diagnostics	Public Health Authority of the Slovak Republik
Slovenia	Department for Public Health Microbiology	National Laboratory of Health, Environment and Food
Spain	Unidad de Enterobacterias	Instituto de Salud Carlos III
Sweden	Unit for Laboratory Surveillance of Bacterial Pathogens	The Public Health Agency of Sweden
The Netherlands	IDS	RIVM

Annex 2. Participation overview EQA-10 and 11

Laboratory	2019 to 2020 (EQA-10)						2020 to 2021 (EQA-11)					
	Participation (min. 1 part)	MLVA		Cluster			Participation (min. 1 part)	MLVA		Cluster		
		STm	SE	PFGE	MLVA	WGS		STm	SE	PFGE	MLVA	WGS
19	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	X	X		
96	X			X			X			X		
100	X	X	X			X	X	X	X			X
106	X			X		X	X			X		X
108	X	X	X			X	X					X
127	X			X		X	X			X		X
128												
129	X					X	X					X
130							X			X		
132												
134	X					X	X					X
135*	X	X	X			X	X					X
138							X				X	
140												
142	X	X	X	X	X	X	X	X	X	X	X	X
144	X	X	X	X	X		X	X	X			
145												
147	X	X	X		X	X	X	X	X		X	X
148	X					x	X					X
149	X					X	X	X	X			X
150	X					X						
Total number of participants	**19**	**8**	**9**	**7**	**5**	**15**	**20**	**8**	**8**	**6**	**4**	**14**

**: previously laboratory 77*

Annex 3. Scores of MLVA results S. Typhimurium

Test isolates no. /allel

Lab. no.	STm1					STm2					STm3					STm4					STm5				
	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	13	-2	-2	211	3	12	9	-2	211	3	15	11	-2	211	4	4	12	8	211
19	2	17	-2	-2	211	3	13	-2	-2	211	3	12	9	-2	211	3	15	11	-2	211	4	4	12	8	211
55	2	17	-2	-2	211	3	13	-2	-2	211	3	12	9	-2	211	3	15	11	5	211	4	4	12	8	211
92	2	17	-2	-2	211	3	13	-2	-2	211	3	12	9	-2	211	3	15	11	-2	211	4	4	12	8	211
100	2	17	-2	-2	211	3	13	-2	-2	211	3	12	9	-2	211	3	15	11	-2	211	4	4	12	8	211
142	2	17	-2	-2	211	3	13	-2	-2	211	3	12	9	-2	211	3	15	11	-2	211	4	-2	12	8	211
144	2	17	-2	-2	211	3	13	-2	-2	211	3	12	9	-2	211	3	15	11	-2	211	4	4	12	8	211
147	2	17	-2	19	211	3	13	-2	-2	211	3	12	9	-2	211	3	15	11	-2	211	4	4	12	8	211
149	2	17	-2	-2	211	3	13	-2	-2	211	3	12	9	-2	211	3	15	11	-2	211	4	4	12	8	211

Lab. no.	STm6					STm7					STm8					STm9					STm10				
	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	3	14	13	22	311	3	11	11	-2	211	2	9	-2	13	212	5	15	7	-2	111	2	20	12	9	212
19	3	14	13	22	311	3	11	11	-2	211	2	9	-2	13	212	5	15	7	-2	111	2	20	12	9	212
55	3	14	13	22	311	3	11	11	-2	211	2	9	-2	13	212	5	15	7	-2	111	2	20	12	9	212
92	3	14	13	22	311	3	11	11	-2	211	2	9	-2	13	212	5	15	7	-2	111	2	20	12	9	212
100	3	14	13	22	311	3	11	11	-2	211	2	9	-2	13	212	5	15	7	-2	111	2	20	12	9	212
142	3	14	13	22	311	3	11	11	-2	211	2	9	22	13	212	5	15	7	-2	111	2	20	12	9	212
144	3	14	13	22	311	3	11	11	-2	211	2	9	-2	13	212	5	15	7	-2	111	2	20	12	9	212
147	3	14	13	22	311	3	11	11	-2	211	2	9	-2	13	212	5	15	7	-2	111	2	20	12	9	212
149	3	14	13	22	311	3	11	11	-2	211	2	9	-2	13	212	5	15	7	-2	111	2	20	12	9	212

Purple: repeat isolates in EQA-4 to -11
Pink: incorrect results.

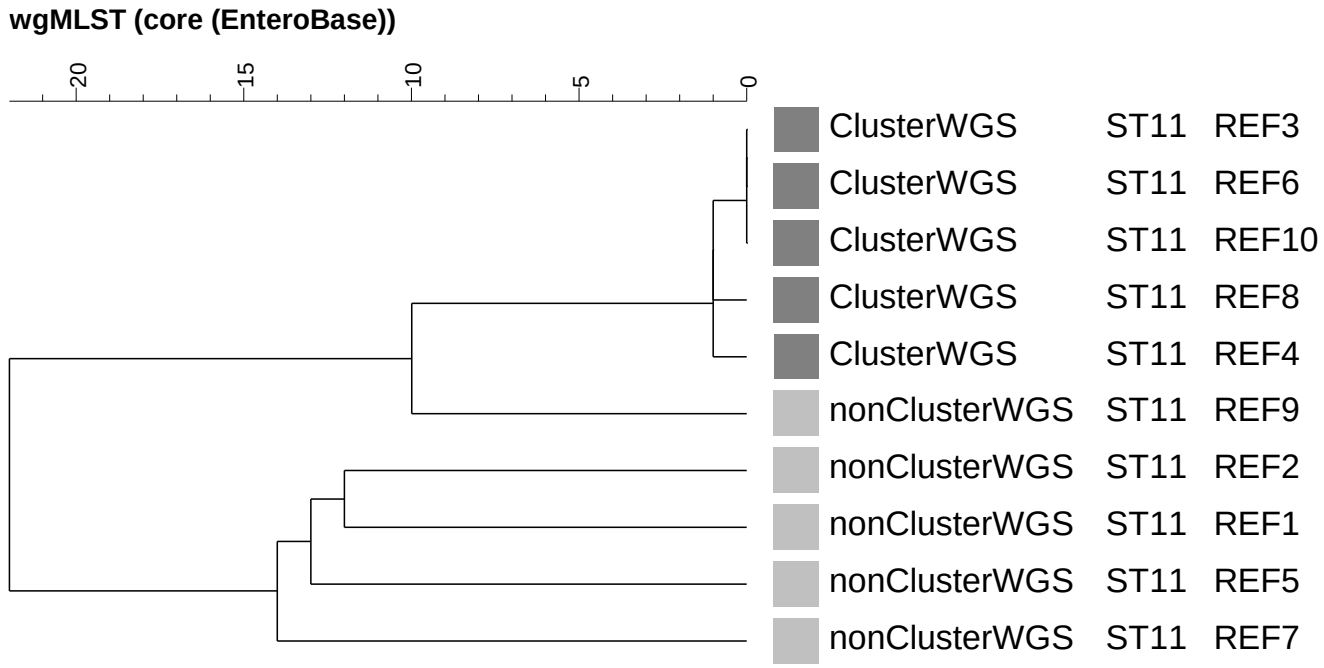
Annex 4. Scores of MLVA results *S.* Enteritidis

	Test isolates no. /allel																								
	SE1					SE2					SE3					SE4					SE5				
Lab. no.	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	9	7	3	2	3	9	4	4	1	2	10	7	3	2	3	10	5	4	1	2	9	7	3	1
19	2	9	7	3	2	3	9	4	4	1	2	10	7	3	2	3	10	5	4	1	2	9	7	3	1
55	2	9	7	3	2	3	9	4	4	1	2	10	7	3	2	3	10	5	2	1	2	9	7	3	1
92	2	9	7	3	2	3	9	4	4	1	2	9	7	3	2	3	10	5	4	1	2	9	7	3	1
100	2	9	7	3	2	3	9	4	4	1	2	10	7	3	2	3	10	5	4	1	2	9	7	3	1
142	2	9	7	3	2	3	9	4	4	1	2	10	7	3	2	3	10	5	4	1	2	9	7	3	1
144	2	9	7	3	2	3	9	4	4	1	2	10	7	3	2	3	10	5	4	1	2	9	7	3	1
147	2	9	7	3	2	3	9	4	4	1	2	10	7	3	2	3	10	5	4	1	2	9	7	3	1
149	2	9	7	3	2	3	9	4	4	1	2	10	7	3	2	3	10	5	4	1	2	9	7	3	1

	Test isolates no. /allel																								
	SE6					SE7					SE8					SE9					SE10				
Lab. no.	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	7	12	7	1	3	16	5	5	2	2	4	4	4	1	3	11	4	4	1	1	10	7	3	2
19	2	7	12	7	1	3	16	5	5	2	2	4	4	4	1	3	11	4	4	1	1	10	7	3	2
55	2	7	12	7	1	3	16	5	5	2	2	4	4	4	1	3	11	4	4	1	1	10	7	3	2
92	2	7	12	7	1	3	16	5	5	2	2	4	4	4	1	3	11	4	4	1	1	10	7	3	2
100	2	7	12	7	1	3	16	5	5	2	2	4	4	4	1	3	11	4	4	1	1	10	7	3	2
142	2	7	12	7	1	3	16	5	5	2	2	4	4	4	1	3	11	4	4	1	1	10	7	3	2
144	2	7	12	7	1	3	16	5	5	2	2	4	4	4	1	3	11	4	4	1	1	10	7	3	2
147	2	7	12	7	1	3	16	5	5	2	2	4	4	4	1	3	11	4	4	1	1	10	7	3	2
149	2	7	12	7	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 -11
Pink: incorrect results.

Annex 5. EQA provider cluster analysis based on WGS-derived data



Single linked dendrogram of core genome multi-locus sequence typing (cgMLST) profiles of Salmonella EQA-11 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 6. Reported cluster of closely related isolates based on PFGE-derived data

Laboratory	Reported cluster	Corresponding REF isolates	Correct
Provider		REF3, REF4, REF6, REF8, REF10	
92	5042, 5237, 5269, 5382, 5459	REF4, REF8, REF10, REF9, REF3	No
96	5424, 5535, 5658, 5704, 5885, 5952, 5983	REF6, REF3, REF5, REF9, REF10, REF4, REF8	No
106	5013, 5087, 5171, 5308, 5715, 5872	REF4, REF3, REF8, REF10, REF6, REF9	No
127	5393, 5404, 5453, 5565, 5667, 5945	REF10, REF3, REF9, REF6, REF4, REF8	No
130	5806, 5117, 5517, 5242, 5869, 5083, 5280	REF8, REF5, REF10, REF9, REF4, REF2, REF3	No
142	5184, 5281, 5309, 5547, 5616, 5632, 5782	REF9, REF10, REF3, REF8, REF6, REF4, REF5	No

Annex 7. Reported sequencing details

Sequencing performed	Protocol (library preparation)	Commercial kit	Sequencing platform
In own laboratory	Commercial kits	NexteraXT	NextSeq
In own laboratory	Commercial kits	Nextera™ XT DNA Library Preparation Kit	NextSeq
In own laboratory	Commercial kits	Illumina Nextera DNA Prep	MiSeq
In own laboratory	Commercial kits	Illumina Nextera DNA Flex	MiSeq
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	Nextera DNA Flex	NextSeq
Externally	Commercial kits	Nextera XT DNA preparation kit	MiSeq
In own laboratory	Commercial kits	Nextera XT	MiSeq
In own laboratory	Commercial kits	Nextera Flex Illumina	MiniSeq Illumina
In own laboratory	Commercial kits	Illumina DNA prep kit	NextSeq
In own laboratory	Commercial kits	Nextera (Illumina)	MiSeq
In own laboratory	Commercial kits	NexteraXT (Illumina)	NextSeq
In own laboratory	Commercial kits	Nextera XT DNA Library Kit, Illumina	MiSeq
In own laboratory	Commercial kits	Kapa HyperPlus (Roche)	NextSeq

Annex 8. Reported cluster of closely related isolates based on MLVA-derived data

Laboratory	Reported cluster	Corresponding to REF isolates	Correct
Provider		REF3, REF4, REF6, REF8, REF10	
55	5133, 5244, 5313, 5473, 5539, 5617, 5944, 5954	REF2, REF8, REF5, REF1, REF9, REF3, REF7, REF4	No
138	5040, 5043, 5347, 5372, 5469, 5489, 5705, 5781, 5790, 5757	REF9, REF7, REF2, REF5, REF3, REF8, REF10, REF4, REF6, REF1	No
142	5184, 5281, 5309, 5462, 5537, 5547, 5580, 5616, 5632, 5782	REF9, REF10, REF3, REF7, REF1, REF8, REF2, REF6, REF4, REF5	No
147	5112, 5126, 5183, 5403, 5425, 5548, 5583, 5850, 5926, 5993	REF2, REF8, REF9, REF10, REF6, REF3, REF7, REF5, REF4, REF1	No

** Laboratory 138 had a typing error REF1 is 5957 and not 5757*

Annex 9. Reported MLVA profile data

				Laboratory ID			
Isolate number	ST	MLVA scheme	Provider	55	138	142	147
REF1	11	*S.* Enteritidis	3-10-5-4-1	3-10-5-4-1	2-11-7-7-2	3-10-5-4-1	3-10-5-4-1
REF2	11	*S.* Enteritidis	3-10-5-4-1	3-10-5-4-1	2-11-7-7-2	3-10-5-4-1	3-10-5-4-1
REF3‡#	11	*S.* Enteritidis	3-10-5-4-1	3-10-5-4-1	2-11-7-7-2	3-10-5-4-1	3-10-5-4-1
REF4‡	11	*S.* Enteritidis	3-10-5-4-1	3-10-5-4-1	2-11-7-7-2	3-10-5-4-1	3-10-5-4-1
REF5	11	*S.* Enteritidis	3-10-5-4-1	3-10-5-4-1	2-11-7-7-2	3-10-5-4-1	3-10-5-4-1
REF6‡	11	*S.* Enteritidis	3-10-5-4-1	3-NA-5-4-1	2-11-7-7-2	3-10-5-4-1	3-10-5-4-1
REF7	11	*S.* Enteritidis	3-10-5-4-1	3-10-5-4-1	2-11-7-7-2	3-10-5-4-1	3-10-5-4-1
REF8‡	11	*S.* Enteritidis	3-10-5-4-1	3-10-5-4-1	2-11-7-7-2	3-10-5-4-1	3-10-5-4-1
REF9	11	*S.* Enteritidis	3-10-5-4-1	3-10-5-4-1	2-11-7-7-2	3-10-5-4-1	3-10-5-4-1
REF10‡#	11	*S.* Enteritidis	3-10-5-4-1	3-NA-5-4-1	2-11-7-7-2	3-10-5-4-1	3-10-5-4-1

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

Annex 10. Reported cluster of closely related isolates based on WGS-derived data

Laboratory	Reported cluster	Corresponding to REF isolates	Correct
Provider		REF3, REF4, REF6, REF8, REF10	
19	5226, 5552, 5666, 5904, 5932	REF4, REF10, REF3, REF6, REF8	Yes
36	5135, 5260, 5668, 5773, 5915	REF10, REF4, REF6, REF8, REF3	Yes
49	5074, 5177, 5732, 5808	REF10, REF3, REF8, REF6	No
100	5369, 5599, 5611, 5655, 5907	REF3, REF10, REF8, REF4, REF6	Yes
106	5087, 5171, 5308, 5715, 5013	REF3, REF8, REF10, REF6, REF4	Yes
108	5077, 5105, 5199, 5692, 5886	REF10, REF4, REF6, REF8, REF3	Yes
127	5393, 5404, 5565, 5667, 5945	REF10, REF3, REF6, REF4, REF8	Yes
129	5272, 5467, 5677, 5888, 5891	REF10, REF4, REF8, REF6, REF3	Yes
134	5512, 5513, 5545, 5689, 5794	REF4, REF6, REF8, REF10, REF3	Yes
135	5089, 5262, 5401, 5434, 5533	REF10, REF8, REF3, REF4, REF6	Yes
142	5281, 5309, 5547, 5616, 5632	REF10, REF3, REF8, REF6, REF4	Yes
147	5126, 5403, 5425, 5548, 5926	REF8, REF10, REF6, REF3, REF4	Yes
148	5426, 5461, 5553, 5847, 5951	REF3, REF6, REF4, REF10, REF8	Yes
149	5321, 5619, 5750, 5811, 5861	REF8, REF10, REF3, REF4, REF6	Yes

Annex 11. Reported SNP distance and allelic differences

SNP distances

			Laboratory ID			
Isolate number	ST	Provider	19*	108	127*	127**
REF1	11	53	55	59	55	85
REF2	11	66	68	71	68	116
REF3‡#	11	0	0	0	0	0
REF4‡	11	1	1	1	1	1
REF5	11	50	51	55	50	59
REF6‡	11	0	0¤	0	0	0
REF7	11	60	62	64	62	109
REF8‡	11	0	0	0	0¤	0¤
REF9	11	16	17	18	79	17
REF10‡#	11	0¤	0	0¤	0	0

Allelic differences

				Laboratory ID												
Isolate number	ST	Provider	19	36	49	49*	100	106	127	129	134	135	142	147	148	149
REF1	11	29	29	26	29	44	30	29	36	25	26	26	30	27	34	26
REF2	11	30	30	29	30	48	39	38	51	31	28	28	34	29	52	28
REF3‡#	11	0	0	0	0	1	0¤	0¤	1	0	0	0	0	0	1	0
REF4‡	11	1	1	1	2	2	1	1	1	1	1	2	1	1	3	1
REF5	11	28	28	26	28	42	40	39	50	23	26	26	30	27	50	26
REF6‡	11	1	0¤	0	0	1	0	0	0	0	0	0	0	0	0¤	0
REF7	11	32	32	32	33	51	42	41	50	34	32	32	34	33	51	32
REF8‡	11	1	0	0	0	2	0	0	0¤	0	0	0	0	0¤	1	0¤
REF9	11	10	10	10	10	14	10	9	12	9	10	11	11	11	13	10
REF10‡#	11	0¤	0	0¤	0¤	0¤	0	0	0	0¤	0¤	0¤	0¤	0	1	0

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

Annex 12. Additional reported QC parameters

Lab ID	1		2		3		4	
	Parameter	Threshold	Parameter	Threshold	Parameter	Threshold	Parameter	Threshold
36	Size of the single read files (*.fastq.gz)	Single read files <50.000 kb will be discarded immediately without further processing	Percent good cgMLST targets	95% (if all other parameters are sufficient)	N50	This is part of the assembly quality check	Contamination	This is already covered by the confirmation of the organism
49	Core %	*Salmonella* 90-96%= warn >=97%= pass	NrBAF multiple	> 20 = warn				
100	N50	100 000	Contig count	Less than 150 contigs	SAV	Cluster density, clusters passing filter and Q30 score were all according to Illumina recommendations	Contamination	Contamination Check (Mesh screen) Seqsphere
127	Number of N bases	< 3 %						
129	the number of good cgMLST loci	>90%	Contig Count					
135	N50	>10 000 bp	GC%	GC% between 51.6 and 52.3%	phred score	>30	Contamination	CheckM <4%
142	Phred score	28						
147	no. of contigs	200 bases (contigs shorter than 200 bases have to be ignored)						
148	N50	> 20 Kb (in house and in Enterobase)	Proportion of scaffolding placeholders (N's)	< 3%	Sickle 1.33 to trip the ends of short reads (FASTQ) of base calls with low quality scores	Argument : score = 10		

Annex 13. Calculated qualitative/quantitative parameters

		Laboratory 19									
Parameters	Ranges*	5001	5016	5122	5226	5481	5541	5552	5666	5904	5932
Detected species	{Se}	*Se*	*Se*	*Se*	*Se*	*Se*	*Se*	*Se*	*Se*	*Se*	*Se*
% Species 1		*95.1*	*95.0*	*95.5*	*95.0*	*95.0*	*95.0*	*95.0*	*94.4*	*95.0*	*95.3*
% Species 2	{<5%}	*0.0*	*0.1*	*0.1*	*0.0*	*0.0*	*0.1*	*0.0*	*0.0*	*0.3*	*0.1*
Unclassified reads (%)		*4.8*	*4.9*	*4.4*	*4.9*	*4.9*	*4.9*	*5.0*	*5.5*	*4.6*	*4.5*
Length at >25 x min. coverage (Mbp)	{>45 ∧ <53}	*4.7*	*4.7*	*4.7*	*4.7*	*4.7*	*4.7*	*4.7*	*4.5*	*4.8*	*4.7*
Length [1-25] x min. coverage (kbp)	{<250}	*5.1*	*2.6*	*4.3*	*9.1*	*14.9*	*10.3*	*10.6*	*127.2*	*5.6*	*4.9*
No. of contigs at 25 x min. coverage	{>0}	*124*	*97*	*114*	*85*	*152*	*135*	*107*	*303*	*108*	*115*
No. of contigs [1-25] x min. coverage	{<1000}	*7*	*2*	*5*	*8*	*16*	*12*	*14*	*112*	*7*	*6*
Average coverage	{>50}	*92*	*85*	*82*	*83*	*86*	*77*	*96*	*34*	*94*	*91*
No. of reads (x 1000)		*3047*	*2828*	*2685*	*2799*	*2837*	*2580*	*3226*	*1129*	*3174*	*3038*
Average read length		*144*	*143*	*145*	*143*	*145*	*143*	*143*	*145*	*145*	*145*
Average insert size		*297*	*282*	*291*	*289*	*292*	*284*	*284*	*265*	*280*	*262*
N50 (kbp)		*68*	*82*	*81*	*133*	*57*	*59*	*81*	*25*	*79*	*72*

		Laboratory 36									
Parameters	Ranges*	5132	5135	5260	5263	5497	5588	5668	5769	5773	5915
Detected species	{Se}	*Se*	*Se*	*Se*	*Se*	*Se*	*Se*	*Se*	*Se*	*Se*	*Se*
% Species 1		*97.2*	*97.3*	*97.4*	*97.4*	*97.1*	*97.2*	*96.7*	*97.3*	*97.2*	*97.1*
% Species 2	{<5%}	*0.2*	*0.0*	*0.1*	*0.1*	*0.1*	*0.1*	*0.5*	*0.0*	*0.0*	*0.0*
Unclassified reads (%)		*2.5*	*2.6*	*2.5*	*2.4*	*2.7*	*2.7*	*2.7*	*2.6*	*2.7*	*2.7*
Length at >25 x min. coverage (Mbp)	{>45 ∧ <53}	*4.7*	*4.7*	*4.7*	*4.7*	*4.7*	*4.7*	*4.8*	*4.7*	*4.7*	*4.7*
Length [1-25] x min. coverage (kbp)	{<250}	*0.0*	*0.0*	*6.9*	*3.8*	*11.5*	*0.6*	*3.5*	*3.1*	*0.8*	*3.3*
No. of contigs at 25 x min. coverage	{>0}	*32*	*39*	*47*	*60*	*29*	*30*	*62*	*60*	*40*	*58*
No. of contigs [1-25] x min. coverage	{<1000}	*0*	*0*	*5*	*4*	*5*	*1*	*3*	*3*	*1*	*2*
Average coverage	{>50}	*140*	*102*	*124*	*133*	*128*	*145*	*122*	*163*	*128*	*125*
No. of reads (x 1000)		*4513*	*3292*	*4003*	*4259*	*4154*	*4672*	*4005*	*5258*	*4122*	*4028*
Average read length		*148*	*148*	*148*	*148*	*147*	*147*	*148*	*148*	*149*	*148*
Average insert size		*346*	*380*	*334*	*342*	*336*	*326*	*350*	*333*	*384*	*362*
N50 (kbp)		*348*	*284*	*215*	*168*	*401*	*406*	*170*	*168*	*269*	*167*

		Laboratory 49									
Parameters	Ranges*	5074	5177	5201	5523	5729	5732	5747	5762	5808	5834
Detected species	{Se}	*Se*	*Se*	*Se*	*Se*	*Se*	*Se*	*Se*	*Se*	*Se*	*Se*
% Species 1		*98.3*	*97.6*	*98.1*	*98.1*	*96.0*	*97.5*	*98.3*	*98.3*	*98.1*	*98.4*
% Species 2	{<5%}	*0.0*	*0.0*	*0.0*	*0.1*	*0.0*	*0.0*	*0.0*	*0.0*	*0.4*	*0.1*
Unclassified reads (%)		*1.6*	*2.4*	*1.9*	*1.8*	*3.9*	*2.4*	*1.6*	*1.6*	*1.4*	*1.5*
Length at >25 x min. coverage (Mbp)	{>45 ∧ <53}	*4.7*	*4.7*	*3.9*	*4.7*	*4.7*	*4.7*	*4.7*	*4.7*	*4.8*	*4.7*
Length [1-25] x min. coverage (kbp)	{<250}	*0.0*	*0.0*	*859.9*	*0.0*	*0.0*	*7.8*	*0.0*	*0.0*	*0.0*	*0.0*
No. of contigs at 25 x min. coverage	{>0}	*23*	*24*	*22*	*18*	*30*	*26*	*22*	*23*	*25*	*30*
No. of contigs [1-25] x min. coverage	{<1000}	*0*	*0*	*3*	*0*	*0*	*1*	*0*	*0*	*0*	*0*
Average coverage	{>50}	*70*	*57*	*31*	*42*	*41*	*43*	*75*	*97*	*60*	*81*
No. of reads (x 1000)		*1267*	*1056*	*535*	*726*	*900*	*907*	*1245*	*1932*	*1100*	*1610*
Average read length		*266*	*260*	*283*	*281*	*219*	*230*	*287*	*241*	*268*	*240*
Average insert size		*331*	*345*	*419*	*398*	*269*	*268*	*445*	*273*	*327*	*273*
N50 (kbp)		*490*	*490*	*471*	*490*	*406*	*312*	*490*	*490*	*490*	*406*

		Laboratory 100									
Parameters	**Ranges***	**5008**	**5197**	**5369**	**5383**	**5599**	**5611**	**5655**	**5745**	**5907**	**5924**
Detected species	{Se}	*Se*	*Se*	*Se*	*Se*	*Se*	*Se*	*Se*	*Se*	*Se*	*Se*
% Species 1		*98.5*	*98.2*	*98.3*	*98.2*	*98.2*	*98.1*	*98.3*	*98.0*	*97.9*	*97.5*
% Species 2	{<5%}	*0.0*	*0.0*	*0.1*	*0.0*	*0.0*	*0.0*	*0.0*	*0.0*	*0.3*	*0.0*
Unclassified reads (%)		*1.4*	*1.7*	*1.6*	*1.8*	*1.7*	*1.8*	*1.7*	*1.9*	*1.7*	*2.4*
Length at >25 x min. coverage (Mbp)	{>45 ∧ <53}	*4.7*	*4.7*	*4.7*	*4.7*	*4.7*	*4.7*	*4.7*	*4.7*	*4.8*	*4.7*
Length [1-25] x min. coverage (kbp)	{<250}	*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}	*28*	*26*	*26*	*28*	*26*	*24*	*25*	*28*	*28*	*22*
No. of contigs [1-25] x min. coverage	{<1000}	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*0*
Average coverage	{>50}	*77*	*127*	*126*	*103*	*94*	*107*	*98*	*103*	*93*	*119*
No. of reads (x 1000)		*1570*	*2606*	*2616*	*2138*	*1958*	*2219*	*2007*	*2121*	*1936*	*2530*
Average read length		*235*	*234*	*233*	*233*	*234*	*235*	*236*	*236*	*238*	*227*
Average insert size		*310*	*308*	*305*	*304*	*303*	*307*	*314*	*315*	*340*	*288*
N50 (kbp)		*413*	*406*	*407*	*406*	*406*	*490*	*406*	*490*	*407*	*493*

		Laboratory 106									
Parameters	**Ranges***	**5013**	**5087**	**5171**	**5308**	**5328**	**5703**	**5715**	**5854**	**5871**	**5872**
Detected species	{Se}	*Se*	*Se*	*Se*	*Se*	*Se*	*Se*	*Se*	*Se*	*Se*	*Se*
% Species 1		*97.9*	*97.8*	*97.9*	*98.0*	*98.0*	*98.2*	*97.5*	*97.9*	*98.0*	*97.6*
% Species 2	{<5%}	*0.0*	*0.0*	*0.0*	*0.1*	*0.1*	*0.1*	*0.4*	*0.0*	*0.1*	*0.0*
Unclassified reads (%)		*2.0*	*2.1*	*2.0*	*1.9*	*1.9*	*1.7*	*2.0*	*2.0*	*2.0*	*2.4*
Length at >25 x min. coverage (Mbp)	{>45 ∧ <53}	*4.7*	*4.7*	*4.7*	*4.7*	*4.7*	*4.7*	*4.8*	*4.7*	*4.7*	*4.7*
Length [1-25] x min. coverage (kbp)	{<250}	*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}	*24*	*25*	*27*	*23*	*25*	*22*	*28*	*24*	*24*	*24*
No. of contigs [1-25] x min. coverage	{<1000}	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*0*
Average coverage	{>50}	*305*	*223*	*405*	*346*	*367*	*362*	*343*	*317*	*376*	*388*
No. of reads (x 1000)		*9710*	*7121*	*12943*	*11062*	*11668*	*11485*	*11162*	*10079*	*11955*	*12433*
Average read length		*150*	*150*	*150*	*149*	*149*	*149*	*149*	*149*	*149*	*149*
Average insert size		*470*	*496*	*420*	*412*	*400*	*377*	*414*	*421*	*437*	*492*
N50 (kbp)		*490*	*490*	*490*	*490*	*490*	*490*	*490*	*490*	*490*	*490*

		Laboratory 108									
Parameters	**Ranges***	**5077**	**5105**	**5108**	**5199**	**5207**	**5692**	**5819**	**5886**	**5899**	**5912**
Detected species	{Se}	*Se*	*Se*	*Se*	*Se*	*Se*	*Se*	*Se*	*Se*	*Se*	*Se*
% Species 1		*98.1*	*98.3*	*98.3*	*96.9*	*98.3*	*98.2*	*98.2*	*98.2*	*98.0*	*98.3*
% Species 2	{<5%}	*0.6*	*0.4*	*0.5*	*1.6*	*0.4*	*0.5*	*0.6*	*0.4*	*0.7*	*0.5*
Unclassified reads (%)		*1.3*	*1.3*	*1.2*	*1.4*	*1.2*	*1.3*	*1.3*	*1.3*	*1.3*	*1.2*
Length at >25 x min. coverage (Mbp)	{>45 ∧ <53}	*4.7*	*4.7*	*4.7*	*4.8*	*4.7*	*4.7*	*4.7*	*4.7*	*4.7*	*4.7*
Length [1-25] x min. coverage (kbp)	{<250}	*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} #	*679*	*640*	*555*	*1113*	*372*	*673*	*498*	*671*	*642*	*616*
No. of contigs [1-25] x min. coverage	{<1000} #	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*0*
Average coverage	{>50}	*119*	*120*	*119*	*106*	*111*	*116*	*116*	*115*	*118*	*118*
No. of reads (x 1000)		*2000*	*2000*	*2000*	*2000*	*2000*	*2000*	*2000*	*2000*	*2000*	*2000*
Average read length		*284*	*287*	*283*	*258*	*263*	*279*	*276*	*277*	*283*	*284*
Average insert size	#	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*0*
N50 (kbp)	#	*12*	*12*	*15*	*7*	*20*	*12*	*16*	*13*	*12*	*13*

		Laboratory 127									
Parameters	Ranges*	5275	5305	5393	5404	5453	5565	5667	5824	5922	5945
Detected species	{Se}	*Se*	*Se*	*Se*	*Se*	*Se*	*Se*	*Se*	*Se*	*Se*	*Se*
% Species 1		*95.4*	*95.3*	*92.9*	*96.0*	*96.8*	*96.7*	*95.9*	*95.4*	*92.8*	*96.8*
% Species 2	{<5%}	*0.0*	*0.0*	*0.0*	*0.1*	*0.0*	*0.1*	*0.0*	*0.0*	*0.1*	*0.0*
Unclassified reads (%)		*4.6*	*4.6*	*7.0*	*3.9*	*3.1*	*3.2*	*4.1*	*4.5*	*7.0*	*3.2*
Length at >25 x min. coverage (Mbp)	{>45 ∧ <53}	*4.7*	*4.7*	*4.7*	*4.7*	*4.7*	*4.8*	*4.7*	*4.7*	*4.7*	*4.7*
Length [1-25] x min. coverage (kbp)	{<250}	*1.0*	*0.0*	*0.0*	*30.0*	*6.8*	*8.8*	*3.2*	*0.0*	*1.4*	*0.0*
No. of contigs at 25 x min. coverage	{>0}	*87*	*89*	*52*	*102*	*74*	*109*	*74*	*94*	*65*	*27*
No. of contigs [1-25] x min. coverage	{<1000}	*2*	*0*	*0*	*16*	*6*	*11*	*4*	*0*	*2*	*0*
Average coverage	{>50}	*116*	*115*	*143*	*51*	*74*	*70*	*115*	*160*	*95*	*95*
No. of reads (x 1000)		*3905*	*3833*	*4948*	*1641*	*2407*	*2338*	*3774*	*5413*	*3173*	*3104*
Average read length		*142*	*145*	*142*	*148*	*148*	*147*	*147*	*141*	*145*	*147*
Average insert size		*358*	*406*	*362*	*453*	*467*	*460*	*452*	*350*	*404*	*477*
N50 (kbp)		*111*	*102*	*165*	*72*	*126*	*87*	*136*	*106*	*137*	*278*

		Laboratory 129									
Parameters	Ranges*	5272	5312	5351	5467	5515	5677	5798	5888	5891	5902
Detected species	{Se}	*Se*	*Se*	*Se*	*Se*	*Se*	*Se*	*Se*	*Se*	*Se*	*Se*
% Species 1		*97.3*	*98.0*	*98.0*	*95.5*	*95.8*	*95.9*	*96.0*	*95.5*	*95.8*	*96.1*
% Species 2	{<5%}	*0.0*	*0.1*	*0.0*	*0.1*	*0.0*	*0.0*	*0.0*	*0.3*	*0.1*	*0.1*
Unclassified reads (%)		*2.6*	*1.9*	*2.0*	*4.4*	*4.1*	*4.0*	*3.9*	*4.2*	*4.2*	*3.8*
Length at >25 x min. coverage (Mbp)	{>45 ∧ <53}	*4.6*	*4.7*	*4.7*	*4.7*	*4.7*	*4.7*	*4.7*	*4.8*	*4.7*	*4.7*
Length [1-25] x min. coverage (kbp)	{<250}	*50.1*	*11.6*	*22.0*	*4.7*	*2.9*	*4.4*	*0.5*	*2.9*	*4.3*	*1.8*
No. of contigs at 25 x min. coverage	{>0}	*179*	*142*	*136*	*91*	*69*	*87*	*57*	*68*	*58*	*65*
No. of contigs [1-25] x min. coverage	{<1000}	*43*	*13*	*17*	*5*	*3*	*4*	*1*	*3*	*4*	*2*
Average coverage	{>50}	*46*	*104*	*86*	*123*	*123*	*122*	*97*	*123*	*95*	*102*
No. of reads (x 1000)		*1480*	*3306*	*2731*	*4025*	*3991*	*4000*	*3159*	*4105*	*3091*	*3317*
Average read length		*149*	*149*	*149*	*149*	*149*	*149*	*149*	*149*	*149*	*149*
Average insert size		*471*	*408*	*437*	*400*	*383*	*386*	*394*	*360*	*393*	*389*
N50 (kbp)		*47*	*60*	*62*	*92*	*165*	*109*	*156*	*119*	*141*	*127*

		Laboratory 134									
Parameters	Ranges*	5512	5513	5522	5545	5559	5566	5678	5689	5794	5910
Detected species	{Se}	*Se*	*Se*	*Se*	*Se*	*Se*	*Se*	*Se*	*Se*	*Se*	*Se*
% Species 1		*97.2*	*92.2*	*95.2*	*96.7*	*95.7*	*97.4*	*97.8*	*97.7*	*97.3*	*98.3*
% Species 2	{<5%}	*0.4*	*3.3*	*2.0*	*0.9*	*1.4*	*0.5*	*0.2*	*0.5*	*0.7*	*0.0*
Unclassified reads (%)		*2.3*	*3.9*	*2.5*	*2.3*	*2.7*	*2.0*	*2.0*	*1.8*	*2.0*	*1.7*
Length at >25 x min. coverage (Mbp)	{>45 ∧ <53}	*4.7*	*4.8*	*4.7*	*4.7*	*4.6*	*0.0*	*4.7*	*4.7*	*4.7*	*4.7*
Length [1-25] x min. coverage (kbp)	{<250}	*0.0*	*57.9*	*0.0*	*0.0*	*58.1*	*4711.3*	*0.0*	*0.0*	*0.0*	*0.0*
No. of contigs at 25 x min. coverage	{>0}	*30*	*50*	*28*	*30*	*39*	*7*	*28*	*30*	*32*	*27*
No. of contigs [1-25] x min. coverage	{<1000}	*0*	*9*	*0*	*0*	*5*	*62*	*0*	*0*	*0*	*0*
Average coverage	{>50}	*58*	*39*	*94*	*67*	*33*	*21*	*98*	*54*	*46*	*80*
No. of reads (x 1000)		*1981*	*1571*	*3302*	*2253*	*1142*	*696*	*3235*	*1776*	*1502*	*2554*
Average read length		*151*	*151*	*151*	*151*	*151*	*151*	*151*	*151*	*151*	*149*
Average insert size		*236*	*148*	*194*	*256*	*218*	*266*	*243*	*273*	*285*	*321*
N50 (kbp)		*396*	*238*	*490*	*421*	*270*	*169*	*490*	*418*	*338*	*490*

		Laboratory 135									
Parameters	**Ranges***	**5037**	**5045**	**5073**	**5089**	**5262**	**5401**	**5431**	**5434**	**5533**	**5894**
Detected species	{Se}	*Se*	*Se*	*Se*	*Se*	*Se*	*Se*	*Se*	*Se*	*Se*	*Se*
% Species 1		*98.2*	*97.6*	*97.7*	*98.1*	*98.1*	*97.9*	*97.8*	*97.0*	*97.8*	*97.6*
% Species 2	{<5%}	*0.5*	*0.7*	*0.5*	*0.5*	*0.5*	*0.4*	*0.6*	*0.8*	*0.5*	*0.7*
Unclassified reads (%)		*1.2*	*1.5*	*1.7*	*1.3*	*1.3*	*1.5*	*1.4*	*2.0*	*1.5*	*1.5*
Length at >25 x min. coverage (Mbp)	{>45 ∧ <53}	*4.7*	*4.7*	*4.7*	*4.7*	*4.7*	*4.7*	*4.7*	*4.7*	*4.8*	*4.7*
Length [1-25] x min. coverage (kbp)	{<250}	*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}	*24*	*26*	*26*	*27*	*26*	*29*	*24*	*26*	*31*	*23*
No. of contigs [1-25] x min. coverage	{<1000}	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*0*
Average coverage	{>50}	*148*	*139*	*189*	*203*	*232*	*165*	*122*	*148*	*291*	*144*
No. of reads (x 1000)		*4740*	*4511*	*6087*	*6569*	*7496*	*5332*	*3922*	*4914*	*9595*	*4653*
Average read length		*151*	*151*	*151*	*151*	*151*	*151*	*151*	*151*	*151*	*151*
Average insert size		*384*	*391*	*379*	*353*	*346*	*346*	*375*	*397*	*368*	*389*
N50 (kbp)		*490*	*695*	*463*	*418*	*490*	*338*	*695*	*490*	*490*	*490*

		Laboratory 142									
Parameters	**Ranges***	**5184**	**5281**	**5309**	**5462**	**5537**	**5547**	**5580**	**5616**	**5632**	**5782**
Detected species	{Se}	*Se*	*Se*	*Se*	*Se*	*Se*	*Se*	*Se*	*Se*	*Se*	*Se*
% Species 1		*97.3*	*97.6*	*97.9*	*96.5*	*97.4*	*96.4*	*98.1*	*97.1*	*97.5*	*97.4*
% Species 2	{<5%}	*0.1*	*0.2*	*0.1*	*0.2*	*0.2*	*0.2*	*0.1*	*0.4*	*0.2*	*0.2*
Unclassified reads (%)		*2.4*	*2.0*	*1.7*	*2.9*	*2.2*	*3.2*	*1.6*	*1.9*	*2.1*	*2.2*
Length at >25 x min. coverage (Mbp)	{>45 ∧ <53}	*4.7*	*4.7*	*4.7*	*4.7*	*4.7*	*4.7*	*4.7*	*4.8*	*4.7*	*4.7*
Length [1-25] x min. coverage (kbp)	{<250}	*0.0*	*4.3*	*0.0*	*0.0*	*0.0*	*0.0*	*0.0*	*0.5*	*0.0*	*0.0*
No. of contigs at 25 x min. coverage	{>0}	*18*	*22*	*22*	*18*	*17*	*20*	*23*	*24*	*16*	*24*
No. of contigs [1-25] x min. coverage	{<1000}	*0*	*3*	*0*	*0*	*0*	*0*	*0*	*1*	*0*	*0*
Average coverage	{>50}	*46*	*65*	*59*	*36*	*59*	*43*	*103*	*70*	*50*	*67*
No. of reads (x 1000)		*904*	*1280*	*1169*	*711*	*1152*	*868*	*2168*	*1414*	*996*	*1323*
Average read length		*246*	*244*	*243*	*245*	*245*	*242*	*226*	*242*	*244*	*243*
Average insert size		*515*	*460*	*442*	*558*	*488*	*535*	*340*	*426*	*469*	*466*
N50 (kbp)		*490*	*490*	*490*	*490*	*490*	*490*	*490*	*490*	*694*	*490*

		Laboratory 147									
Parameters	**Ranges***	**5112**	**5126**	**5183**	**5403**	**5425**	**5548**	**5583**	**5850**	**5926**	**5993**
Detected species	{Se}	*Se*	*Se*	*Se*	*Se*	*Se*	*Se*	*Se*	*Se*	*Se*	*Se*
% Species 1		*97.3*	*93.9*	*96.3*	*96.0*	*97.2*	*97.3*	*95.9*	*95.7*	*96.9*	*97.7*
% Species 2	{<5%}	*0.1*	*0.1*	*0.1*	*0.1*	*0.4*	*0.1*	*0.2*	*0.1*	*0.0*	*0.1*
Unclassified reads (%)		*2.6*	*5.9*	*3.5*	*3.8*	*2.2*	*2.5*	*3.7*	*4.1*	*3.0*	*2.1*
Length at >25 x min. coverage (Mbp)	{>45 ∧ <53}	*4.7*	*4.7*	*4.7*	*4.7*	*4.8*	*4.7*	*4.7*	*4.7*	*4.7*	*4.7*
Length [1-25] x min. coverage (kbp)	{<250}	*0.0*	*3.9*	*0.0*	*0.0*	*0.0*	*0.0*	*0.0*	*3.4*	*0.0*	*0.0*
No. of contigs at 25 x min. coverage	{>0}	*27*	*70*	*29*	*22*	*31*	*32*	*27*	*36*	*33*	*27*
No. of contigs [1-25] x min. coverage	{<1000}	*0*	*6*	*0*	*0*	*0*	*0*	*0*	*1*	*0*	*0*
Average coverage	{>50}	*164*	*44*	*93*	*98*	*139*	*73*	*67*	*83*	*109*	*145*
No. of reads (x 1000)		*3393*	*856*	*1680*	*1774*	*2598*	*1292*	*1167*	*1516*	*2012*	*2796*
Average read length		*233*	*252*	*268*	*270*	*263*	*274*	*277*	*266*	*263*	*249*
Average insert size		*266*	*415*	*386*	*403*	*330*	*391*	*438*	*389*	*343*	*294*
N50 (kbp)		*406*	*125*	*376*	*694*	*431*	*352*	*382*	*281*	*406*	*439*

		Laboratory 148									
Parameters	Ranges*	5155	5205	5211	5426	5461	5553	5825	5847	5946	5951
Detected species	{Se}	*Se*	*Se*	*Se*	*Se*	*Se*	*Se*	*Se*	*Se*	*Se*	*Se*
% Species 1		*96.2*	*96.1*	*95.8*	*96.1*	*95.9*	*96.0*	*95.9*	*96.1*	*96.2*	*96.3*
% Species 2	{<5%}	*0.1*	*0.1*	*0.1*	*0.0*	*0.2*	*0.1*	*0.1*	*0.0*	*0.1*	*0.1*
Unclassified reads (%)		*3.7*	*3.7*	*4.0*	*3.8*	*3.8*	*3.9*	*3.9*	*3.8*	*3.7*	*3.6*
Length at >25 x min. coverage (Mbp)	{>45 ∧ <53}	*4.7*	*4.7*	*4.6*	*4.7*	*4.8*	*4.7*	*4.6*	*4.7*	*4.7*	*4.7*
Length [1-25] x min. coverage (kbp)	{<250}	*10.8*	*3.4*	*0.0*	*0.0*	*5.7*	*0.0*	*4.5*	*10.0*	*0.0*	*1.1*
No. of contigs at 25 x min. coverage	{>0}	*197*	*159*	*201*	*168*	*151*	*193*	*188*	*144*	*116*	*113*
No. of contigs [1-25] x min. coverage	{<1000}	*15*	*5*	*0*	*0*	*6*	*0*	*4*	*13*	*0*	*1*
Average coverage	{>50}	*218*	*250*	*235*	*238*	*222*	*196*	*193*	*209*	*243*	*257*
No. of reads (x 1000)		*6908*	*7953*	*7429*	*7604*	*7214*	*6225*	*6097*	*6728*	*7720*	*8216*
Average read length		*150*	*149*	*149*	*149*	*150*	*149*	*150*	*149*	*149*	*149*
Average insert size		*366*	*345*	*356*	*339*	*368*	*357*	*375*	*336*	*335*	*348*
N50 (kbp)		*48*	*54*	*46*	*64*	*62*	*49*	*46*	*78*	*89*	*97*

		Laboratory 149									
Parameters	Ranges*	5274	5321	5359	5365	5428	5619	5750	5811	5861	5930
Detected species	{Se}	*Se, Pt*	*Se, Pt*	*Se*	*Se*	*Se, Pt*	*Se*	*Se, Pt*	*Se, Pt*	*Se, Pt*	*Se, Pt*
% Species 1		*78.8*	*85.4*	*88.5*	*88.8*	*85.5*	*86.6*	*86.2*	*83.0*	*75.3*	*86.3*
% Species 2	{<5%}	*8.7*	*5.6*	*4.2*	*3.9*	*5.6*	*4.9*	*5.6*	*6.5*	*10.7*	*5.4*
Unclassified reads (%)		*11.4*	*8.4*	*6.8*	*6.9*	*8.4*	*8.1*	*7.4*	*10.1*	*12.4*	*7.7*
Length at >25 x min. coverage (Mbp)	{>45 ∧ <53}	*4.7*	*4.7*	*4.7*	*4.7*	*4.7*	*4.7*	*4.7*	*4.7*	*4.8*	*4.7*
Length [1-25] x min. coverage (kbp)	{<250}	*6.8*	*0.0*	*0.0*	*0.0*	*0.0*	*0.0*	*11.5*	*4.3*	*0.0*	*30.0*
No. of contigs at 25 x min. coverage	{>0}	*36*	*34*	*38*	*31*	*32*	*33*	*33*	*38*	*36*	*35*
No. of contigs [1-25] x min. coverage	{<1000}	*4*	*0*	*0*	*0*	*0*	*0*	*1*	*1*	*0*	*9*
Average coverage	{>50}	*59*	*93*	*112*	*114*	*91*	*94*	*94*	*89*	*70*	*101*
No. of reads (x 1000)		*2445*	*3702*	*4278*	*4204*	*3560*	*3618*	*3683*	*3551*	*3177*	*3929*
Average read length		*151*	*151*	*151*	*151*	*151*	*151*	*151*	*151*	*151*	*151*
Average insert size		*264*	*201*	*199*	*209*	*213*	*208*	*213*	*225*	*282*	*208*
N50 (kbp)		*260*	*359*	*376*	*376*	*359*	*376*	*284*	*284*	*305*	*270*

Se: Salmonella enterica
Pt: Pseudomonas tolaasii
#: QC values unreliable due to assembly issues for Ion Torrent data

Annex 14. Accessing Genomes

	Participant		EQA provider*	
Lab ID	Cluster	Genome 1	Contamination	Cluster
EQA provider		A cluster isolate (REF3/REF10) mixed with a *Citrobacter* (approx. 10%).	Yes	Yes
19	No	Genome 1 is likely contaminated with Citrobacter indicated by fail in Kraken analysis (8% Citrobacter identified), wrong genome size, low N50, many contigs and too many loci with multiple consensus. However, the core % is high (100%) and based on the cgMLST comparision, Genome 1 could be suspected as closely related isolate to the outbreak (with 10 AD to the representative outbreak genome). The isolate should therefore be restreaked for pure culture and resequenced for conformation.	Yes	Yes
36	No	For genome 1 a potential contamination with Citrobacter freundii was identified (Mash screen in SeqSphere). Further, the SeqSphere assembly statistics revealed an improper approx. genome size of 7.3 Mbases scattered over 9,357 contigs with an average contig length of 778 bp and N50 (assembled) of 1,147. As a consequence, the program detected only 34% good cgMLST targets. Genome 1 was therefore excluded from the analysis. However, if reference mapping instead of de novo assembly was applied (and therefore filtering out the reads of the contaminant) this genome could be analysed and would be considered as part of the outbreak cluster (0 AD to the representative genome 5135).	Yes	Yes
49	No	Analysis of genome 1 suggests it is contaminated with a non-*Salmonella* isolate, possibly Citrobacter freundii. The genome size was very large (>9MB), N50 was low, there was a high number of contigs, high numbers of bases Non ACGT and bases N. Testing with rMLST in PubMLST showed presence of *Salmonella* (95%) and Citrobacter freundii (5%). This genome had a core genome of 100% and there were just 6 cgMLST differences between it and the representative isolate from the cluster. However, to obtain a reliable result the isolate would need to be purity plated and re-extracted prior to re-sequencing to ensure testing of a pure culture.	Yes	Yes
100	Yes	Genome 1 has 0 AD from the representative isolate.	No	Yes
106	Yes	When performing the analysis without excluding samples with missing values in more than 10% of distance columns this genome belongs to the cluster as our cluster threshold is < 5 allelic differences and genome number 1 is 0 allelic difference. However, we would suggest to sequence this isolate again. This sample has missing values in more than 10% of distance columns. About 7% of the reads belongs to Citrobacter freundii and therefore many contigs belong to this species. This contamination yields a genome size of 8.6 Mb and 64.8% of good targets	Yes	Yes
108	Yes	Genome1 is 1 SNP from the representative cluster isolate. The genome size is too large and the results suggest contamination with Citrobacter freundii. In routine analysis we would reculture and rerun the sample.	Yes	Yes
127	No	Genome 1 should be excluded from the analysis because of sequence length (> 9,7 Mb) and many loci with multiple alleles. It seems that contamination may be present.	Yes	-
129	Yes	No allelic differences compared to index isolate in the cluster.	No	Yes
134	Yes	The assembly of Genome 1 was done by bwa (because not enough targets were found by using SKESA). With 98.8% of good targets found and an AD = 2 from the reference strain (5689), Genome 1 is considered a part of the cluster.	No	Yes
135	No	Genome 1 does not meet quality criteria the genome size, number of contigs and contamination is too high.% coverage of cgMLST alleles is too low therefore it is not possible to reliably state ifit is part of the cluster or not. isolate should be checked for contamination and WGS repeated from the start.	Yes	-
142	No	Possible contamination (mixed culture), assembly size 9.4Mb, 46 non-called cgMLST loci	Yes	-
147	No	8 AD to cluster representative isolate 5126 quality of sequencing is not good only 54.2% good targets, 1377 missing values genome size too big for *Salmonella* (9.8 Mb) possibly due to contamination	Yes	Yes
148	Yes	Genome 1 has the same HC5 as the reference reference Isolate 5 (5461), and it is at 1 allelic distance.	No	Yes
149	Yes	Genome 1 has good quality (99,2% good targets) and cluster together with 5321 (0 alleles differences).The sequences also contain a Citrobacter freundii-genome with low coverage, but that didn't seem to have an effect on the quality of the cgMLST data for *Salmonella*.	Yes	Yes

	Participant		EQA provider*	
Lab ID	Cluster	Genome 2	Contamination	Cluster
EQA provider		A cluster isolate (REF3/REF10) mixed with a *Salmonella* ST34 (approx. 20%), same species contamination.	Yes	Yes
19	No	Genome 2 is likely contaminated with other *Salmonell*a, here not indicated by fail in Kraken analysis, but by wrong genome size, low N50, many contigs and too many loci with multiple consensus. The core % is high (100%) and based on the cgMLST comparison, Genome 2 is not suspected as a closely related isolate to the outbreak (more than 200 AD to the representative outbreak genome). The isolate should however be restreaked for pure culture and resequenced for conformation.	Yes	No
36	No	For genome 2 SeqSphere identified only 10% good targets. Although no contamination had been detected by the software, the reads provided only 607/1,000 shared hashes with *Salmonella* enterica. Assembly statistics revealed an approx. genome size of only 3.9 Mbases on 6,859 contigs with an average contig length of 569 bp and N50 (assembled) of 632. Genome 2 was therefore excluded from the analysis.	Yes	-
49	No	Analysis of Genome 2 shows suggests that this sequence possibly contains a mixture of 2 or more *Salmonella*. The genome size (5.7MB) is slightly large for a *Salmonella*, the N50 is small there are a lot of non-ACGT and N bases called and there are >3500 assembly free multiple calls. rMLST analysis only detected *Salmonella*. Purity plates would be required to ensure pure cultures of *Salmonella* were sequenced and analysed to determine whether one of them matched the cluster.	Yes	-
100	No	Genome 2 has too low percentage of good targets to perform cgMLST cluster analysis. Genome size is too big (6.1 Mb). KmerFinder was used and two different organisms were identified. Contamination is possible.	Yes	-
106	No	When performing the analysis without excluding samples with missing values in more than 10% of distance columns this genome does not belong to the cluster as our cluster threshold is < 5 allelic differences and genome number 1 is 304 allelic difference. However, we would suggest to re-sequence this isolate. This sample has missing values in more than 10% of distance columns. It appears to be contaminated with a different *Salmonella* serotype which yields to a bad assembly with a genome size of 5.8% fragmented in more than 4000 contigs.	Yes	No
108	Yes	Genome2 is 2 SNPs from the representative cluster isolate. The genome size is too large, suggesting that the sample is contaminated, possibly with other *Salmonella*. Sequence type cannot be defined in our analysis. In routine analysis we would reculture and rerun the sample.	Yes	Yes
127	No	Genome 2 did not pass quality control because of low N50.The sample should be remade.	Yes	-
129	No	Poor quality: Different serotype than other isolates in the cluster or possible contamination with several serotypes (based on SeqSero: S.Essen). Average coverage and amount of data was good. However, contiq count was too high (3481). Did not get MLST type and the percentage of good targets of cgMLST for S.Enteritidis and S. enterica was too low. Need to start typing again from culture and DNA-extraction.	Yes	-
134	No	The sequence quality of Genome 2 is really bad. The genome assembly suggests a mix of 2 genomes of S. enteritidis (same species but different genotypes as the genome size is above the expected size). Only 14,2% of good targets are found and based on the distance matrix with representative strains from the cluster, at least 29 alleles are different. Genome 2 is not part of the cluster	Yes	No
135	No	Genome 2 does not meet quality criteria the genome size, number of contigs and contamination is too high. N50, completeness and % coverage of cgMLST alleles are too low. Therefore, it is not possible to reliably state if it is part of the cluster or not. Isolate should be checked for contamination and WGS repeated from the start.	Yes	-
142	No	Only 65% of cgMLST loci called, poor quality (high sequence duplication levels)	Yes	-
147	No	Too many AD to cluster representative isolate 5126 (971AD) quality of sequencing is not good only 82.7% good targets, 522 missing values genome size too big for *Salmonella* (5.8 Mb) possibly due to contamination	Yes	No
148	No	Genome 2 could not be analysed because N50 and length of the genome did not pass the Quality controls of Enterobase.	Yes	-
149	No	The quality of the sequence is too low (13,7% good targets). We cannot evaluate if genome 2 is part of the cluster or not. The genome has very high coverage, but also too high genome length. It might be more than one *Salmonella*-genome in the sample. We recomend to try to get pure cultures of the *Salmonella* strains before another DNA extraction.	Yes	-

	Participant		EQA provider*	
Lab ID	Cluster	Genome 3	Quality issue	Cluster
EQA provider		A cluster isolate (REF3/REF10) with altered coverage (reduced to 10x).	Yes	Yes
19	No	Genome 3 is of poor quality. The read coverage is low, resulting in a low core % and poor genome assembly (low N50, many contigs, many unknown bases, wrong genome size). The genome can therefore not be used for comparison and should be resequenced for better genome quality.	Yes	-
36	No	Due to insufficient data (very few reads and therefore too low coverage of only 10-fold) SeqSphere identified (assembly-based) only 50% good targets (40% of those with warnings). Genome 3 was excluded from the analysis as well. Although mapping-based the percentage of "good" cgMLST targets rose to 95,5% (indicating that the genome does not belong to the outbreak cluster --> 21 AD to the representative genome 5135), we would still exclude this genome from the analysis due to the extremely low coverage (and therefore the number of "good" targets with warnings).	Yes	No
49	No	Genome 3 could not be considered as part of the cluster as the core genome as the coverage (x10) and core genome (44%) would be far too low to get meaningful results from the data. The DNA extract should be measured and, if too low, re-extracted and wgs performed again.	Yes	-
100	No	Genome 3 has too low percentage of good targets to perform cgMLST cluster analysis. Also, the coverage is too low (10x).	Yes	-
106	No	When performing the analysis without excluding samples with missing values in more than 10% of distance columns this genome does not belong to the cluster as our cluster threshold is < 5 allelic differences and genome number 1 is 32 allelic difference. However, we would suggest to re-sequence this isolate. This sample has missing values in more than 10% of distance columns. This genome has less than 200K reads which produces a genome fragmented in about 700 contigs (more than usual) although we are able to cover 92% of the genome.	Yes	No
108	No	The sequence does not pass our quality parameters and cannot be included in the cluster assessment. The coverage is too low but the results suggest a *S*. Enteritidis ST11. In routine analysis we would rerun the sample and evaluate if it is a part of the outbreak cluster.	Yes	-
127	No	Genome 3 did not pass quality control because of low average coverage and low N50.The sample should be remade.	Yes	-
129	No	Poor quality: average coverage was only 9 and amount of sequence data too low. Also, contiq count was too high (3764). Did not get MLST type and the percentage of good targets of cgMLST for *S*.Enteritidis and *S*. enterica was too low. This strain should be rerun to get more data.	Yes	-
134	Yes	Only 50,2 % of good targets can be used for the comparison analysis. Of the 1505 targets available, only 21 are discrimant targets for classifying this strain including 4 displaying specific alleles characterizing this cluster. In an outbreak situation, I would not exclude this strain from the cluster but definitively need to re-isolate, re-extract en re-run this sample to conclude.	Yes	Yes
135	No	Genome 3 does not meet quality criteria: genome size, N50, average coverage and % coverage of cgMLST alleles are too low. Number of contigs and GC% are too high. Therefore, it is not possible to reliably state if it is part of the cluster or not WGS can be repeated from isolated DNA if the quality of this DNA meets criteria.	Yes	-
142	No	Low coverage, low number of reads, assembly length only 1.2Mb	Yes	-
147	No	Too many AD to cluster representative isolate 5126 (106 AD) coverage of sequence is low (10x average coverage unassembled) 96.1% good targets, 116 missing values --> quality of sequence average genome size is ok for *Salmonella* (4.7 Mb)	Yes	No
148	No	Genome 3 could not be analysed because N50 and the number of contigs of the genome did not pass the Quality controls of Enterobase.	Yes	-
149	No	The quality of the sequence is too low, with 44,0% good targets, quite fragmented and very low coverage. We recommend to do library preparation and sequencing again.	Yes	-

	Participant		EQA provider*	
Lab ID	Cluster	Genome 4	QC Accepted	Cluster
EQA provider		A non-cluster isolate (REF9), good quality of reads. 10 allelic difference to the cluster (REF3).	Yes	No
19	No	Genome 4 has 10 allele differences (10 AD) to the cluster representative genome and is therefore not considered part of the cluster.	Yes	No
36	No	The sequence reads were of good quality and included in our analysis. In cgMLST, genome 4 is indistinguishable from 5769 but 10-11 alleles different from the identified cluster.	Yes	No
49	No	The quality of Genome 4 was good and was compared with a representative isolate from the cluster. This showed that it differed from the cluster by 10 cgMLST so would not be included in the cluster. However, it was an exact match by cgMLST with 5747.	Yes	No
100	No	Genome 4 has 10 AD from the representative isolate and is not part of the cluster.	Yes	No
106	No	Our cluster threshold is < 5 allelic differences and genome number 4 is 9 allelic difference	Yes	No
108	No	The sample clusters with 5912 and is 19 SNPs from the representative cluster isolate.	Yes	No
127	No	Genome 4 is not a member of the outbreak cluster. Based on the SNP analysis, the 81 SNPs between genome 4 and representative sequence of the outbreak cluster (5945) exists.	Yes	No
129	No	Too many allelic differences compared to isolates in the cluster (AD to index=10)	Yes	No
134	No	98.5% of good targets are identified with this genome and coverage = 87. Genome 4 has identical cgMLST type than strian 5566 which is not part of the cluster.	Yes	No
135	No	Genome 4 meets all quality criteria, but differs 10 alleles from the closest cluster isolate	Yes	No
142	No	>7 AD (12AD)	Yes	No
147	No	12 AD to cluster representative isolate 5126 coverage (84x) and quality of sequence are good 99.2% good targets 23 missing values genome size is ok (4.7 Mb) cluster with isolate 5183 --> only 1 AD	Yes	No
148	No	Genome 4 has a different HC5 to the reference Isolate 5 (5461), and it is at 15 alleles distance. It could be considered as related, but not as part of the cluster.	Yes	No
149	No	Genome 4 has good quality (98,5% good targets) but has 10 alleles difference from 5321and is therefore not part of the cluster.	Yes	No

	Participant		EQA provider*	
Lab ID	Cluster	Genome 5	QC Accepted	Cluster
EQA provider		A non-custer isolate (REF1), good quality of reads assembled with SKESA to a FASTA file. 29 allelic difference to the REF3 in the cluster.	Yes	No
19	No	Genome 5 has 29 allele differences (29 AD) to the cluster representative genome and is therefore not considered part of the cluster.	Yes	No
36	No	The provided assembly file was of good quality and included in our analysis. In cgMLST, genome 5 is indistinguishable from 5132 but 26-27 alleles different from the identified cluster.	Yes	No
49	No	Genome 5 was a de novo sequence in fasta format. The quality of Genome 5 was good and was compared with a representative isolate from the cluster. This showed that it differed from the cluster by 29 cgMLST so would not be included in the cluster. However it was an exact match with 5762 by cgMLST.	Yes	No
100	No	Genome 5 has 30 AD from the representative isolate and is not part of the cluster.	Yes	No
106	No	We had a technical problem a dowloading this genome and could not be analysed	-	-
108	No	The sample clusters with 5207 and is 57 SNPs from the representative cluster isolate. Sequences in FASTA format receives a lower coverage (<20x) in our analysis.	Yes	No
127	No	Genome 5 is not a member of the outbreak cluster. Based on the SNP analysis, the 12 SNPs between genome 5 and representative sequence of the outbreak cluster (5945) exists.	Yes	No
129	No	Because no fastq reads were available, comparison with fasta was necessary. Too many allelic differences compared to isolates in the cluster (AD to index=25)	Yes	No
134	No	98.9% of good targets are identified with this genome which has the same cgMLST type than the strain 5910 (not part of the cluster).	Yes	No
135	No	Genome 5 meets all quality criteria (phred score excluded), but differs 26 alleles from the closest cluster isolate	Yes	No
142	No	>7 AD (34AD)	Yes	No
147	No	Too many AD to cluster representative isolate 5126 (27 AD) quality of sequence is good 98.9% good targets 33 missing values genome size is ok (4.7 Mb) cluster with isolate 59933 --> 0 AD	Yes	No
148	No	Genome 5 has a different HC5 and HC10 to the reference Isolate 5 (5461), and it is at 50 alleles distance.	Yes	No
149	No	Genome 5 has good quality (98,9% good targets) but has 26 alleles difference from 5321 and is therefore not part of the cluster.	Yes	No

	Participant		EQA provider*	
Lab ID	**Cluster**	**Genome 6**	**QC Accepted**	**Cluster**
EQA provider		A cluster isolate (REF3/REF10) good quality of reads.	Yes	Yes
19	Yes	Genome 6 has 0 allele differences (0 AD) to the cluster representative genome and is therefore considered part of the cluster.	Yes	Yes
36	Yes	The reads were of good quality and included in our analysis. cgMLST-based genome 6 belongs to the identified cluster (0-1 AD).	Yes	Yes
49	Yes	The quality of genome 6, while not perfect and having a core genome of just 95%, was sufficient to analyse. The genome was indistinguishable from the cluster by cgMLST so could be considered as part of the cluster.	Yes	Yes
100	Yes	Genome 6 has 0 AD from the representative isolate.	Yes	Yes
106	Yes	Our cluster threshold is < 5 allelic differences and genome number 5 is 0 allelic difference	Yes	Yes
108	Yes	Genome6 is 1 SNP from the representative cluster isolate.	Yes	Yes
127	Yes	Genome 6 is genetically linked to the outbreak cluster. Based on the SNP analysis, the 0 SNPs between genome 6 and the representative sequence of the outbreak cluster (5945) exists.	Yes	Yes
129	Yes	No allelic differences compared to index isolate in the cluster.	Yes	Yes
134	Yes	The sequencing quality of this genome is good, 98.8% of good targets identified and coverage =107. Genome 6 has the same cgMLST profile than the representative strain from the cluster (5689).	Yes	Yes
135	Yes	Genome 6 meets all quality criteria genome 6 is identical (0 alleles difference) to the representative isolate. Further isolate information needed to determine if there is an epidemiological link to confirm this genetic link.	Yes	Yes
142	Yes	0 AD	Yes	Yes
147	Yes	0 AD to cluster representative isolate 5126 coverage (106x) and quality of sequence are good 99.2% good targets 23 missing values genome size is ok (4.7 Mb)	Yes	Yes
148	Yes	Genome 6 has the same HC5 as the reference reference Isolate 5 (5461), and it is at 1 allelic distance.	Yes	Yes
149	Yes	Genome 6 has good quality (98,3% good targets) and has 0 alleles difference from 5321 and is therefore part of the cluster.	Yes	Yes

**Evaluated by the EQA provider, the "Cluster" result was changed from submitted cluster data if the description by the participant indicated differences.*

Annex 15. EQA-11 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-11 2020-2021

Dear Participant
Welcome to the eleventh External Quality Assessment (EQA-11) scheme for typing of *Salmonella* in 2020-2021.
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
- Czech Republic
- 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
- UK

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 ID's

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 137

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 21

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. Submitting cluster results

(State one answer only)

- Cluster analysis based on MLVA – Go to 22
- Do not wish to submit any cluster results based on MLVA analysis – Go to 32

22. Cluster analysis based on MLVA data

23. 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

24. MLVA scheme used:

Please indicate serovar and/or protocol

25. Please list the loci in scheme used

26. Locus 1:

27. Locus 2:

28. Locus 3:

29. Locus 4:

30. Locus 5:

31. 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					

32. Submitting cluster results

(State one answer only)

- Cluster analysis based on WGS data – Go to 33
- Do not wish to submit any cluster results based on WGS data - Go to 137

33. Cluster analysis based on WGS data

34. 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 36
- Allele based – Go to 43
- Other – Go to 35

35. If another analysis is used please describe your approach: - Go to 50

36. Please report the used SNP-pipeline

(reference if publicly available or in-house pipeline)

37. Please select the approach used for the SNP analysis

(State one answer only)

- Reference based – Go to 38
- Assembly based – Go to 41

38. 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)

39. Please indicate the read mapper used

(e.g. BWA, Bowtie2)

40. Please indicate the variant caller used

(e.g. SAMtools, GATK)

41. Please indicate the assembler used

(e.g. SPAdes, Velvet)

42. Please specify the variant caller used

(e.g. NUCMER)

43. Please select tools used for the allele analysis

(State one answer only)

- BioNumerics – Go to 45
- SeqSPhere – Go to 45
- Enterobase – Go to 45
- Other – Go to 44

44. If another tool is used please enter here:

45. Please indicate allele calling method:

(State one answer only)

- Assembly based and mapping based – Go to 46
- Only assembly based – Go to 46
- Only mapping based – Go to 47

46. Please indicate the assembler used

(e.g. SPAdes, Velvet)

47. Please select scheme used for the allele analysis

(State one answer only)

- Applied Math (wgMLST) – Go to 49
- Applied Math (cgMLST/Enterobase) – Go to 49
- Enterobase (cgMLST) – Go to 49
- Other – Go to 48

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

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

50. 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.

51. Please list the ID´s for the isolates included in the cluster of closely related isolates:

Please use semicolon (;) to separate ID´s

52. Select a representative isolate in the cluster

Indicate the isolate ID

53. 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		

54. Analysis of the EQA provided genomes

The six genomes uploaded by the EQA provider should be included in the analysis and evaluated.

Please evaluate this part as a simulation, mimicking a large outbreak situation in your country.
These genomes (1-6) are very important because they might solve the outbreak.
Each of the provided genomes should be assessed whether it could be a part of the cluster defined in first part.
Explain your assessment of each genome in details, please not just suggesting rerunning the sequence, but explain what you observe and what you would suggest as the conclusion.

This part is not evaluated with a final score in the evaluation report, however the EQA provider list the characteristics of the isolates.

55. 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

56. Explain your assessment of genome 1 in details

Please not just suggest rerunning, but explain what you observe and what you would propose as the conclusion.

57. 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

58. Explain your assessment of genome 2 in details

Please not just suggest rerunning, but explain what you observe and what you would propose as the conclusion.

59. 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

60. Explain your assessment of genome 3 in details

Please not just suggest rerunning, but explain what you observe and what you would propose as the conclusion.

61. 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

62. Explain your assessment of genome 4 in details

Please not just suggest rerunning, but explain what you observe and what you would propose as the conclusion.

63. In an outbreak situation, would you consider the EQA provided genome 5 a part of the cluster of closely related isolates?

(State one answer only)

- Yes, genome 5 is a part of the cluster
- No, genome 5 is NOT a part of the cluster

64. Explain your assessment of genome 5 in details

Please not just suggest rerunning, but explain what you observe and what you would propose as the conclusion.

65. In an outbreak situation, would you consider the EQA provided genome 6 a part of the cluster of closely related isolates?

(State one answer only)

- Yes, genome 6 is a part of the cluster
- No, genome 6 is NOT a part of the cluster

66. Explain your assessment of genome 6 in details

Please not just suggest rerunning, but explain what you observe and what you would propose as the conclusion.

67. 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 68
- No – Go to 107

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

(State one answer only)

- SNP based – Go to 70
- Allele based – Go to 77
- Other – Go to 69

69. If another analysis is used please describe your approach: - Go to 84

70. Please report the used SNP-pipeline

(reference if publicly available or in-house pipeline)

71. Please select the approach used for the SNP analysis

(State one answer only)

- Reference based – Go to 72
- Assembly based – Go to 75

72. 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)

73. Please indicate the read mapper used

(e.g. BWA, Bowtie2)

74. Please indicate the variant caller used

(e.g. SAMtools, GATK)

75. Please indicate the assembler used

(e.g. SPAdes, Velvet)

76. Please specify the variant caller used

(e.g. NUCMER)

77. Please select tool used for the allele analysis

(State one answer only)

- BioNumerics – Go to 79
- SeqSPhere – Go to 79
- Enterobase – Go to 79
- Other – Go to 78

78. If another tool is used please list here:

79. Please indicate allele calling method:

(State one answer only)

- Assembly based and mapping based - Go to 80
- Only assembly based - Go to 80
- Only mapping based - Go to 81

80. Please indicate the assembler used

(e.g. SPAdes, Velvet)

81. Please select scheme used for the allele analysis

(State one answer only)

- Applied Math (wgMLST) – Go to 83
- Applied Math (cgMLST/Enterobase) – Go to 83
- Enterobase (cgMLST) – Go to 83
- Other – Go to 82

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

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

84. Additional analysis on data derived from WGS

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

Indicate the isolate ID

86. 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		

87. 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 88
- No – Go to 107

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

(State one answer only)

- SNP based – Go to 90
- Allele based – Go to 97
- Other – Go to 89

89. If another analysis is used please describe your approach: - Go to 104

90. Please report the used SNP-pipeline

(reference if publicly available or in-house pipeline)

91. Please select the approach used for the SNP analysis

(State one answer only)

- Reference based – Go to 92
- Assembly based – Go to 95

92. 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)

93. Please indicate the read mapper used

(e.g. BWA, Bowtie2)

94. Please indicate the variant caller used

(e.g. SAMtools, GATK)

95. Please indicate the assembler used

(e.g. SPAdes, Velvet)

96. Please specify the variant caller used

(e.g. NUCMER)

97. Please select tool used for the allele analysis

(State one answer only)

- BioNumerics - Go to 99
- SeqSPhere - Go to 99
- Enterobase - Go to 99
- Other - Go to 98

98. If another tool is used please enter here:

99. Please indicate allele calling method:

(State one answer only)

- Assembly based and mapping based - Go to 100
- Only assembly based - Go to 100
- Only mapping based - Go to 101

100. Please indicate the assembler used

(e.g. SPAdes, Velvet)

101. Please select scheme used for the allele analysis

(State one answer only)

- Applied Math (wgMLST) - Go to 103
- Applied Math (cgMLST/Enterobase) - Go to 103
- Enterobase (cgMLST) - Go to 103
- Other - Go to 102

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

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

104. Third analysis on data derived from WGS

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

Indicate the isolate ID

106. 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		

107. Additional questions to the WGS part

108. Where was the sequencing performed

(State one answer only)

- In own laboratory
- Externally

109. Protocol used to prepare the library for sequencing:

(State one answer only)

- Commercial kits - Go to 110
- Non-commercial kits - Go to 112

110. Please indicate name of commercial kit:

111. If relevant please list deviation from commercial kit shortly in few bullets: - Go to 113

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

113. The sequencing platform used

(State one answer only)

- Ion Torrent PGM - Go to 115
- Ion Torrent Proton - Go to 115
- Genome Sequencer Junior System (454) - Go to 115
- Genome Sequencer FLX System (454) - Go to 115
- Genome Sequencer FLX+ System (454) - Go to 115
- PacBio RS - Go to 115
- PacBio RS II - Go to 115

- HiScanSQ - Go to 115
- HiSeq 1000 - Go to 115
- HiSeq 1500 - Go to 115
- HiSeq 2000 - Go to 115
- HiSeq 2500 - Go to 115
- HiSeq 4000 - Go to 115
- Genome Analyzer Iix - Go to 115
- MiSeq - Go to 115
- MiSeq Dx - Go to 115
- MiSeq FGx - Go to 115
- ABI SOLiD - Go to 115
- NextSeq - Go to 115
- MinION (ONT) - Go to 115
- Other - Go to 114

114. If another platform is used please list here:

115. 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 to EQA-10 scheme.

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.

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

(State one answer only)

- Yes
- No – Go to 118

117. Procedure used to evaluate confirmation of organism:

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

(State one answer only)

- Yes
- No – Go to 120

119. Procedure or threshold used for coverage:

120. Did you evaluate assembly quality?

(State one answer only)

- Yes
- No – Go to 122

121. Procedure used to evaluate assembly quality:

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

(State one answer only)

- Yes
- No – Go to 124

123. Procedure or threshold used for assembly length:

124. Did you evaluate allele calling result?

(State one answer only)

- Yes
- No – Go to 126

125. Procedure used to evaluate allele calling:

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

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

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

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

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

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

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

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

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

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

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

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

137. Comment(s):

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

138. Thank you for your participation

Thank you for filling out the Submission form for the *Salmonella* EQA-11.

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)

Address:
Gustav III:s boulevard 40, SE-169 73 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 a matter in which, directly or indirectly, they have any personal interest such as to impair their independence. Declarations of interest must be received from any prospective contractor(s) before any contract can be awarded.
www.ecdc.europa.eu/en/aboutus/transparency

HOW TO OBTAIN EU PUBLICATIONS

Free publications:

- one copy:
  via EU Bookshop (http://bookshop.europa.eu);
- more than one copy or posters/maps:
  from the European Union's representations (http://ec.europa.eu/represent_en.htm);
  from the delegations in non-EU countries (http://eeas.europa.eu/delegations/index_en.htm);
  by contacting the Europe Direct service (http://europa.eu/europedirect/index_en.htm) or calling 00 800 6 7 8 9 10 11 (freephone number from anywhere in the EU) (*).

  (*) The information given is free, as are most calls (though some operators, phone boxes or hotels may charge you).

Priced publications:

- via EU Bookshop (http://bookshop.europa.eu).

Publications Office