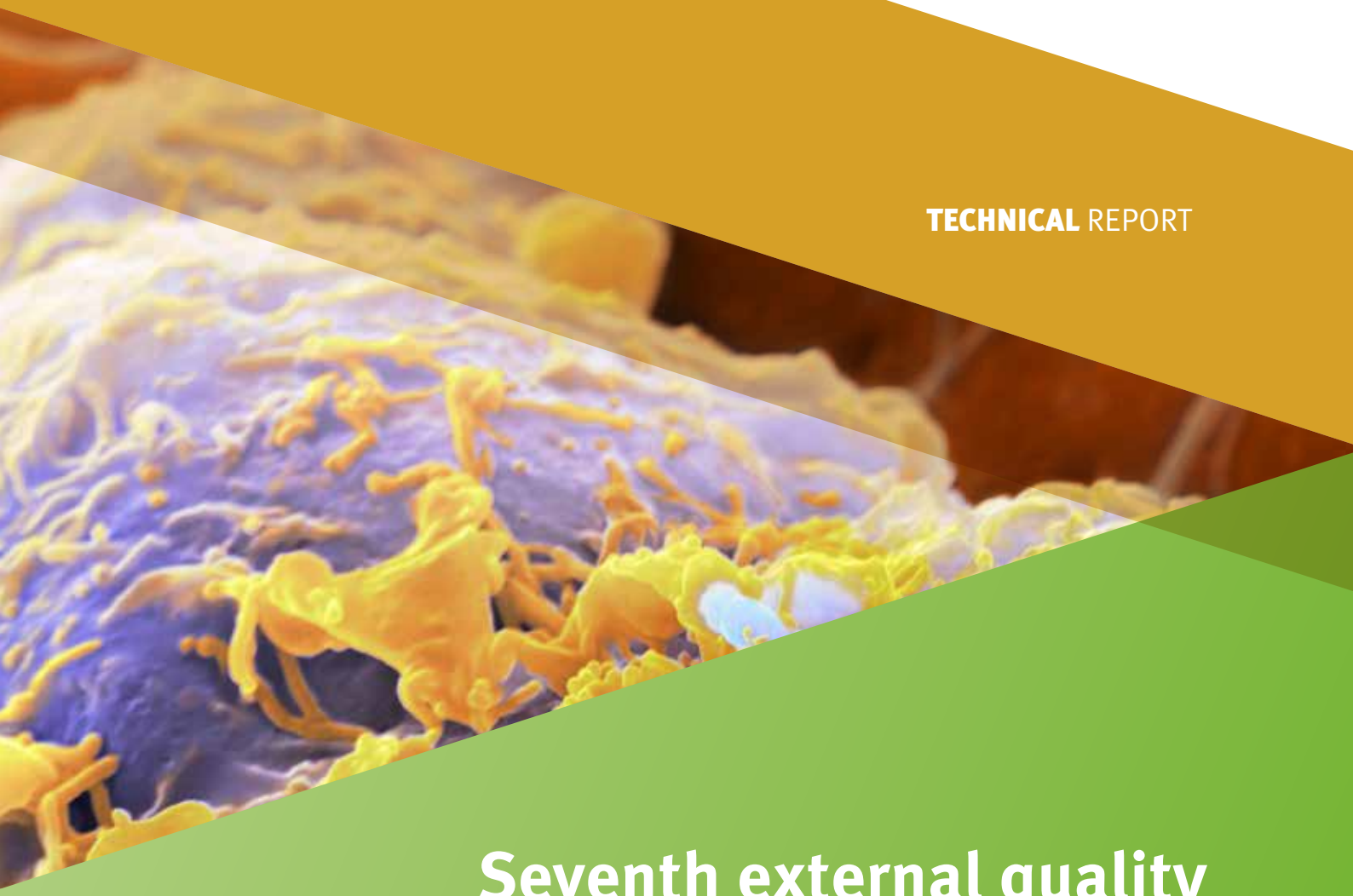
ecdc
EUROPEAN CENTRE FOR DISEASE PREVENTION AND CONTROL



TECHNICAL REPORT

Seventh external quality assessment scheme for *Listeria monocytogenes* typing

www.ecdc.europa.eu

ECDC TECHNICAL REPORT

Seventh external quality assessment scheme for *Listeria monocytogenes* 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 G. Dahl, Gitte Sørensen, Kristoffer Kiil, and Eva Møller Nielsen of the Section for Foodborne Infections at Statens Serum Institut, Copenhagen, Denmark.

Suggested citation: European Centre for Disease Prevention and Control. Seventh external quality assessment scheme for *Listeria monocytogenes* typing. Stockholm: ECDC, 2021.

Stockholm, June 2021

ISBN: 978-92-9498-540-8
DOI: 10.2900/599508
Catalogue number: TQ-02-21-741-EN-N

Contents

Figures

Tables

Annexes

Abbreviations

BN	BioNumerics
cgMLST	Core genome multilocus sequence type
EQA	External quality assessment
EU/EEA	European Union/European Economic Area
EURL	European Union Reference Laboratory
FWD	Food- and waterborne diseases and zoonoses
FWD-Net	Food- and Waterborne Diseases and Zoonoses Network
PFGE	Pulsed-field gel electrophoresis
NPHRL	National Public Health Reference Laboratory
QC	Qualitative control
SNP	Single nucleotide polymorphism
SNV	Single-nucleotide variant based on cgMLST
SSI	Statens Serum Institut
ST	Sequence type
TESSy	The European Surveillance System
wgMLST	Whole genome multilocus sequence type
WGS	Whole-genome sequencing

Executive summary

This report presents the results of the seventh round of the external quality assessment (EQA-7) scheme for *Listeria monocytogenes (L. monocytogenes)* typing organised for national public health reference laboratories (NPHRLs) providing data to the Food- and Waterborne Diseases and Zoonoses Network (FWD-Net) managed by ECDC. Since 2012, the Section for Foodborne Infections at Statens Serum Institut (SSI) in Denmark has arranged the EQA under a framework contract with ECDC. EQA-7 contains serotyping and molecular typing-based cluster analysis.

Human listeriosis is a relatively rare but serious foodborne disease with an EU notification rate of 0.47 cases per 100 000 population in 2018 [3]. The number of human listeriosis cases in the EU increased between 2008 and 2016. In 2017-2018, the level of reported cases was stable.

Since 2007, ECDC has been responsible for the EU-wide surveillance of listeriosis, including facilitating detection, and investigation of foodborne outbreaks. Surveillance data, including basic typing parameters for the isolated pathogen, are reported by European Union/European Economic Area (EU/EEA) countries to The European Surveillance System (TESSy), including molecular typing data. This molecular surveillance system relies on the capacity of laboratories to provide comparable data to FWD-Net. In order to ensure the EQA is linked to the development of surveillance methods used by NPHRLs in the EU/EEA, EQAs 5 to 7 featured a molecular typing-based cluster analysis using either pulsed-field gel electrophoresis (PFGE) and/or whole-genome sequencing (WGS)-derived data.

The objectives of the EQA are to assess the quality and comparability of typing data reported by NPHRLs participating in FWD-Net. Test isolates for the EQA were selected to cover isolates currently relevant for public health in the EU and represent a broad range of clinically relevant types for invasive listeriosis. Two sets of 11 test isolates were selected for serotyping and molecular typing-based cluster analysis. Eighteen laboratories signed up and 17 completed the exercise, representing a decrease in participation from 20 laboratories (15%) for EQA-5, but the same level of participation as for EQA-6. The majority of participants (12/17; 71%) completed the full EQA scheme.

In total, 14 (82%) participants participated in the serotyping part. Molecular serogrouping results were provided by 13 of 15 (93%) participants. Three participants performed both conventional serotyping and molecular serogrouping. The performance of molecular serogrouping was highest, with 100% correct results. For the conventional method, 75% of the participants correctly serotyped all test isolates. One new participant mistyped five of the 11 isolates. Since the first EQA in 2012, a trend towards substituting conventional serotyping with molecular serogrouping has been observed.

Of the 17 laboratories participating in EQA-7, 15 (88%) performed molecular typing-based cluster analysis using a method of their choice. The purpose 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. to correctly categorise the cluster test isolates regardless of the method used, not to follow a specific procedure.

The cluster of closely related isolates was pre-defined by the EQA provider using WGS-derived data. Therefore, as expected, the correct cluster delineation was difficult to obtain by the use of less discriminatory methods, e.g. PFGE. None of the three participants using PFGE did identify the correct cluster. Thirteen laboratories performed cluster analysis using WGS-derived data. Performance was high, with 100% of the participants correctly identifying the cluster of closely related isolates. An allele-based method was preferred since 84% (11/13) used core genome multilocus sequence type (cgMLST), compared with 16% (2/13) using single nucleotide polymorphism (SNP).

In EQA-7, the EQA provider introduced an additional part to the molecular typing-based cluster analysis: an assessment of four EQA provided genomes. This was designed to mimic an urgent outbreak situation, where sequence data may have been produced in other laboratories and the available sequences must be addressed despite, for example, possible poor quality. The majority of participants successfully identified the different characteristics in the modified genomes, and also correctly concluded one cluster isolate as being part of the cluster defined in the cluster analyses part of the EQA and one non-cluster genome as not being part of the cluster of closely related isolates.

In EQAs 5-7, participants were free to choose their preferred analytical method for the WGS-based cluster identification. The conclusion from EQA-5 was that cgMLST has higher consistency compared to SNP analysis. The conclusion was not as obvious in either EQA-6 or EQA-7, since only a few SNP analyses were reported in these schemes compared with six SNP analyses in EQA-5. One participant changed from SNP analysis to cgMLST and identified the cluster correctly, unlike in EQA-6.

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 fosters the development of sufficient capacity within the EU/EEA network for the diagnosis, detection, identification, and characterisation of infectious agents that may threaten public health. ECDC maintains and extend such cooperation and supports the implementation of quality assurance schemes [1].

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

ECDC has outsourced the organisation of EQA schemes for EU/EEA countries in the disease networks. EQAs aim to identify areas of improvement in the 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 the EQA schemes are to:

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

Since 2012, the Section for Foodborne Infections at Statens Serum Institut (SSI) in Denmark has been the EQA provider for the typing of *Salmonella enterica* ssp. *enterica*, Shiga toxin/verocytotoxin-producing *Escherichia coli* (STEC/VTEC), and *L. monocytogenes*. In 2016, SSI was also granted the new round of tenders (2017–2020) for all three pathogens. For lot 3 (*L. monocytogenes*) from 2017, the EQA scheme no longer covers assessment of the PFGE quality. However, it still covers serotyping and includes cluster analysis of *L. monocytogenes.* This report presents the results of the *Listeria* EQA-7.

1.2 Surveillance of listeriosis

Human listeriosis is a relatively rare but serious foodborne disease, with high rates of morbidity, hospitalisation, and mortality in vulnerable populations. The number of human listeriosis cases in the EU increased between 2008 and 2016. Since 2016, the number of confirmed human listeriosis cases has stabilised: from 2 509 cases in 2016, 2 480 in 2017, and 2 549 in 2018, corresponding to a notification rate of 0.47-0.48 cases per 100 000 population [3].

One of the key objectives for ECDC is to improve and harmonise the surveillance system in the EU to increase scientific knowledge of aetiology, risk factors, and burden of food- and waterborne diseases and zoonoses. Surveillance data, including some basic typing parameters for the isolated pathogen, are reported by countries to 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. Since 2012, ECDC has enhanced EU surveillance by incorporating molecular typing data into the reporting of isolate-based data for selected foodborne pathogens. Since March 2019, ECDC has been coordinating WGS-enhanced real-time surveillance of invasive listeriosis within the EU/EEA. The overall aims of integrating molecular typing 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 isolates across EU/EEA countries and contribute to global investigations;
- detect the emergence of new and/or evolving pathogenic isolates;
- support investigations to trace the source of an outbreak and identify new risk factors; and
- aid the study of a particular pathogen's characteristics and behaviour in a community of hosts.

Molecular typing-enhanced surveillance gives countries' users access to EU-wide molecular typing data for the included pathogens. It also gives users the opportunity to perform cluster searches and assess cross-country comparability of EU-level data to determine whether isolates characterised by molecular typing at the national level are part of a multinational cluster that may require cross-border response and collaboration.

1.3 Objectives

EQA schemes offer quality support for those NPHRLs that are performing molecular typing-enhanced surveillance and those that are implementing it into their surveillance system at the national level.

1.3.1 Serotyping

The EQA-7 scheme assessed serotype determination by either conventional antigen-based typing of somatic 'O' antigens and flagellar 'H' antigens or PCR-based molecular serogrouping.

1.3.2 Molecular typing-based cluster analysis

The objective of *L. monocytogenes* EQA-7 was to assess the ability to detect a cluster of closely related isolates. Laboratories could perform analysis using PFGE and/or derived data from WGS. In addition, 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 genomes were manipulated by the EQA provider. In the individual reports, this analysis was not evaluated and therefore not directly commented on, but will be summarised in this report.

2. Study design

2.1 Organisation

Listeria EQA-7 was funded by ECDC and arranged by SSI following ISO/IEC 17043:2010 [4]. EQA-7 included serotyping and molecular typing-based cluster analysis and was carried out from June-December 2019.

Invitations were emailed to ECDC contact points in FWD-Net (26 countries nominated laboratories to participate in the EQA rounds from 2017–2020) by 24 May 2019, with a deadline to respond by 2 June 2019. In addition, an invitation was sent to EU candidate country Turkey. Each laboratory was asked to fill in the reason for participating or not participating.

Eighteen NPHRLs in EU/EEA and EU enlargement countries accepted the invitation to participate, and 17 submitted results (Annex 1). In Annex 2, details of participation in EQA-6 and EQA-7 are listed to give an overview of the trend in the number of participants. The EQA test isolates were sent to participants on 19 June 2019. Participants were asked to submit their results by 9 September 2019 using the online form (Annex 14). 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 was available on the online site.

2.2 Selection of test isolates/genomes

Twenty-eight candidate isolates were analysed by the methods used in the EQA (serotyping and WGS) before and after re-culturing. All candidate isolates remained stable using these methods and the final selection of 19 test isolates; 11 test isolates and four sets of technical duplicates (same isolate culture twice), were made.

Nineteen *L. monocytogenes* test isolates were selected according to the following criteria:

- cover a broad range of the commonly reported clinically relevant isolates of invasive listeriosis in Europe;
- include genetically closely related isolates;
- remain stable during the preliminary test period at the organising laboratory;
- include three 'repeat isolates' from EQA-1 to EQA-6; and
- include three sets of technical duplicates (Annex 4).

Eleven test isolates for serotyping were selected to cover different serotypes/serogroups (1/2a/IIa, 1/2c/IIc, and 4b/IVb). Eleven test isolates for cluster analysis were selected to include isolates with different or varying relatedness and different 7-gene Multi-Locus Sequence Types (ST) (ST8 and ST120). Three of the serotyping isolates were also included in the set for cluster identification.

Table 1. Serotypes/-groups and sequence types of *L. monocytogenes* test isolates/genomes

Method	Number of test isolates	Serogroup/ST	Annex
Only Serotyping	8 isolates Sero1-Sero8	1/2a/IIa x 2 1/2c/IIc x 2 4b/IVb x 4	4
Both Serotyping and Cluster analysis	3 isolates (Sero9/REF1-Sero11/REF3)	1/2a/IIa x 2 1/2a/IIa x 1	4/7
Only Cluster analysis	8 isolates REF4-REF11	7 x ST8 and 1 x ST120	7
	4 genomes REF11*, REF8*, REF12-REF13	4 x ST8 (modified genomes: one contaminated with 15% *Listeria innocua* and one with reduced coverage)	12

Technical duplicate sets were Sero1/3, Sero2/10, Sero4/5 and REF5/REF11 (Annex 4 and Annex 7)
'Repeat isolates' included in EQA-4 to 10 (Sero1/3, Sero4/5 and Sero7)
**modified by the EQA provider.*

To follow the development of each laboratory's performance (the reproducibility), three isolates of different serotypes/-groups were included in EQAs 1 to 7: Sero1 (technical duplicate with isolate Sero3) (4b - IVb) and Sero4 (technical duplicate with isolate Sero5 (1/2c - IIc) and Sero7 (1/2a - IIa).

Based on the WGS-derived data, the cluster of closely related isolates consisted of five *L. monocytogenes* ST8 isolates (including the technical duplicate REF5/REF11). The characteristics of all the *L. monocytogenes* test isolates are listed in Annex 4–12.

2.3 Distribution of isolates/genomes

All 19 test isolates were blinded and shipped on 19 June 2019. The protocol for the EQA exercise and a letter stating the unique isolate IDs were included in the packages and distributed individually to the participants by email on 20 June 2019 as an additional precaution. Ten participants received their dispatched isolates within one day, five within two days, and three received the package five days after the shipment. The packages were shipped from SSI labelled as UN3373 Biological Substance. No participants reported damage to the shipment or errors in the unique isolates' IDs.

On 3 July 2019, instructions regarding the procedure for submitting 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 serotyping part, 11 *L. monocytogenes* isolates were tested to assess the participants' ability to obtain the correct serotype. Participants could choose to perform conventional serological methods and/or PCR-based molecular serogrouping (multiplex PCR according to the protocol suggested by Doumith et al. [5]). The serotypes/-groups were submitted in the online form.

In the cluster analysis part, participants could choose to perform the laboratory part using PFGE (*Apa*I and *Asc*I profiles) and/or WGS-derived data. The participants were instructed to report the IDs of the isolates included in the cluster of closely related isolates by method.

Laboratories performing WGS could use their own analysis pipeline for cluster analysis, e.g. single nucleotide polymorphism analysis (SNP-based) or whole genome multilocus sequence typing (wgMLST)/cgMLST (allele-based), and were asked to submit the isolates identified as a 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 analyses), but the detected cluster had to be based on results from the main analysis. Laboratories reported SNP distance or allelic differences between a selected cluster isolate and each test isolate, and uploaded the raw reads (FASTQ files).

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

2.5 Data analysis

The submitted serotyping and cluster analysis results, as well as the raw reads, were imported to a dedicated *Listeria* EQA-7 BioNumerics (BN) database. The EQA provider accepted one participant submitting data one month after the deadline.

Serotyping results were evaluated according to the percentage of correct results, generating a score from 0-100%. Molecular typing-based cluster analysis was evaluated according to correct or incorrect identification of the expected cluster of closely related isolates, based on a pre-defined categorisation by the organiser based on the WGS-derived data. Cluster analysis based on WGS-derived data was derived on allele-based analysis (cgMLST [6] and SNP analysis (NASP, [7]). The correct number of closely related *L. monocytogenes* isolates by WGS were five ST8 isolates: REF1, REF5, REF8, REF9 and REF11 (REF5 and REF11 were technical duplicates). The EQA provider found at most four allele differences or eight SNPs between any two isolates in the cluster. The remaining six of the cluster test isolates were additional four ST8s and two ST120.

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

Individual evaluation reports were distributed to participants in December 2019 and certificates of attendance in January 2021. If WGS-derived 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.

3. Results

3.1 Participation

Laboratories could participate either in the full EQA scheme or in one part only (serotyping or molecular typing-based cluster analysis). Of the 18 participants who signed up, 17 completed and submitted their results. The majority of the participants (71%, 12/17) completed the full EQA scheme. In total, 14 (82%) participants participated in serotyping and 15 (88%) in cluster analysis. Conventional serotyping results were provided by 29% (4/14) of the participants, and three of these laboratories also performed molecular serogrouping. Molecular serogrouping results were provided by 13 (76%) participants. Of the 14 participants who completed the serotyping/serogrouping part, all reported the reason for participating: laboratory policy to enhance the typing quality in a combination of accreditation needs, institute policy, and/or national policy (Annex 3).

Most participants (87%: 13/15) reported cluster analysis using WGS-derived data, while two (13%) reported using PFGE data only. One participant (7%) submitted cluster data based on both PFGE and WGS (Table 1). Of the 13 laboratories that completed the cluster analysis using WGS, 11 reported the reason for participating: laboratory policy to enhance the typing quality in a combination of accreditation needs, institute policy and/or national policy. The last two reported accreditation needs and institute policy and the other only institute policy. (Annex 3).

Table 1. Number and percentage of laboratories submitting results for each method

	Serotyping				Cluster analysis			
	Conventional only	Molecular only	Both	Total	PFGE only	WGS only	Both	Total
Number of participants	1	10	3	14	2	12	1	15
Percentage of participants	7	71	21	82*	13	80	6	88*

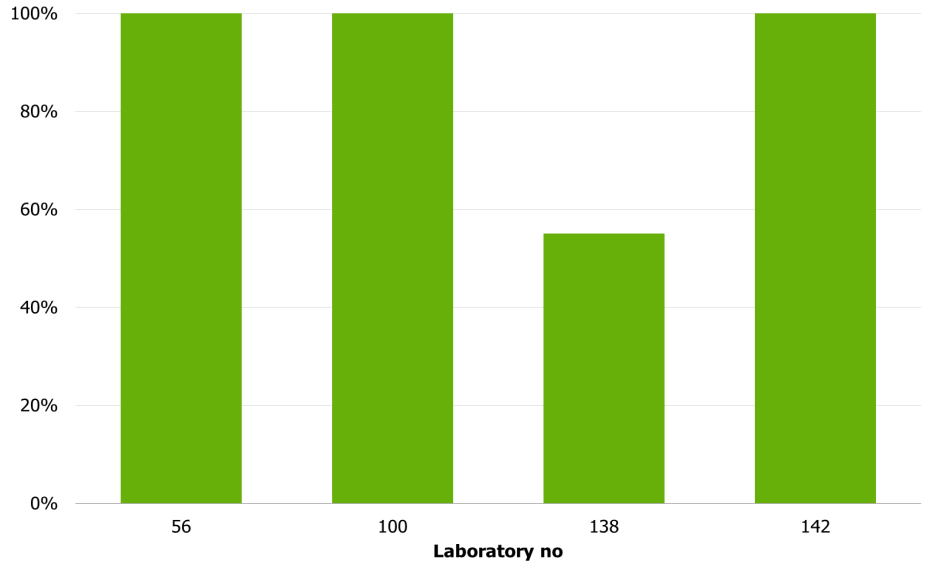
Twelve of the 17 participants (71%) completed both parts (serotyping and cluster analysis) of the EQA.
**: percentage of total number of participating laboratories (17).*

3.3 Serotyping

3.3.1 Conventional serotyping

Four participants performed conventional serotyping of *L. monocytogenes* (Figure 1). Performance was high (90%) as three of the participants correctly serotyped all 11 test isolates. One laboratory (138) had issues with multiple isolates, and only correctly serotyped seven of the 11 isolates. Among the three sets of technical duplicated isolates, laboratory 138 only reported the correct serotype of one isolate in each set (Annex 4).

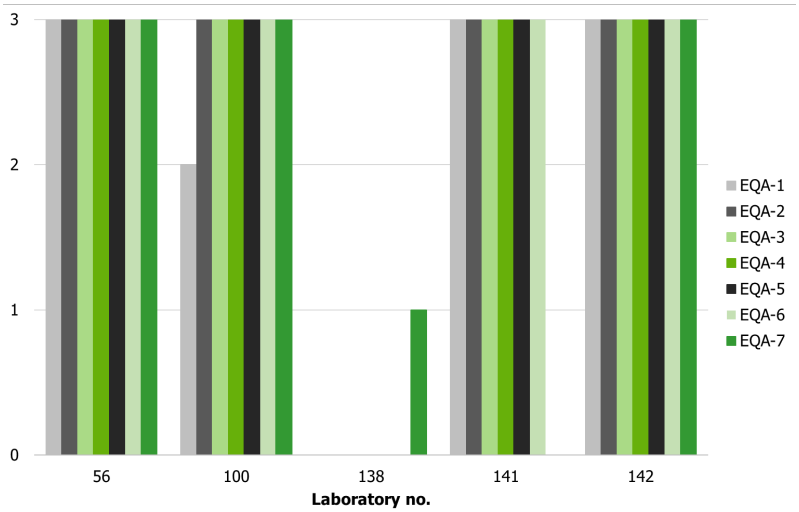
Figure 1. Participant scores for conventional serotyping of 11 test isolates



Arbitrary numbers represent participating laboratories. Bars represent the percentage of correctly assigned serotypes for the 11 test isolates (Sero1–11).

Figure 2 shows the reproducibility of the individual participants' performances on conventional serotyping of the three 'repeat isolates' from EQA-1 to EQA-7 (including some of the technical duplicates). Only laboratories participating in EQA-6 and/or EQA-7 are shown. The reproducibility of conventional serotyping results of the repeat isolates, show stability and high performance among four of the participants participating every year (laboratory 56, 100, 141 and 142). However, laboratory 138, participating for the first time did not report correct serotyping results for the repeat isolates (reporting different results for technical duplicates).

Figure 2. Correct conventional serotyping of three repeat isolates through EQA-1 to 7 (for laboratories participating in EQA-6 and/or EQA-7)

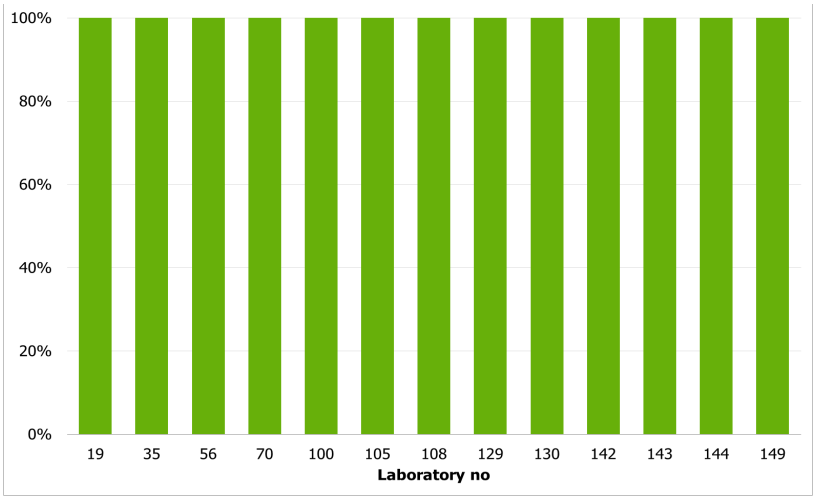


Arbitrary numbers represent the participating laboratories. Bars represent the number of correctly assigned serotypes for the three repeat isolates (Sero1/3, 4/5 and 7).

3.3.2 Molecular serogrouping

Thirteen participants performed molecular serogrouping of *L. monocytogenes* in EQA-7 (Figure 3). In EQA-5, two new laboratories (96 and 130) participated in molecular serogrouping, but in EQA-6 and EQA-7, only laboratory 130 continued the participation. Molecular serogrouping was carried out in accordance with guidelines of Doumith et al. [5] and nomenclature from Doumith et al. [8] was used. All 13 (100%) participants were able to correctly serogroup all 11 EQA test isolates. Three of the 13 participants reported using WGS-based analysis (*in silico* PCR) for molecular serogrouping.

Figure 3. Participant scores for molecular serogrouping of 11 *L. monocytogenes* test isolates

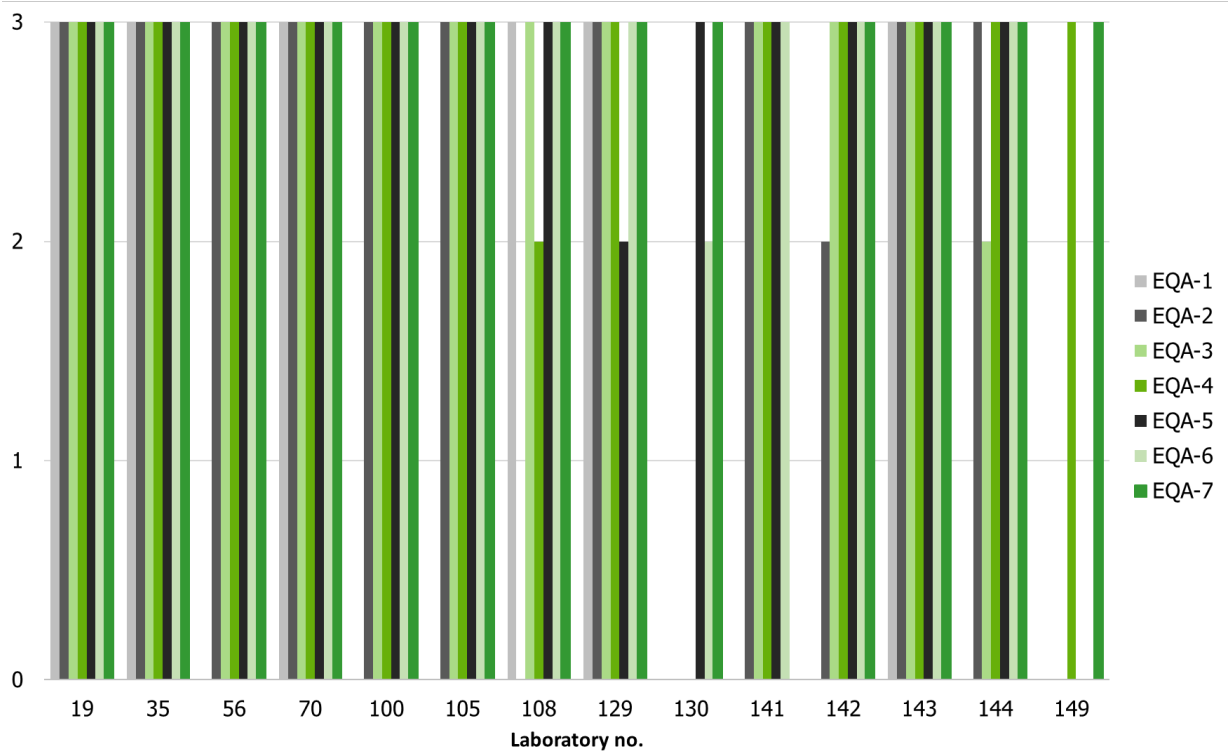


Arbitrary numbers represent the participating laboratories. Bars represent the percentage of correctly assigned serogroups for the 11 test isolates Sero1–11.

Figure 4 shows the individual reproducibility of participants' performances on molecular serogrouping when assessing the three repeat isolates during the seven EQAs. Out of the 14 laboratories that either participated in EQA-6 and/or EQA-7, 64% (9/14) have correctly serogrouped all three repeat isolates when participating during the

seven EQAs. All the laboratories with errors improved their performance and, correctly serogrouped the three reproduced isolates in the following EQA participations.

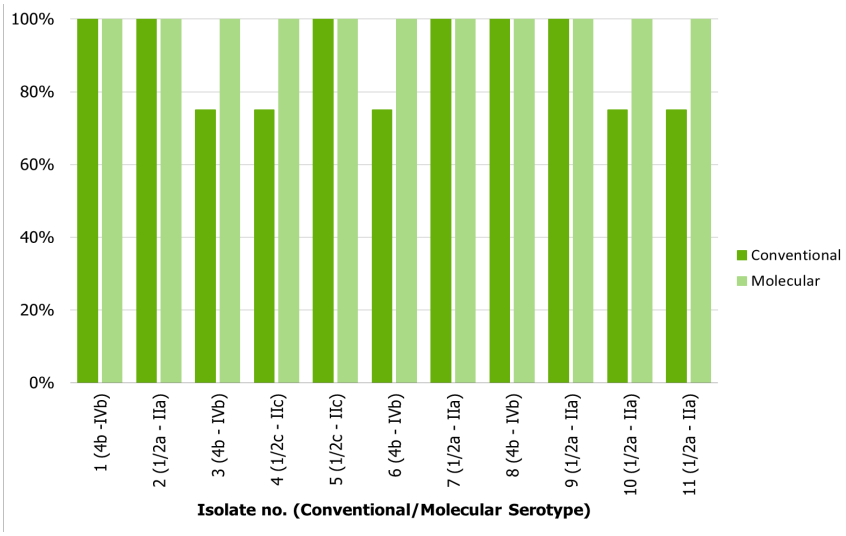
Figure 4. Correct molecular serogrouping of three repeat isolates from EQA-1 to 7 (for laboratories participating in EQA-6 and/or EQA-7)



Arbitrary numbers represent the participating laboratories. Bars represent the number of correctly assigned serogroups for the three repeat isolates (Sero1/3, 4/5 and 7).

Figure 5 show the reported error distributed per isolate. No isolate had more than one laboratory reporting an error, and all errors in this EQA-7 were reported by laboratory 138 using conventional serotyping.

Figure 5. Average score of 11 test isolates



Bars represent the percentage of correctly assigned serotypes/-groups by the participants.

3.4 Molecular typing-based cluster analysis

Participants were tested on their ability to correctly identify the cluster of closely related isolates defined by pre-categorisation from the EQA provider among the 11 cluster test isolates using either PFGE and/or WGS-derived data. The cluster of five test isolates were pre-categorised by the EQA provider. The expected cluster of closely related *L. monocytogenes* ST8 isolates contained four isolates based on WGS-derived data. The EQA provider's cluster analysis of WGS-derived data was based on an allele-based (cgMLST [6]) and SNP analysis (NASP [7]).

The correct cluster based on WGS-derived data contained five ST8 isolates: REF1, REF5, REF8, REF9, and REF11 (REF5 and REF11 were technical duplicates). The EQA provider found at most four allele differences or eight SNPs between any two isolates in the cluster. The rest of the cluster test isolates were additional four ST8s and two ST120. (Annexes 5, 9).

3.4.1 PFGE-derived data

Three (3/17, 18%) participants performed cluster analysis using PFGE-derived data. The cluster categorisation was based on WGS data and therefore as expected the correct cluster delineation was difficult to obtain by the use of a less discriminatory method. All three participants included the five WGS cluster isolates as a part of the correct cluster, however in addition, laboratory 138 and 144 also included REF3, REF6 and REF10 in the cluster, and laboratory 142 added REF6 (ST120).

Table 2 shows the overview of the isolates that each participant included or excluded in their cluster identification.

Table 2. Results of cluster analysis based on PFGE-derived data

		Laboratory ID		
Isolate number	**ST**	**138**	**142**	**144**
REF1‡	ST8	Yes	Yes	Yes
REF2	ST120	No	No	No
REF3	ST8	Yes	No	Yes
REF4	ST8	No	No	No
REF5#‡	ST8	Yes	Yes	Yes
REF6	ST120	Yes	Yes	Yes
REF7	ST8	No	No	No
REF8‡	ST8	Yes	Yes	Yes
REF9‡	ST8	Yes	Yes	Yes
REF10	ST8	Yes	No	Yes
REF11#‡	ST8	Yes	Yes	Yes
Correct cluster identified*		**No**	**No**	**No**
Included the five WGS cluster isolates		**Yes**	**Yes**	**Yes**

‡: closely related isolates predefined by WGS (in grey).
#: technical duplicate isolates (in bold).
**pre-defined categorisation by WGS derived data.*
Annex 9.

3.4.2 WGS-derived data

Reported results from participants

Thirteen participants (13/17, 77%) 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, one HiSeq, two NextSeq, one NovaSeq, one Ion GeneStudio S5 System, and one Ion Torrent. All reported using commercial kits for library preparation. Of the 13 participants, nine (69%) used Illumina's Nextera kit. One participant reported volume changes from the manufactory protocol and one laboratory listed less time for shearing and volume changes from the manufactory protocol (Annex 8).

Performance was high in cluster analysis with WGS-derived data. All 13 participants (100%) correctly identified the cluster of closely related isolates defined by pre-categorisation from the EQA provider among the 11 test isolates (Table 3).

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

		Laboratory ID												
Isolate number	**ST**	**19**	**35**	**56**	**70**	**100**	**105**	**108**	**129**	**135**	**141**	**142**	**146**	**149**
REF1‡	ST8	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
REF2	ST120	No	No	No	No	No	No	No	No	No	No	No	No	No
REF3	ST8	No	No	No	No	No	No	No	No	No	No	No	No	No
REF4	ST8	No	No	No	No	No	No	No	No	No	No	No	No	No
REF5#‡	ST8	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
REF6	ST120	No	No	No	No	No	No	No	No	No	No	No	No	No
REF7	ST8	No	No	No	No	No	No	No	No	No	No	No	No	No
REF8‡	ST8	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
REF9‡	ST8	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
REF10	ST8	No	No	No	No	No	No	No	No	No	No	No	No	No
REF11#‡	ST8	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Main analysis		Allele	Allele	Allele	Allele	Allele	Allele	SNP	Allele	Allele	Allele	Allele	SNP	Allele
Additional analysis 1		SNP				SNV								
Additional analysis 2						SNP								
Cluster- identified		**Yes**	**Yes**	**Yes**	**Yes**	**Yes**	**Yes**	**Yes**	**Yes**	**Yes**	**Yes**	**Yes**	**Yes**	**Yes**

‡: closely related isolates (in grey).
#: technical duplicate isolates (in bold).
**: 'cgMLST' assigned by provider based on reported information (Table 5).*
ST: 7 multilocus sequence type.
Allele: allele-based analysis (cgMLST).
SNV: single-nucleotide variant based on cgMLST.
SNP: single-nucleotide polymorphism (Annex 10).

Laboratories were instructed to report the data analysis used for cluster identification 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 analysis. Laboratories could report results from up to three analyses (one main and one to two additional), but the detected cluster had to be based on results from the main analysis. Only two participants reported additional analysis (laboratories 19 and 100).

Of the 13 participants, two (15%) used SNP as the main analysis, one used an in-house pipeline, and one a published pipeline. Both used a reference-based approach with different isolates as reference. One laboratory used CLC for both read mapper and variant caller, and the other used Burrows-Wheeler Aligner (BWA) as the read mapper and a different variant caller, Genome Analysis Toolkit (GATK; Tables 4–5).

Table 4. Results of SNP-based cluster analysis

Laboratory	**SNP**						
	SNP pipeline	**Approach**	**Reference**	**Read mapper**	**Variant caller**	**Distance within cluster**	**Distance outside cluster¤**
Provider	NASP [8]	Reference-based	REF11 (ST8)	BWA	GATK	0-5	26-868 (983-1008)
Provider	NASP [8]	Reference-based	REF11 (ST8)	BWA	GATK	0-5	26-1007 (not analysed)
Provider	NASP [8] + recombination filter	Reference-based	REF11 (ST8)	BWA	GATK	0-5	26-74 (not analysed)
19*	NASP	Reference-based	REF8 (ST8)	BWA	GATK	0-6	24-75 (not reported)
100*	BioNumerics	Reference-based	Listeria ST121 HE999704.1	Bowtie 2	-	0-5	23-82 (114-181)
108	In-house	Reference based	In-house strain resp ST	CLC assembly cell v.4.4.2	CLC assembly cell v.4.4.2	0-6	26-1059 (not reported)
146	Phenix [a]	Reference based	CP006862	CLC assembly cell v.4.4.2	CLC assembly cell v.4.4.2	0-6	1073-1648 (3075-2608)

¤: reported distance to ST8 isolates (non-ST8; Annex 11)
**: additional analysis*
a: (https://github.com/phe-bioinformatics/PHEnix) and SnapperDB (https://github.com/phe-bioinformatics/snapperdb).

Eleven of the 13 participants used allele-based analysis selected as the main analysis for cluster detection (Table 5). Seven (54%) used only an assembly-based allele calling method, three used both mapping- and assembly-based allele calling, and one used only mapping-based allele calling method (Table 5). All 11 reported using cgMLST, seven (64%) used cgMLST Ruppitsch (1701 loci) [10], three cgMLST Pasteur (1 748 loci), and one an in-house cgMLST scheme with only 1503 loci.

Table 5. Results of allele-based cluster analysis

Laboratory	Allele-based analysis						
	Approach	Allelic calling method	Assembler	Scheme	Number of loci	Difference within cluster	Difference outside cluster¤
Provider	BioNumerics	Assembly- and mapping-based	SPAdes	Applied Math (cgMLST/Pasteur)	1748	0–2	14–43 (60-64)
19	BioNumerics	Assembly-based and mapping-based	SPAdes	Applied Math (cgMLST/Pasteur)	1748	0–2	14–43 (60-64)
35	SeqPhere	Only assembly-based	Velvet	Ruppitsch (cgMLST)	1701	0–4	14–47 (59-67)
56	Other	Only mapping-based MentaLiST	-	Pasteur (cgMLST)	1748	0–4	10-35 (49-66)
70	SeqPhere	Assembly-based and mapping-based	Velvet	Ruppitsch (cgMLST)	1701	0–4	14-47 (60-69)
100	SeqPhere	Only assembly-based	Velvet	Ruppitsch (cgMLST)	1701	0–4	14-55 (53-106)
100*	SNV	-	-	SNVs were calculated based on cgMLST scheme	‘1701’	0-4	14-55 (71-112)
105	SeqPhere	Only assembly-based	SPAdes (Unicycler 0.4.6 (SPAdes v3.12.0 + pilon v1.22)	Ruppitsch (cgMLST)	1701	0–4	14-47 (60-69)
129	SeqPhere	Only assembly-based	Velvet	The Ridom SeqSphere+ software’s Target Definer, cgMLST	1503	0–3	15-36 (51-55)
135	SeqPhere	Only assembly-based	SPAdes	Ruppitsch (cgMLST)	1701	0–4	14-55 (67-106)
141	SeqPhere	Only assembly-based	SPAdes 3.11.1	Ruppitsch (cgMLST)	1701	0–4	14-47 (61-70)
142	BioNumerics	Assembly-and mapping-based	SPAdes	Pasteur (cgMLST)	1748	0–2	14-44 (61-65)
149	SeqPhere	Only assembly-based	Velvet	Ruppitsch (cgMLST)	1701	0–4	14-47 (61-70)

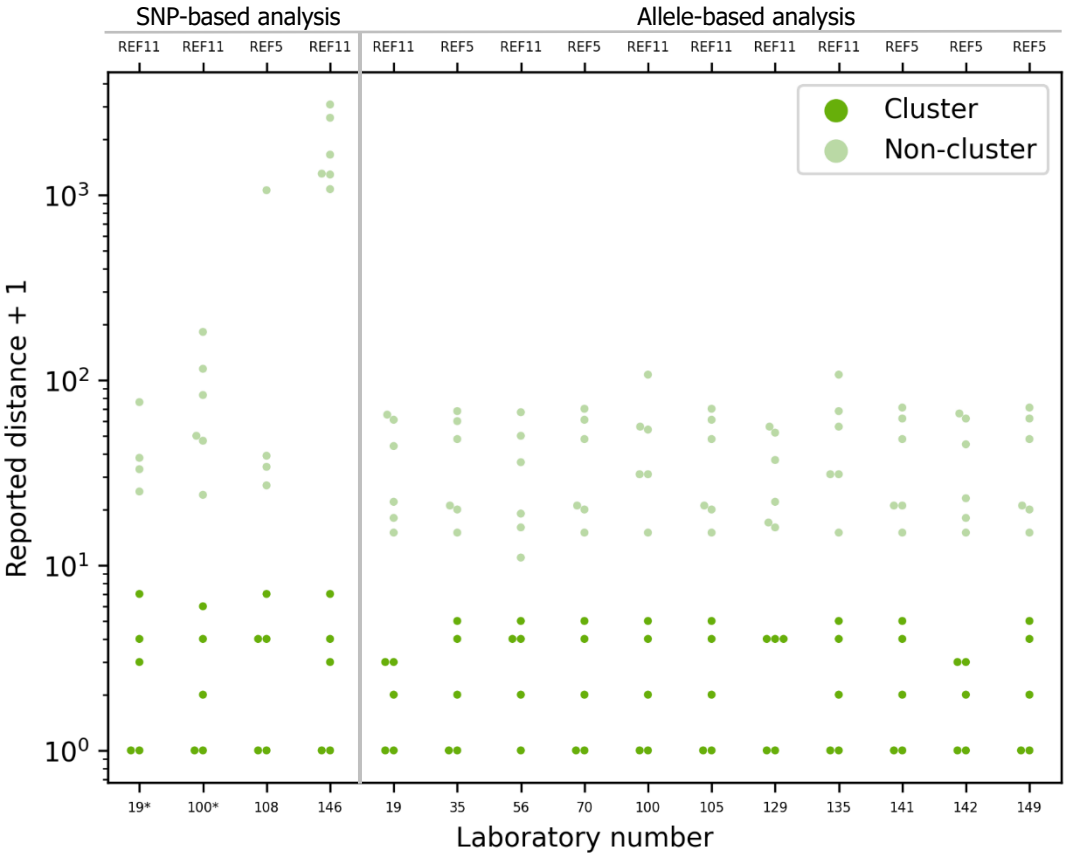
¤: reported distance to ST8 isolates (non-ST8)
SNV: single nucleotide variant
‘1701’ added by the EQA provider, based on the comments of using the cgMLST scheme.
Annex 11.

All 11 laboratories reported up to four allele differences within the correct cluster (Table 5/Figure 6). Two laboratories reported only 0-2 allele difference, one 0-3 and the eight laboratories reported 0-4 allelic difference inside the cluster. All laboratories selected one of the technical duplicates as the cluster representative in the main analysis (six laboratories selected REF5 and seven used REF11 (Figure 6)).

Four of the test isolates (REF3, REF4, REF7 and REF10) were also ST8, but not predefined by the EQA provider as part of the cluster. Based on cgMLST, the 11 laboratories reported allele differences to the selected cluster isolate at 10-55 for this group of isolates. Two test isolates (REF2 and REF6) were ST120 and allele differences to the selected cluster isolate at 49-106 were reported. The only SNV analysis reported 71-112 allele differences. (Annex 11).

Laboratories 19, 56, and 142 used the same cgMLST scheme as the EQA provider (cgMLST/Pasteur) [6], and all the others except one laboratory (129) used the Ruppitsch scheme [10] with 1701 loci.

Figure 6. Reported SNP distances or allelic differences for each test isolate to selected cluster representative isolate



SNP: single nucleotide polymorphism.
Selected cluster representative marked as REF.
Dark green: reported cluster of closely related isolates.
Light green: not reported as part of cluster.
The additional analysis of laboratory 100 using SNV analysis was not included in the figure.

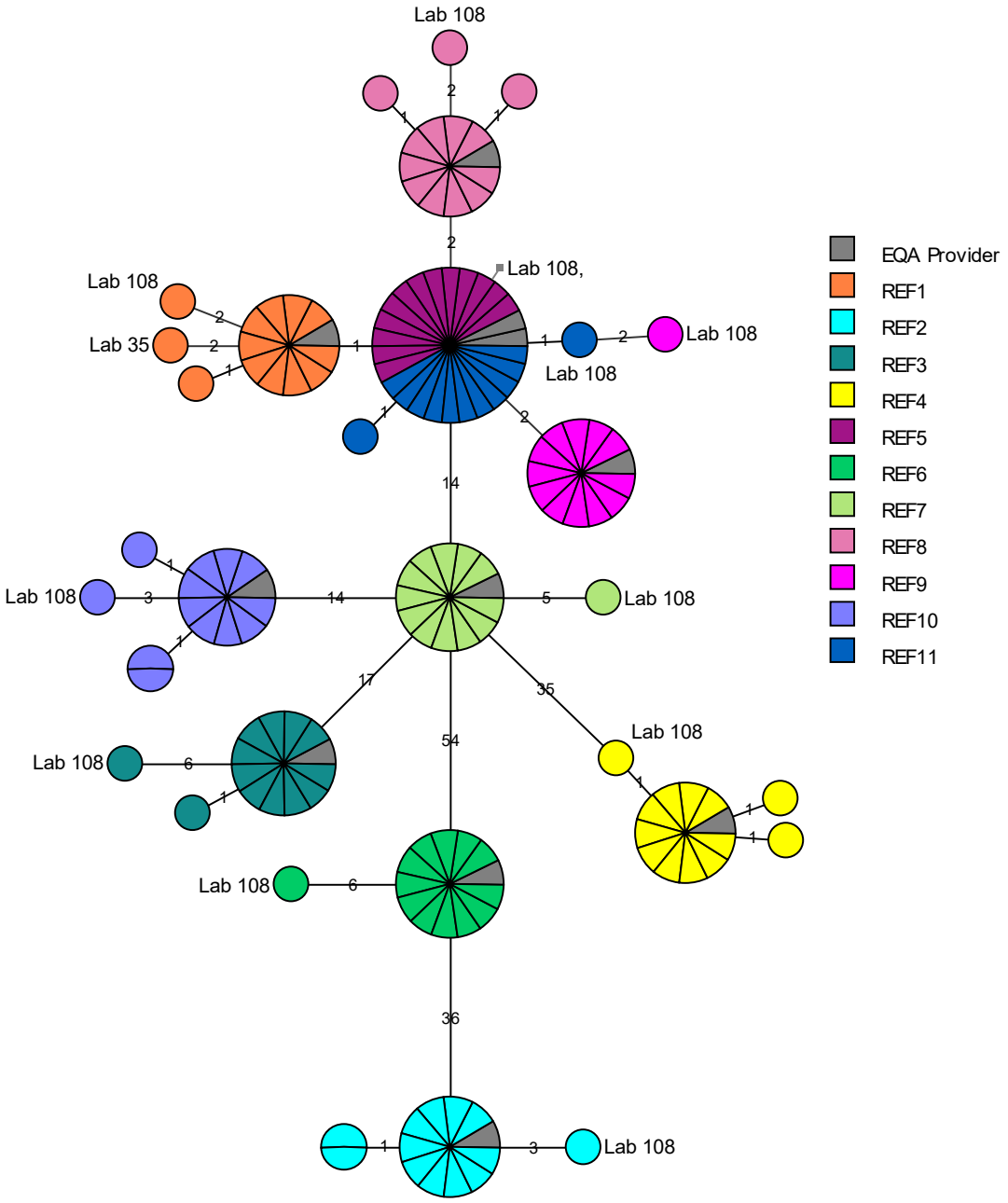
The reported SNP differences within the correct cluster (Figure 6) were between 0-6 for both the laboratories using SNP as the main analysis and the laboratories using SNP as the additional analysis, despite not including all isolates and using different reference sequences. A clear separation between cluster and non-cluster isolates was obtained by all laboratories, however the distance reported by laboratory 146 was much higher than the one reported by other laboratories. When the EQA provider analysed all ST8 sequences by SNP analysis and hereafter filtered the recombination [9], the results were similar to both the additional analysis of laboratory 19 and 100, despite the laboratories did not inform of any removal of recombination areas.

Analysis of raw reads uploaded by participants

In addition to the reported cluster identification, participants submitted their FASTQ files to be evaluated by the EQA provider. The FASTQ files were uploaded to an Applied Maths calculation engine for allele calling (Institut Pasteur) [6] and evaluated by the EQA provider's in-house quality control (QC) pipeline [11].

The overall cgMLST analysis, shown in the minimum spanning tree (MST) based on submitted raw reads (FASTQ files) from 13 laboratories reveals clear clustering of the results for each test isolate (Figure 7).

Figure 7. Minimum spanning tree of core genome multilocus sequence typing participant FASTQ files

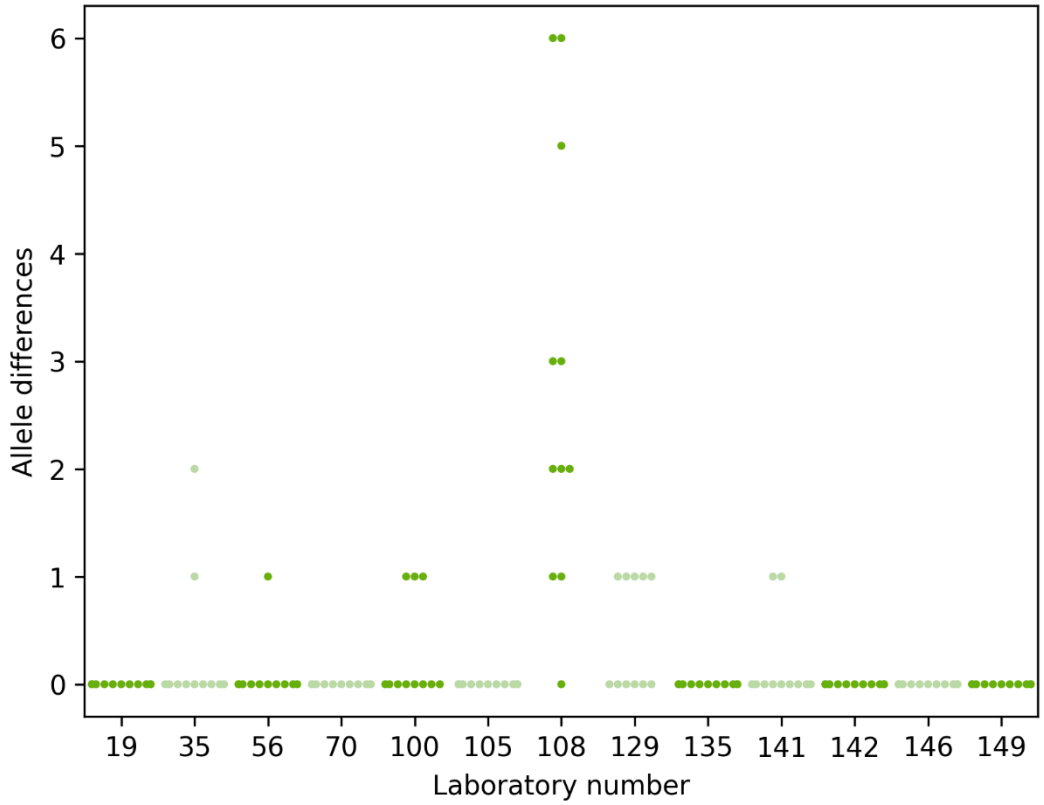


Minimum spanning tree (MST) in log scale of core genome multi-locus sequence typing (cgMLST) [6] based on submitted raw reads (FASTQ files). Each of the REF1–11 test isolates have a different colour. REF results from the EQA provider are in grey.

The allele differences in Figure 7 do not exactly match those illustrated in the individual reports and consequently those in Figure 8, as all are based on the same data. This discrepancy is caused by loci being dropped if they do not pass QC in all isolates in the analysis. Joint analysis thus contains fewer loci.

For each laboratory's sequences, cgMLST was performed on the submitted raw reads (FASTQ files), applying Applied Maths allele calling with the Pasteur scheme [6]. For each laboratory, a hierarchical single linkage clustering was performed on the submitted data along with the EQA provider's reference isolates. Figure 8 shows the allele differences between each submitted sequence and the corresponding reference.

Figure 8. Participant allele difference from reference result (EQA provider) for each test isolate



Allele difference of participant isolates from corresponding REF isolates (EQA provider) based on submitted raw reads (FASTQ files).

For 120 of 143 results (84%), no difference was identified. For 14 results (10%), a difference of one allele from the REF isolate was calculated and a difference of 2-10 alleles was seen for nine results (6%) in Figure 8. Results from laboratory 108 showed allele difference for 10 of 11 isolates, eight isolates with a difference of 2-6 alleles.

Separately, the laboratories responded to QC parameters used to evaluate their data. Both coverage and confirmation of the genus were the most widely used QC parameters, with 85% of the laboratories using these parameters (Table 6). Participants used different thresholds of coverage, ranging from 20-50 x coverage. Different programs used for the contamination check of the genus were reported. The number of good cgMLST loci was used as a QC parameter by 77% of the participants, with thresholds ranging from 89-99%. Q score and genomic size were used by 46% and 69% of the participants respectively. A few laboratories reported additional parameters (Annex 12). For the full QC evaluation of all isolates, see Annex 13.

Table 6. Summary of selected QC parameters reported by participants

Laboratory	Confirmation of genus	Coverage	Q score (Phred)	Genome size	Number of good cgMLST loci
19	Kraken and <5% contamination with others	Min. x25	No	2,8-3,1Mb	Min. 99% core% and max. 15 loci with multiple consensus. But no actual threshold are employed on regular basis for either.
35	No	No	No	No	90% good cgMLST loci
56	No mismatches in the alignment with the seven housekeeping genes of MLST panel	50	22	2,8 Mb	No
70	A specific task in SeqSphere targets the prs gene	50x but if it's less, the number of targets found should be >95%	No	length of contigs assembled < ref genome + 10%	cgMLST alleles found and called >95%
100	KmerFinder 3,1[a]	40x	90	app. 2,9 Mb	0.98
105	Assembled genomes were aligned against a *Listeria monocytogenes* genome (threshold: >90% nucleotide identity)	depth of coverage >45x	Trimming was performed with trimmomatic, removing 3 nucleotides with Phred <10 or an average Phred <15 in a sliding window of 4 nucleotides. Sequences with a length <70 bases were removing too	<=3.3Mb	>=95%
108	Assembled genomes aligned against *L. monocytogenes*	>20x	No	+/- 20%	No
129	presence of prfA gene (LIP)	>29	No	No	>89
135	No	>30	30	2.8-4Mb	>10%
141	JSpecies	30x	No	~2.8-3.0 Mb	min. 98% good targets
142	Kraken	>30x	>28	>3.3Mb	0,95
146	KmerID software[b]	>30 average coverage across the genome	Quality score >30 following trimming of reads	No	No
149	KRAKEN	No	No	No	> 95% good targets
% of laboratories using QC parameter	**85%**	**85%**	**46%**	**69%**	**77%**

a: Center for Genomic Epidemiology
b: https://github.com/phe-bioinformatics/kmerid

For each laboratory, the submitted raw reads (FASTQ files) were evaluated by the EQA provider's in-house quality control pipeline [11]. Table 7 shows the QC parameters and range of QC values per laboratory. For the full QC evaluation of all isolates, see Annex 13.

According to the QC parameters, sequencing quality was uniformly good. Overall, coverage was high.

Table 7. Results of raw reads submitted by participants evaluated by EQA provider QC pipeline summarised by laboratory

		Laboratory ID												
Parameters	**Ranges***	**19**	**35¤**	**56**	**70**	**100**	**105**	**108**	**129**	**135**	**141**	**142**	**146**	**149**
Detected species	{*Lm*}	*Lm*	*Lm*	*Lm*	*Lm*	*Lm*	*Lm*	*Lm*	*Lm*	*Lm*	*Lm*	*Lm*	*Lm*	*Lm*
% Species 1		91.9-95.5	98.1-98.6	97.0-98.7	97.8-98.1	90.4-99.3	98.1-98.7	98.7-99.0	96.2-98.6	94.3-97.5	95.5-97.4	97.5-98.9	94.3-95.3	88.4-95.8
% Species 2	{<5%}	0.1-0.2	0.0	0.0-0.1	0.1-0.2	0.0-3.4	0.0-0.1	0.0-0.1	0.0	0.0-1.3	0.0-0.1	0.0-0.3	0.0-0.1	1.3-7.2
Unclassified reads (%)		4.3-7.8	1.3-1.8	1.3-3.0	1.6-1.9	0.6-4.3	1.3-1.8	1.0-1.2	1.3-3.7	2.4-4.5	2.4-4.3	1.0-2.2	4.6-5.7	1.6-4.5
Length at >25 x min. coverage (Mbp)	{>28 ∧ <31}	2.9-3.0	2.9-3.0	2.4-2.9	2.9-3.0	2.9-3.1	2.9-3.1	2.5-2.6	2.9-3.1	2.9-3.1	2.9-3.1	2.9-3.1	2.9-3.1	2.9-3.1
Length [1-25] x min. coverage (kbp)	{<250}	0.0-2.8	0.0-77.1	0.0	0.0-86.6	0.0-84.0	0.0	0.0-9.3	0.0-1.6	0.0-8.0	0.0-0.5	0.0-7.4	0.0-27.6	0.0-0.7
No. of contigs at 25 x min. coverage	{>0}	27-67	29-118	143-3222 #	14-21	16-36	15-20	3215-3789 #	16-58	19-82	19-61	15-22	14-51	16-31
No. of contigs [1-25] x min. coverage	{<1000}	0-4	0-62	0 #	0-1	0-113	0	0-39 #	0-2	0-9	0-1	0-5	0-21	0-2
Average coverage	{>50}	118-204	47-230	154-524	32-60	48-114	169-448	55-99	140-264	248-339	76-113	50-114	99-200	49-196
No. of reads (x 1000)		2548-4436	956-4574	1515-6023	663-1216	613-1490	3464-8827	546-1000	2786-5231	5166-7459	937-1371	637-1443	3028-6063	590-2537
Average read length		131-145	145-148	264-318	151-151	232-244	147-148	290-307	147-149	132-148	230-256	230-245	98-100	245-285
Average insert size		179-267	NA	NA #	300-336	318-386	324-408	NA#	315-383	252-400	258-315	353-439	203-318	288-467
N50 (kbp)		76-342	45-425	1 -41#	304-439	418-580	426-580	1 #	100-460	72-425	114-525	425-580	99-439	298-580

**: indicative QC range*
Lm: L. monocytogenes
NA: not available
#: QC values unreliable due to assembly issues for Ion Torrent data
¤: single end reads (Annex 13).

Assessment of the EQA provided genomes

All participants were invited to perform an additional assessment of the cluster analysis. Four provided genomes were asked to be checked for possible relatedness to the cluster, thus mimicking an urgent outbreak situation where rerunning the sequence is not possible and the sequences must be assessed despite poor quality, etc.

The four provided genomes should be individually assessed and compared with the already produced data in the cluster analysis. The participants had to determine whether or not 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 the raw reads: the four genomes were a mix of cluster isolates with contamination, reduced coverage or as an assembly (Fasta file), and a non-cluster isolate with good quality (Table 8).

For genome 1 (good quality but contaminated), 92% (12/13) of laboratories correctly described the contamination present in genome 1, two even detected the added species *L. innocua.* Due to the observed contamination most of the laboratories did not proceed to an analysis of genome 1. The two laboratories that performed the analysis, suggested the isolate to be a part of the already identified cluster. One laboratory did not give any comments to genome 1.

For genome 2, (low coverage), 92% (12/13) of laboratories correctly observed poor quality in genome 2, seven used the low average coverage to disregard the genome, and other laboratories used different QC parameters. Therefore, none of the 12 laboratories suggested the isolate to be a cluster isolate, but only one laboratory listed details of the analysis and laboratory 108 did not describe any observation of a low coverage, but suggested Genome 2 was a different ST (ST1419).

For genome 3, (good quality), 85%% (11/13) of laboratories were able to use the Fasta file in their analysis, and 10/11 correctly identified genome 3 as part of the cluster of closely related isolates. One laboratory (35) had misunderstood the question about the cluster analysis and two laboratories could not use the Fasta file in their analysis.

For genome 4, (good quality), 100% (all 13 laboratories) correctly identified genome 4 as a non-cluster isolate.

Table 8. Results of the participants' assessment of the EQA provided genomes

Genome	Characteristics	Characteristics identified by participants	Yes	No	Not analysed
1	A cluster isolate (REF5/REF11) mixed with a *Listeria innocua* (approx. 15%)	Contamination observed	12	0	1
2	A cluster isolate (REF8) with altered coverage (reduced to 10x)	Poor quality observed	12	1	0
3	A cluster isolate (REF12) good quality of reads assembled with SKESA to a FASTA file. Zero allelic difference to the REF5 in the cluster	Suggested to be a cluster isolate	10	1	2
4	A non-cluster isolate (REF13), good quality of reads. 28 allelic difference to the cluster isolate (REF5)	Suggested to be a cluster isolate	0	13	0

Annex 12.

4. Discussion

4.1 Serotyping/-grouping

Fourteen (82%) laboratories participated in the serotyping part of EQA-7 and 13 (93%) provided molecular serogrouping results.

4.1.1 Conventional serotyping

The number of participants decreased from 10 laboratories in EQA-1 to four in EQA-6 and EQA-7, highlighting the transition towards the use of molecular serogrouping. Comparing the conventional serotyping results from EQA-1 to 7, the results showed stable and high performances among participants (EQA-1 to 7: 94%; 87%; 91%; 97%; 98%; 100%; 89%).

4.1.2 Molecular serogrouping

Since EQA-2, the number of participants in the molecular serogrouping has ranged between 13 and17 participants. In EQA-6 and EQA-7 the number of participants were the same (N=13). In both EQA-6 and EQA-7, only three laboratories reported the use of *in silico* PCR (WGS) serogrouping, but seven of the remaining participants who used the conventional PCR did participate in the cluster analysis. This suggests that the participants are able to use WGS, but chose not to, likely due to cost, as some participants have suggested reducing the number of isolates for the serotyping/-grouping part. The three remaining participants who also used conventional PCR did not participate in the cluster analysis. The performance of the laboratories with regard to molecular serogrouping was very high in EQA-7, with a score of 100% correct. The general performance among the participating laboratories has been high over the years from EQA-1 to 7: 98%; 94%; 94%; 94%; 99%; 97%; 100%.

The switch from the conventional serotyping to molecular serogrouping has reached a level where the molecular serogrouping can be seen as the best practise in the NPHRLs in the EU/EEA.

4.2 Molecular typing-based cluster analysis

In EQA-5 to EQA-7, PFGE was no longer an independent part, but was added as a possible method of choice for cluster identification. This EQA-7 is contemporary with the development of surveillance methods used by NPHRLs in the EU/EEA. This adjustment of the EQA appears to be well accepted by the countries, as 15 of the 17 laboratories (88%) participated in the cluster analysis. Only two laboratories participated in cluster identification using PFGE as a sole method, while one laboratory participated in the cluster identification using both PFGE- and WGS-derived data.

4.2.1 PFGE-derived data

Of the 15 laboratories participating in the cluster analysis, three (20%) performed cluster analysis using PFGE-derived data. As the criteria of the pre-defined cluster was based on WGS-derived data, the correct cluster delineation was difficult to obtain by the use of less discriminatory method. None of the participants only identified the five cluster isolates (defined by WGS), but included one or more isolates. Laboratory 138 and 144 included the same additional isolates (REF3, REF6 and REF10) in the cluster, whereas laboratory 142 only included REF6 (ST120). Laboratory 142 also performed cluster analysis on WGS-derived data (with correct result). The number of participants only submitting cluster analysis based on PFGE-derived data has decreased from three (EQA-5 and EQA-6) to two in EQA-7, while 76% (13/17) has submitted analysis based on WGS-derived data.

4.2.3 WGS-derived data

Thirteen of the 17 laboratories (76%) performed cluster analysis using WGS-derived data. In EQA-5, 12 laboratories participated in WGS-based cluster analysis. In EQA-6, one laboratory stopped and one new started. In EQA-7, all participants from EQA-6 continued and one new started. Performance was very high, with all laboratories (100%) correctly identifying the cluster of closely related isolates. As in EQA-5 and EQA-6, only one laboratory reported the use of external assistance for sequencing and the majority (11/13) also reported using an Illumina platform. All reported using commercial kits for library preparation.

The newest participant performed allelic-based analysis as the main analysis, but provided also a SNP analysis and an additional cgMLST analysis (SNV). One participant changed their main analysis from SNP in EQA-6 to allele analysis in EQA-7. The rest of the participants (11 laboratories) used the same main analyses as in the previous year (two SNP and nine allele). The two laboratories using SNP analysis reported distances (0-6 SNPs) comparable to those reported using allele-based methods (0-4 ADs). Regardless of sequencing method, analysis method and even scheme, all participants achieved very similar results. One of the two laboratories using Ion Torrent data,

changed their analysis from SNP (EQA-6) to cgMLST (EQA-7). In EQA-7, they achieved a clear separation of the isolates (cluster isolates and non-cluster isolates), as opposed to the results in EQA-6 when using their SNP based method on Ion Torrent data.

In general, differences measured outside the cluster were 10-55 alleles, despite not being analysed using the same scheme or assembly tool. In the SNP analysis, more variations were seen outside the cluster (See Table 5). No clear explanation could be identified for the variable results, but the laboratories used different reference sequences, different analysing tools, and some were selective about which isolates to include in the analysis and the EQA provider suspect some of the laboratories removed recombination areas, before submitting the results.

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

In order to compare the quality of the raw data, the EQA provider analysed the submitted raw reads to obtain selected QC parameters. Using single-end sequencing on the Illumina platform, laboratory 35 reduced the run time without apparent negative consequences for any QC parameters or the cluster analysis.

The EQA provider's analysis of the submitted raw data showed that when using a standardised cgMLST analysis, a very high concordance was obtained (Figure 8). The maximum differences between any two cluster isolates were four alleles for most of the participants and five for a single participant (all using Illumina data, data not shown). This is, however, very similar to the distances reported by the participants. For Ion Torrent data, the maximum allelic difference between any two cluster isolates was five for laboratory 56 and seven for laboratory 108 when data were analysed by the EQA provider (data not shown). In the EQA provider's analysis of the participants' sequences, the majority of the results that were two allele differences or more were caused by laboratory 108 (Figure 8). The EQA provider's analysis is not optimised for Ion Torrent data, making correct assembly difficult, but the analysis performed well on the Ion Torrent data from laboratory 56 with one allele difference for one isolate. Thus, the observed allele differences may be method artefacts, but the use of Ion Torrent data can complicate the communication and investigation of multi country outbreaks if only an allele-based method is used.

From the EQA data, the number of NPHRLs that have access to WGS capacity, at least for the EQA test isolates, is slowly increasing, and the level of laboratories only performing PFGE is decreasing. This suggests that the WGS can be seen as the preferred method of surveillance in NPHRLs in the EU/EEA.

In the assessment of the additional EQA provided genomes, most of the participants successfully identified the quality deficiencies in the modified genomes, both the contaminated genome and the one with low coverage. In addition, the participants also correctly concluded the cluster isolate (Fasta file) as being part of the cluster and the non-cluster genome as not being part of the cluster of closely related isolates. Unfortunately, most of the participants, when observing the contamination or the low coverage, did not proceed in assessing if the data were of any use at all. Most responded that they would need to rerun the sequence, and therefore said no to the question 'Is this genome a part of the cluster?'. The wording of the questions in the online form was apparently not formed in such a way that the participants understood that the EQA provider wished they would process with an analysis despite the observed obstacles.

5. Conclusions

Seventeen laboratories participated in the EQA-7 scheme, with 14 (82%) performing serotyping and 15 (88%) cluster identification.

Most laboratories (71%, 10/14) performed only molecular serogrouping and 21% (three) molecular serogrouping in combination with the conventional serotyping, and only 7% (1) with only conventional serotyping. In general, the trend towards substituting conventional serotyping with molecular was observed through the seven EQAs. The average quality of conventional serotyping (89%) was in the range of the previous EQAs. The performance of molecular serogrouping were highest, with 100% in EQA-7. The general conclusion is that serotyping performance is high with both methods, but slightly higher for molecular serogrouping. The switch from the conventional serotyping to molecular serogrouping has reached a level where the molecular serogrouping can be seen as the best practise in NPHRLs in the EU/EEA.

Three laboratories participated using PFGE for cluster analysis of which two solely used PFGE-derived data for analysis. As the cluster pre-categorisation was based on WGS data, it was expected that the correct cluster delineation was difficult to obtain by the use of less discriminatory methods, e.g. PFGE. All three participants included more isolates in the cluster than was expected by WGS.

Thirteen laboratories performed cluster analysis using WGS-derived data. The performance was very high, with all participants correctly identifying the cluster of closely related isolates.

In the WGS, an allele-based method was preferred, as 85% (11/13) used cgMLST compared to 15% (2/13) using SNP as the main reported for cluster analysis. In general, the reported cgMLST results were at a comparable level of allelic difference (0–4) within the cluster isolates despite analysis with different schemes. Slightly higher variation but similar results were obtained by the two laboratories using SNP analysis that reported the correct cluster (0–6 SNP distance inside the cluster). Both the SNP and the allele method provided a good differentiation between the cluster and non-cluster isolates; therefore, both methods seem to work for cluster detection. Nevertheless, standardised cgMLST analyses leave little room for error, resulting in good inter-laboratory comparability.

Most of the participants were able to identify the different characteristics (and modifications) of the EQA-provided genomes. For the two genomes with good quality, most participants were able to correctly conclude if they were part of the cluster. For the genomes with contamination or low coverage, most participants did not proceed with cluster analysis and therefore could not assess if the data would suggest the genome was a part of the cluster or not.

The current EQA scheme for *L. monocytogenes* typing is the seventh 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 sequences of good quality and comparable typing results for cross-border cluster detections. From March 2019, ECDC launched the possibility to submit WGS variables for *L. monocytogenes* to TESSy to be used for EU-wide surveillance and cross-sectoral comparison. Public Health Institutes can submit *L. monocytogenes* sequences directly or DNA samples for WGS performed by ECDC's contractor. Since the start, 16 EU/EEA countries have submitted sequences from 840 concurrent isolates either as part of the routine real-time submissions or as part of investigations of foodborne events. By June 2020, 28 cgMLST-based clusters had been identified and one of the clusters was escalated to an urgent inquiry.

6. Recommendations

6.1 Laboratories

Laboratories are recommended to use 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.

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

6.2 ECDC and FWD-Net

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

ECDC will search for the possibility of carrying out a validation study of the most used sequencing technologies in NPHRLs.

6.3 EQA provider

The evaluation of the provided genome sequences was a success, almost all the participant performed the analysis and identified the modifications introduced by the EQA provider. For the following EQA round, the EQA provider will expand this part of the EQA and rephrase the aim of the part and put emphasise on the importance of assessing despite a contamination – however of course concluded with the upmost caution.

References

1. European Parliament, European Council. 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. Off J Eur U. 2004 Apr 30;L142:1-11. Available at: http://ecdc.europa.eu/en/aboutus/Key%20Documents/0404_KD_Regulation_establishing_ECDC.pdf
2. European Parliament, European Council. Decision No 1082/2013/EU of the European Parliament and the Council of 22 October 2013 on serious cross-border threats to health and repealing Decision No 2119/98/EC (Text with EEA relevance). Off JEur U. 2013 Nov 5;L293:1-15. Available at: http://ec.europa.eu/health/preparedness_response/docs/decision_serious_crossborder_threats_22102013_en.pdf
3. European Food Safety Authority, European Centre for Disease Prevention and Control. The European Union One Health 2018 Zoonoses Report. EFSA J. 19 November 2019. Available at: https://efsa.onlinelibrary.wiley.com/doi/pdf/10.2903/j.efsa.2019.5926
4. International Organization for Standardization. ISO/IEC 17043:2010 – Conformity assessment -- General requirements for proficiency testing. Geneva: ISO; 2015. Available at: http://www.iso.org/iso/catalogue_detail.htm?csnumber=29366
5. Doumith M, Buchrieser C, Glaser P, Jacquet C, Martin P. Differentiation of the Major *Listeria monocytogenes* Serovars by Multiplex PCR. J Clin Microbiol. 2004 Aug;42(8):3819-22.
6. Moura A, Criscuolo A, Pouseele H, Maury MM, Leclercq A, Tarr C, et al. Whole genome-based population biology and epidemiological surveillance of *Listeria monocytogenes*. Nat Microbiol. 2016 Oct 10;2:16185. Available at: http://www.nature.com/articles/nmicrobiol2016185
7. Doumith M, Jacquet C, Gerner-Smidt P, Graves LM, Loncarevic S, Mathisen T, et al. Multicenter validation of a multiplex PCR assay for differentiating the major *Listeria monocytogenes* serovars 1/2a, 1/2b, 1/2c, and 4b: toward an international standard. J Food Prot. 2005 Dec;68(12):2648-50.
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. Available at: http://www.microbiologyresearch.org/content/journal/mgen/10.1099/mgen.0.000074
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. Ruppitsch W, Pietzka A, Prior K, Bletz S, Fernandez HL, Allerberger F, et al. Defining and Evaluating a Core Genome Multilocus Sequence Typing Scheme for Whole-Genome Sequence-Based Typing of *Listeria monocytogenes*. J Clin Microbiol. 2015 Sep;53(9):2869-76.
11. Statens Serum Institut. SerumQC "Bifrost" [Internet, software package]. Copenhagen: SSI; 2020 [cited 24 January 2020]. Available at: https://github.com/ssi-dk/bifrost

Annex 1. List of participants

Country	Laboratory	National institute
Austria	National Research Laboratory *Listeria* Austria	AGES – Austrian Agency for Health and Food Safety
Belgium	National Reference Centre *Listeria* Belgium	Sciensano
Denmark	Diagnostic and Typing of Gastrointestinal Bacteria	Statens Serum Institut
Finland	Expert Microbiology	National Institute for Health and Welfare
France	National Reference Centre and WHO Collaborating Centre for *Listeria*	Institut Pasteur
Germany	Binational German-Austrian Consiliary Laboratory for *Listeria*	Robert Koch Institute
Greece	National Reference Centre for *Salmonella* and Other Enteropathogenes	National School of Public Health
Hungary	Department of Phage-Typing and Molecular Epidemiology	National Public Health Institute
Italy	Department of Food Safety, Nutrition and Veterinary Public Health	Istituto Superiore di Sanità
Latvia	National Microbiology Reference laboratory	Infectology Centre of Latvia
Luxembourg	Epidemiology and Microbial Genomics	Laboratoire National de Sante
The Netherlands	Centre for Infectious Research, Diagnostics and Screening	National Institute for Public Health and the Environment
Norway	National Reference Laboratory for Enteropathogenic Bacteria	Norwegian Institute of Public Health
Slovenia	Department for Public Health Microbiology	National Laboratory of Health, Environment and Food
Spain	*Neisseria*, *Listeria* and *Bordetella* Unit	National Centre for Microbiology, Instituto de Salud Carlos III
Sweden	Microbiology	Folkhälsomyndigheten
United Kingdom	Gastrointestinal Bacteria Reference Unit	Public Health England

Annex 2. Participation overview EQA-6 and 7

	2018 to 2019 (EQA-6)					2019 to 2020 (EQA-7)				
	All[#]	Serotyping		Cluster		All[#]	Serotyping		Cluster	
Laboratory		Conventional	Molecular	PFGE	WGS		Conventional	Molecular	PFGE	WGS
19	X		X	X	X	X		X		X
35	X		X		X	X		X		X
56	X	X	X	X	X	X	x	X		X
70	X		X		X	X		X		X
100	X	X	X			X	x	X		X
105	X		X		X	X		X		X
108	X		X		X	X		X		X
129	X		X		X	X		X		X
130	X		X	X		X		X		
135*	X				X	X				X
138	X			X		X	X		X	
141	X	X	X	X	X	X				X
142	X	X	X	X	X	X	X	X	X	X
143	X		X			X		X		
144	X		X	X		X		X	X	
146	X				X	X				X
149	X				X	X		X		X
Number of participants	**17**	4	13	7	12	**17**	4	13	3	13

[#]: participating in at least one part.
**: previously laboratory 77.*

Annex 3. Reason(s) for participating in EQA

	Serotyping				Cluster			
Lab ID	**Accreditation needs**	**Institute policy**	**National policy**	**Enhance typing quality**	**Accreditation needs**	**Institute policy**	**National policy**	**Enhanced typing quality**
19				X				X
35		X				X		
56				X				X
70				X				X
100		X	X	X		X	X	X
105	X			X	X			X
108	X	X			X	X		
129	X			X	X			X
130[a]	X	X	X	X	X	X	X	X
135	*Serotyping is not relevant to our laboratory					X		X
138			X				X	
141	*Serotyping is not relevant to our laboratory. We only perform WGS.				X	X		X
142	X			X	X			X
143	X			X	*Not covered by accreditation			
144[b]	X			X	X			X
146	*Inferred from NGS				X		X	X
149	X	X	X	X	X	X	X	X
Number of participants	**8**	**4**	**4**	**11**	**9**	**7**	**5**	**13**

Laboratory 114: repeatedly stated that participation in this EQA is important for Accreditation needs. However not submitted any results in EQA-9 or EQA-7.
X: selected as reason for participating
**: reasons given when not participating.*
a: WGS is not yet established in our Institute.
b: We are already using WGS for typing Listeria isolates, but at the moment our Institution is under reorganisation and some reagents are missing.

Annex 4. Serotyping result scores

Conventional serotyping

		Laboratory ID			
Isolate number	Provider	56	100	138	142
Sero1[#1]	4b	4b	4b	4b	4b
Sero2[#2]	1/2a	1/2a	1/2a	1/2a	1/2a
Sero3[#1]	4b	4b	4b	4e	4b
Sero4[#3]	1/2c	1/2c	1/2c	3a	1/2c
Sero5[#3]	1/2c	1/2c	1/2c	1/2c	1/2c
Sero6	4b	4b	4b	4c	4b
Sero7	1/2a	1/2a	1/2a	1/2a	1/2a
Sero8	4b	4b	4b	4b	4b
Sero9	1/2a	1/2a	1/2a	1/2a	1/2a
Sero10[#2]	1/2a	1/2a	1/2a	1/2c	1/2a
Sero11	1/2a	1/2a	1/2a	1/2b	1/2a

Molecular serogrouping

		Laboratory ID												
Isolate	Provide	19	35	56	70	10	105	108	129	130	142	143	144	149
Sero1[#1]	IVb	IVb	IVb	IVb	IVb	IVb	IVb	IVb	IVb	IVb	IVb	IVb	IVb	IVb
Sero2[#2]	IIa	IIa	IIa	IIa	IIa	IIa	IIa	IIa	IIa	IIa	IIa	IIa	IIa	IIa
Sero3[#1]	IVb	IVb	IVb	IVb	IVb	IVb	IVb	IVb	IVb	IVb	IVb	IVb	IVb	IVb
Sero4[#3]	IIc	IIc	IIc	IIc	IIc	IIc	IIc	IIc	IIc	IIc	IIc	IIc	IIc	IIc
Sero5[#3]	IIc	IIc	IIc	IIc	IIc	IIc	IIc	IIc	IIc	IIc	IIc	IIc	IIc	IIc
Sero6	IVb	IVb	IVb	IVb	IVb	IVb	IVb	IVb	IVb	IVb	IVb	IVb	IVb	IVb
Sero7	IIa	IIa	IIa	IIa	IIa	IIa	IIa	IIa	IIa	IIa	IIa	IIa	IIa	IIa
Sero8	IVb	IVb	IVb	IVb	IVb	IVb	IVb	IVb	IVb	IVb	IVb	IVb	IVb	IVb
Sero9	IIa	IIa	IIa	IIa	IIa	IIa	IIa	IIa	IIa	IIa	IIa	IIa	IIa	IIa
Sero10[#2]	IIa	IIa	IIa	IIa	IIa	IIa	IIa	IIa	IIa	IIa	IIa	IIa	IIa	IIa
Sero11	IIa	IIa	IIa	IIa	IIa	IIa	IIa	IIa	IIa	IIa	IIa	IIa	IIa	IIa
		A	A	A	A	A	A	B	B	A	A	A	A	B

A: PCR-based serotyping, B: WGS-based serotyping
Purple: repeat isolates in EQA-1 to 7
#: set of technical duplicates 1, 2 and 3.
Pink shading: incorrect results

Annex 5. Reported cluster of closely related isolates based on PFGE-derived data

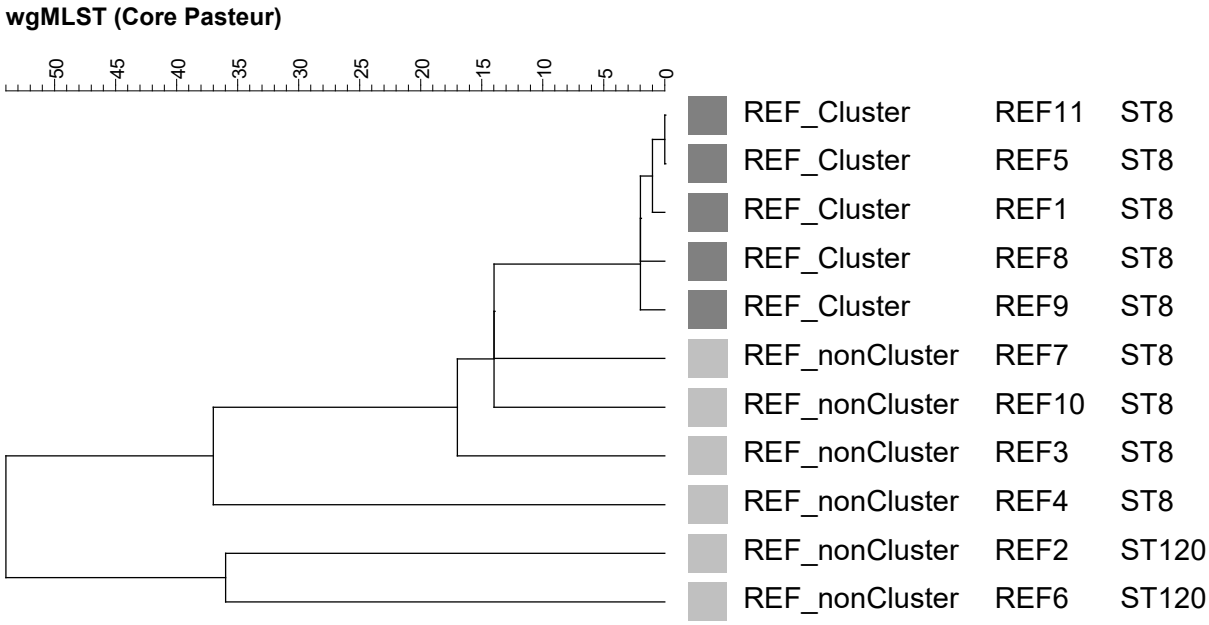
Lab ID	Reported cluster	Corresponding REF isolates	Correct
Provider		REF1, REF5, REF8, REF9, and REF11 (5 and 11 technical duplicates)	
138	1044, 1493, 1549, 2118, 2144, 2415, 1168, 1325	REF5, REF9, REF11, REF8, REF6, REF10, REF3, REF1	No
142	1141, 1238, 1275, 1500, 2065, 2391	REF5, REF11, REF9, REF1, REF6, REF8	No
144	1706, 1922, 1973, 2119, 2400, 2562, 1055, 1742	REF11, REF5, REF9, REF10, REF6, REF8, REF1, REF3	No

Annex 6. Reported sequencing details

Sequencing performed	Protocol (library prep)	Commercial kit	Sequencing platform
In own laboratory	Commercial kits	Nextera XT	MiSeq
In own laboratory	Commercial kits	NEBNext® Fast DNA Fragmentation & Library Prep Set for Ion Torrent® New England Biolabs **	Ion GeneStudio S5 System
In own laboratory	Commercial kits	MiSeq reagent kit v2	MiSeq
In own laboratory	Commercial kits	Nextera XT	HiSeq 2500
In own laboratory	Commercial kits	Nextera XT DNA Library Prep Kit (Illumina)	MiSeq
In own laboratory	Commercial kits	Nextera	MiSeq
In own laboratory	Commercial kits	Ion Xpress TM Plus Fragment Library Kit for AB Library Builder TM System	IonTorrent S5XL
Externally	Commercial kits	Nextera xt	Illumina NovaSeq
In own laboratory	Commercial kits	Nextera XT DNA Library Kit, Illumina	MiSeq
In own laboratory	Commercial kits	Nextera XT	NextSeq
In own laboratory	Commercial kits	Kapa HyperPlus (Kapa Biosystems)	MiSeq
In own laboratory	Commercial kits	nextera DNA flex library prep (illumina)	NextSeq
In own laboratory	Commercial kits	Nextera Flex Illumina *	MiniSeq

** We use half the volume of reagents for each step of the protocol*
***The shearing was carried out for 15 minutes at 25 degrees instead than from 20 minutes. The reason is that we used a 400bp sequencing protocol. The reaction was performed in half of the volume suggested by the manufacturer's instructions, starting from 100 ng of DNA.*

Annex 7. EQA provider cluster analysis based on WGS-derived data



Single linked dendrogram of core genome multi-locus sequence typing (cgMLST) profiles of Listeria *EQA-7 isolates (cgMLST, Pasteur, Moura et al 2016).*
Analysed in BioNumerics: maximum distance of 200 exceeded; results clipped.
Dark grey: cluster isolates
Light grey: outside cluster isolates.

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

Lab ID	Reported cluster	Corresponding to REF isolates	Correct
Provider		REF1, REF5, REF8, REF9 and REF11 (5 and 11 technical duplicates)	
19	1490, 1463, 1189, 2327, 1948	REF11, REF5, REF1, REF8, REF9	Yes
35	1206, 1330, 1424, 1466, 2978	REF1, REF9, REF5, REF11, REF8	Yes
56	1315, 1961, 1645, 2422, 1687	REF11, REF5, REF9, REF8, REF1	Yes
70	1690, 1672, 2716, 1685, 1794	REF5, REF11, REF8, REF1, REF9	Yes
100	1065, 1698, 1365, 2086, 1656	REF11, REF5, REF9, REF8, REF1	Yes
105	1112, 1643, 1030, 1635, 2788	REF11, REF5, REF9, REF1, REF8	Yes
108	1504, 1995, 1943, 1406, 2207	REF5, REF11, REF9, REF1, REF8	Yes
129	1387, 1215, 1347, 1744, 2450	REF1, REF11, REF5, REF9, REF8	Yes
135	2446, 1760, 1172, 1799, 1005	REF8, REF9, REF11, REF5, REF1	Yes
141	1309, 1395, 1418, 1649, 2677	REF5, REF9, REF11, REF1, REF8	Yes
142	1275, 2391, 1238, 1141, 1500	REF9, REF8, REF11, REF5, REF1	Yes
146	1630, 1763, 1778, 2050, 1356	REF11, REF5, REF9, REF8, REF1	Yes
149	1987, 1003, 1372, 1938, 2582	REF1, REF5, REF11, REF9, REF8	Yes

Annex 9. Reported SNP distance and allelic differences

SNP distances

					Laboratory ID			
Isolate number	ST	Provider	Provider	Provider + Recombination filter	19*	100*	108	146
REF1‡	ST8	3	3	3	3	3	3	3
REF2	ST120	983	9999	9999	9999	181	9999	2608
REF3	ST8	33	33	33	32	46	33	1287
REF4	ST8	868	1007	74	75	82	1059	1073
REF5#‡	ST8	0	0	0	0	0	0	0
REF6	ST120	1008	9999	9999	9999	114	9999	3075
REF7	ST8	26	26	26	24	23	26	1301
REF8‡	ST8	5	5	5	6	5	6	6
REF9‡	ST8	2	2	2	2	1	3	2
REF10	ST8	41	38	38	37	49	38	1648
REF11#‡	ST8	0	0	0	0¤	0	0	0

Allelic distances

			Laboratory ID											
Isolates number	ST	Provider	19	35	56	70	100	100*[a]	105	129	135	141	142	149
REF1‡	ST8	1	1	3	3	3	3	3	3	3	3	3	1	3
REF2	ST120	64	64	67	66	69	106	112	69	55	106	70	65	70
REF3	ST8	21	21	20	10	20	30	30	20	21	30	20	22	20
REF4	ST8	43	43	47	35	47	55	55	47	36	55	47	44	47
REF5#‡	ST8	0	0	0	1	0	0	0	0	0	0	0	0	0
REF6	ST120	60	60	59	49	60	53	71	60	51	67	61	61	61
REF7	ST8	14	14	14	15	14	14	14	14	16	14	14	14	14
REF8‡	ST8	2	2	4	3	4	4	4	4	3	4	4	2	4
REF9‡	ST8	2	2	1	4	1	1	1	1	3	1	1	2	1
REF10	ST8	17	17	19	18	19	30	30	19	15	30	20	17	19
REF11#‡	ST8	0¤	0¤	0	0	0	0	3	0	0	0	0	0	0

**: additional analysis*
‡: closely related isolates
#: technical duplicate isolate
a: SNVs were calculated based on cgMLST scheme
¤: isolate used as cluster representative by participant
9999: isolates not included in analysis by participant
ST: sequence type

Annex 10. Additional reported QC parameters

QC		Laboratory ID				
		19	56	100	141	146
1	Parameter	N50 - Available from QC analysis but no threshold	MLST	N50	No. of contigs	Contamination
	Threshold	-	no mismatches	400000	200 bases (contigs shorter than 200 bases have to be ignored)	Check using Kraken software (https://github.com/phe-bioinformatics/kmerid)
2	Parameter	Number of contigs - Available from QC analysis but no threshold		contig count		Read depth
	Threshold	-		less than 35 contigs		Require >10 reads per base
3	Parameter	Number of unidentified bases (N) - Available from QC analysis but no threshold		contamination		Mapping quality
	Threshold	-		BWA Mapping with Salmonella, 0 contigs assembled		>Q30 following mapping to reference
4	Parameter			SAV		MLST allele mapping
	Threshold			cluster density, clusters passing filter and Q30 score were all according to Illumina recommendations		Check the % of non-consensus bases of each MLST allele - more than 10% indicates a mixture/ contamination
5	Parameter					Mean consensus depth to call SNPs for each MLST allele
	Threshold					>90% consensus to call a SNP in MLST alleles

Annex 11. Calculated qualitative/quantitative parameters

		Laboratory 19										
Quali-/Quantitative	Ranges*	1170	1189	1241	1463	1490	1948	2056	2264	2327	2606	2684
Detected species	{Lm}	Lm	Lm	Lm	Lm	Lm	Lm	Lm	Lm	Lm	Lm	Lm
% Species 1		95.1	94.9	95.1	91.9	95.4	95.5	95.3	94.4	94.7	95.4	95.1
% Species 2	{<5%}	0.1	0.1	0.1	0.2	0.1	0.2	0.2	0.1	0.1	0.1	0.1
Unclassified reads (%)		4.7	5.0	4.7	7.8	4.3	4.3	4.4	5.4	5.1	4.4	4.7
Length at >25 x min. coverage (Mbp)	{>28 ∧ <31}	2.9	3.0	2.9	3.0	3.0	2.9	3.0	2.9	3.0	2.9	3.0
Length [1-25] x min. coverage (kbp)	{<250}	0.0	0.3	2.8	0.6	0.8	2.5	2.6	0.0	1.0	0.6	0.0
No. of contigs at 25 x min. coverage	{>0}	29	64	42	52	45	55	67	35	40	57	27
No. of contigs [1-25] x min. coverage	{<1000}	0	1	4	1	1	4	4	0	2	1	0
Average coverage	{>50}	184	176	150	154	150	168	118	204	181	146	178
No. of reads (x 1000)		3902	3900	3137	3604	3233	3452	2548	4436	4022	2969	3826
Average read length		142	138	142	131	143	143	145	136	138	144	141
Average insert size		243	219	259	179	243	247	267	245	254	263	238
N50 (kbp)		329	87	153	99	144	87	76	201	173	99	342

		Laboratory 35 ¤										
Quali-/Quantitative	Ranges*	1206	1330	1424	1466	1529	1751	2202	2491	2567	2673	2978
Detected species	{Lm}	Lm	Lm	Lm	Lm	Lm	Lm	Lm	Lm	Lm	Lm	Lm
% Species 1		98.6	98.3	98.6	98.2	98.6	98.5	98.5	98.5	98.3	98.1	98.4
% Species 2	{<5%}	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Unclassified reads (%)		1.3	1.7	1.3	1.7	1.4	1.5	1.5	1.5	1.7	1.8	1.5
Length at >25 x min. coverage (Mbp)	{>28 ∧ <31}	2.9	2.9	3.0	3.0	2.9	2.9	3.0	2.9	2.9	3.0	3.0
Length [1-25] x min. coverage (kbp)	{<250}	77.1	1.6	5.0	6.1	7.9	1.3	3.4	0.0	0.0	6.0	13.7
No. of contigs at 25 x min. coverage	{>0}	74	36	44	76	118	71	67	29	35	50	45
No. of contigs [1-25] x min. coverage	{<1000}	62	4	8	9	10	2	2	0	0	9	12
Average coverage	{>50}	90	204	116	144	56	100	47	166	230	193	185
No. of reads (x 1000)		1826	4014	2388	2998	1111	2013	956	3308	4574	4073	3813
Average read length		148	147	147	145	147	147	147	148	145	145	146
Average insert size		NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
N50 (kbp)		76	343	237	127	45	90	114	425	202	339	334

		Laboratory 56										
Quali-/Quantitative	Ranges*	1315	1452	1645	1687	1710	1961	2254	2262	2328	2422	2539
Detected species	{Lm}	*Lm*	*Lm*	*Lm*	*Lm*	*Lm*	*Lm*	*Lm*	*Lm*	*Lm*	*Lm*	*Lm*
% Species 1		*98.5*	*98.7*	*97.9*	*98.5*	*98.4*	*98.5*	*97.0*	*98.1*	*98.2*	*98.1*	*98.1*
% Species 2	{<5%}	*0.0*	*0.0*	*0.0*	*0.0*	*0.0*	*0.0*	*0.0*	*0.0*	*0.1*	*0.0*	*0.0*
Unclassified reads (%)		*1.5*	*1.3*	*2.1*	*1.5*	*1.5*	*1.5*	*3.0*	*1.9*	*1.7*	*1.7*	*1.9*
Length at >25 x min. coverage (Mbp)	{>28 ∧ <31}	*2.6*	*2.8*	*2.9*	*2.9*	*2.9*	*2.7*	*2.9*	*2.7*	*2.8*	*2.4*	*2.9*
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*	*0.0*
No. of contigs at 25 x min. coverage #	{>0}	*2047*	*775*	*213*	*608*	*229*	*2272*	*143*	*1189*	*1672*	*3222*	*389*
No. of contigs [1-25] x min. coverage #	{<1000}	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*0*
Average coverage	{>50}	*281*	*310*	*197*	*242*	*308*	*206*	*370*	*380*	*238*	*154*	*524*
No. of reads (x 1000)		*2791*	*3109*	*2097*	*2411*	*3053*	*2044*	*4159*	*3851*	*2479*	*1515*	*6023*
Average read length		*315*	*297*	*275*	*306*	*301*	*313*	*264*	*290*	*300*	*318*	*264*
Average insert size #		*NA*	*NA*	*NA*	*NA*	*NA*	*NA*	*NA*	*NA*	*NA*	*NA*	*NA*
N50 (kbp) #		*2*	*6*	*24*	*8*	*26*	*2*	*41*	*4*	*3*	*1*	*14*

		Laboratory 70										
Quali-/Quantitative	**Ranges***	**1370**	**1672**	**1685**	**1690**	**1794**	**1929**	**2280**	**2540**	**2699**	**2716**	**2845**
Detected species	{Lm}	Lm	Lm	Lm	Lm	Lm	Lm	Lm	Lm	Lm	Lm	Lm
% Species 1		97.9	97.9	98.0	98.0	97.9	97.9	98.1	97.8	97.8	98.1	97.9
% Species 2	{<5%}	0.2	0.1	0.1	0.1	0.1	0.2	0.2	0.2	0.1	0.1	0.1
Unclassified reads (%)		1.7	1.8	1.7	1.7	1.9	1.8	1.6	1.7	1.8	1.6	1.9
Length at >25 x min. coverage (Mbp)	{>28 ∧ <31}	2.9	3.0	3.0	3.0	2.9	3.0	2.9	3.0	3.0	3.0	2.9
Length [1-25] x min. coverage (kbp)	{<250}	0.0	0.0	0.0	0.0	0.0	0.0	0.0	86.6	0.0	0.0	0.0
No. of contigs at 25 x min. coverage	{>0}	17	17	20	21	14	18	21	19	16	21	16
No. of contigs [1-25] x min. coverage	{<1000}	0	0	0	0	0	0	0	1	0	0	0
Average coverage	{>50}	47	51	47	50	38	44	37	32	60	43	50
No. of reads (x 1000)		926	1046	957	1013	750	882	724	663	1216	877	973
Average read length		151	151	151	151	151	151	151	151	151	151	151
Average insert size		309	331	332	331	321	336	328	336	319	320	300
N50 (kbp)		439	439	439	425	439	396	425	304	439	439	439

		Laboratory 100										
Quali-/Quantitative	**Ranges***	**1065**	**1365**	**1371**	**1425**	**1656**	**1698**	**2030**	**2054**	**2086**	**2885**	**2961**
Detected species	{Lm}	Lm	Lm	Lm	Lm	Lm	Lm	Lm	Lm	Lm	Lm	Lm
% Species 1		99.1	99.3	90.4	99.1	99.2	99.2	99.1	99.3	99.0	99.2	99.2
% Species 2	{<5%}	0.1	0.0	3.4	0.0	0.0	0.0	0.1	0.0	0.0	0.0	0.0
Unclassified reads (%)		0.7	0.6	4.3	0.8	0.7	0.7	0.8	0.6	0.8	0.8	0.7
Length at >25 x min. coverage (Mbp)	{>28 ∧ <31}	3.0	2.9	2.9	3.0	3.0	3.0	3.1	2.9	3.0	2.9	3.0
Length [1-25] x min. coverage (kbp)	{<250}	6.1	6.0	84.0	0.0	0.0	4.3	0.0	0.0	0.0	0.0	0.0
No. of contigs at 25 x min. coverage	{>0}	21	18	36	20	21	19	21	16	19	18	16
No. of contigs [1-25] x min. coverage	{<1000}	2	4	113	0	0	3	0	0	0	0	0
Average coverage	{>50}	48	78	109	94	114	85	95	85	62	76	113
No. of reads (x 1000)		613	954	1490	1144	1440	1062	1235	1013	807	927	1404
Average read length		238	238	243	244	240	242	238	243	232	241	242
Average insert size		346	323	373	386	347	369	323	377	318	371	372
N50 (kbp)		425	425	580	418	441	439	439	540	440	439	580

		Laboratory 105										
Quali-/Quantitative	**Ranges***	**1030**	**1112**	**1240**	**1635**	**1643**	**1679**	**2047**	**2680**	**2788**	**2942**	**2953**
Detected species	{Lm}	Lm	Lm	Lm	Lm	Lm	Lm	Lm	Lm	Lm	Lm	Lm
% Species 1		98.7	98.5	98.2	98.3	98.2	98.5	98.3	98.3	98.3	98.4	98.1
% Species 2	{<5%}	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.1
Unclassified reads (%)		1.3	1.4	1.7	1.6	1.7	1.4	1.7	1.6	1.6	1.6	1.8
Length at >25 x min. coverage (Mbp)	{>28 ∧ <31}	2.9	3.0	3.0	3.0	3.0	2.9	2.9	3.0	3.0	2.9	3.1
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	0.0
No. of contigs at 25 x min. coverage	{>0}	16	17	20	18	16	15	17	16	18	15	17
No. of contigs [1-25] x min. coverage	{<1000}	0	0	0	0	0	0	0	0	0	0	0
Average coverage	{>50}	448	432	301	225	169	326	209	235	353	364	343
No. of reads (x 1000)		8775	8827	6064	4607	3464	6480	4173	4745	7262	7117	7139
Average read length		148	148	148	148	148	148	147	148	147	148	148
Average insert size		341	377	363	406	400	356	408	396	324	370	407
N50 (kbp)		580	440	426	440	440	580	439	438	440	540	440

		Laboratory 108										
Quali-/Quantitative	Ranges*	1208	1406	1504	1704	1943	1995	2040	2207	2512	2877	2904
Detected species	{Lm}	Lm	Lm	Lm	Lm	Lm	Lm	Lm	Lm	Lm	Lm	Lm
% Species 1		98.9	98.9	98.9	98.8	98.9	98.9	98.7	98.9	98.9	99.0	98.9
% Species 2	{<5%}	0.0	0.0	0.0	0.0	0.0	0.0	0.1	0.0	0.0	0.0	0.0
Unclassified reads (%)		1.0	1.1	1.0	1.2	1.0	1.1	1.2	1.0	1.1	1.0	1.0
Length at >25 x min. coverage (Mbp)	{>28 ∧ <31}	2.5	2.6	2.6	2.5	2.5	2.6	2.6	2.6	2.5	2.5	2.6
Length [1-25] x min. coverage (kbp)	{<250}	7.4	2.5	1.3	9.1	2.1	2.5	3.8	8.6	9.3	0.0	0.7
No. of contigs at 25 x min. coverage #	{>0}	3460	3461	3477	3562	3441	3546	3789	3555	3700	3215	3282
No. of contigs [1-25] x min. coverage #	{<1000}	32	11	6	39	9	11	17	39	38	0	3
Average coverage	{>50}	55	76	90	63	99	81	93	83	55	93	82
No. of reads (x 1000)		547	767	899	630	957	835	1000	836	546	919	825
Average read length		302	302	307	301	303	299	290	306	302	296	299
Average insert size #		NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
N50 (kbp) #		1	1	1	1	1	1	1	1	1	1	1

		Laboratory 129										
Quali-/Quantitative	Ranges*	1071	1211	1215	1347	1387	1744	2284	2397	2450	2702	2743
Detected species	{Lm}	Lm	Lm	Lm	Lm	Lm	Lm	Lm	Lm	Lm	Lm	Lm
% Species 1		98.0	97.9	96.5	96.2	97.9	96.9	96.2	98.6	98.3	98.6	98.3
% Species 2	{<5%}	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Unclassified reads (%)		2.0	2.1	3.4	3.7	2.0	3.1	3.7	1.4	1.6	1.3	1.6
Length at >25 x min. coverage (Mbp)	{>28 ∧ <31}	3.0	2.9	3.0	3.0	3.0	2.9	3.1	2.9	3.0	2.9	3.0
Length [1-25] x min. coverage (kbp)	{<250}	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.6	0.0	0.0
No. of contigs at 25 x min. coverage	{>0}	18	17	30	24	24	31	22	25	58	20	16
No. of contigs [1-25] x min. coverage	{<1000}	0	0	0	0	0	0	0	0	2	0	0
Average coverage	{>50}	151	140	212	150	238	264	150	218	166	153	183
No. of reads (x 1000)		3036	2786	4374	3096	4898	5231	3156	4264	3419	3023	3702
Average read length		149	148	149	148	148	147	148	148	147	149	149
Average insert size		355	349	366	378	315	337	363	329	339	383	348
N50 (kbp)		425	460	225	359	359	183	410	249	100	261	431

		Laboratory 135										
Quali-/Quantitative	Ranges*	1005	1172	1286	1760	1799	1801	2170	2246	2249	2444	2911
Detected species	{Lm}	Lm	Lm	Lm	Lm	Lm	Lm	Lm	Lm	Lm	Lm	Lm
% Species 1		96.5	97.3	96.2	95.5	97.3	97.2	96.1	96.7	97.5	96.3	94.3
% Species 2	{<5%}	0.0	0.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.3
Unclassified reads (%)		3.4	2.6	3.8	4.5	2.6	2.8	3.8	3.2	2.4	3.7	3.2
Length at >25 x min. coverage (Mbp)	{>28 ∧ <31}	3.0	3.0	3.0	2.9	3.0	2.9	2.9	3.0	3.1	3.0	3.0
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	8.0
No. of contigs at 25 x min. coverage	{>0}	32	23	25	47	82	26	32	22	19	39	41
No. of contigs [1-25] x min. coverage	{<1000}	0	0	0	0	0	0	0	0	0	0	9
Average coverage	{>50}	309	285	266	339	304	332	318	297	299	306	248
No. of reads (x 1000)		6748	5970	5734	7459	6510	6898	6839	6527	6377	6800	5166
Average read length		139	145	139	132	142	142	135	138	145	135	148
Average insert size		292	352	297	252	307	323	258	285	329	256	400
N50 (kbp)		237	322	270	149	72	414	189	359	425	163	304

		Laboratory 141										
Quali-/Quantitative	**Ranges***	**1063**	**1309**	**1395**	**1418**	**1649**	**1655**	**2219**	**2322**	**2395**	**2585**	**2677**
Detected species	{Lm}	Lm	Lm	Lm	Lm	Lm	Lm	Lm	Lm	Lm	Lm	Lm
% Species 1		96.9	95.5	96.8	96.9	97.4	95.8	96.2	96.4	97.0	96.3	96.1
% Species 2	{<5%}	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.1	0.0
Unclassified reads (%)		3.0	4.3	3.1	2.9	2.4	4.1	3.7	3.5	2.9	3.5	3.8
Length at >25 x min. coverage (Mbp)	{>28 ∧ <31}	2.9	3.0	2.9	3.0	3.0	3.0	2.9	2.9	3.0	3.1	3.0
Length [1-25] x min. coverage (kbp)	{<250}	0.0	0.0	0.0	0.5	0.4	0.0	0.0	0.0	0.5	0.0	0.0
No. of contigs at 25 x min. coverage	{>0}	26	28	19	26	61	22	19	22	37	29	25
No. of contigs [1-25] x min. coverage	{<1000}	0	0	0	1	1	0	0	0	1	0	0
Average coverage	{>50}	106	80	98	107	83	110	113	98	102	84	76
No. of reads (x 1000)		1249	986	1194	1345	1115	1369	1371	1176	1233	1044	937
Average read length		256	251	241	245	230	247	243	250	253	252	250
Average insert size		306	315	283	285	258	285	289	306	298	312	315
N50 (kbp)		282	317	525	259	114	359	425	414	175	225	300

		Laboratory 142										
Quali-/Quantitative	**Ranges***	**1050**	**1141**	**1238**	**1275**	**1500**	**1554**	**2065**	**2300**	**2391**	**2488**	**2715**
Detected species	{Lm}	*Lm*	*Lm*	*Lm*	*Lm*	*Lm*	*Lm*	*Lm*	*Lm*	*Lm*	*Lm*	*Lm*
% Species 1		*98.0*	*98.9*	*97.5*	*98.2*	*98.5*	*98.2*	*98.5*	*97.7*	*98.3*	*98.5*	*97.7*
% Species 2	{<5%}	*0.1*	*0.1*	*0.2*	*0.2*	*0.0*	*0.2*	*0.1*	*0.3*	*0.1*	*0.1*	*0.1*
Unclassified reads (%)		*1.9*	*1.0*	*2.2*	*1.5*	*1.4*	*1.5*	*1.3*	*1.9*	*1.5*	*1.3*	*2.0*
Length at >25 x min. coverage (Mbp)	{>28 ∧ <31}	*2.9*	*3.0*	*3.0*	*2.9*	*3.0*	*3.0*	*2.9*	*2.9*	*3.0*	*3.0*	*3.1*
Length [1-25] x min. coverage (kbp)	{<250}	*0.5*	*0.0*	*0.0*	*1.1*	*0.0*	*7.4*	*0.0*	*2.2*	*0.0*	*0.0*	*0.5*
No. of contigs at 25 x min. coverage	{>0}	*22*	*19*	*18*	*16*	*20*	*20*	*18*	*15*	*19*	*17*	*18*
No. of contigs [1-25] x min. coverage	{<1000}	*1*	*0*	*0*	*2*	*0*	*5*	*0*	*4*	*0*	*0*	*1*
Average coverage	{>50}	*83*	*114*	*53*	*71*	*57*	*85*	*85*	*61*	*62*	*61*	*50*
No. of reads (x 1000)		*1061*	*1443*	*666*	*862*	*731*	*1082*	*1062*	*763*	*791*	*778*	*637*
Average read length		*230*	*239*	*245*	*239*	*238*	*236*	*237*	*236*	*238*	*235*	*244*
Average insert size		*353*	*382*	*439*	*386*	*380*	*362*	*371*	*382*	*380*	*357*	*436*
N50 (kbp)		*456*	*440*	*440*	*580*	*440*	*425*	*439*	*425*	*440*	*563*	*439*

		Laboratory 146										
Quali-/Quantitative	**Ranges***	**1339**	**1356**	**1630**	**1721**	**1763**	**1778**	**2050**	**2261**	**2529**	**2854**	**2967**
Detected species	{Lm}	*Lm*	*Lm*	*Lm*	*Lm*	*Lm*	*Lm*	*Lm*	*Lm*	*Lm*	*Lm*	*Lm*
% Species 1		*95.3*	*95.2*	*95.3*	*94.3*	*94.7*	*95.3*	*95.3*	*95.2*	*95.1*	*95.3*	*94.4*
% Species 2	{<5%}	*0.0*	*0.0*	*0.0*	*0.0*	*0.0*	*0.0*	*0.0*	*0.0*	*0.0*	*0.0*	*0.1*
Unclassified reads (%)		*4.6*	*4.8*	*4.6*	*5.7*	*5.3*	*4.7*	*4.7*	*4.7*	*4.8*	*4.7*	*5.4*
Length at >25 x min. coverage (Mbp)	{>28 ∧ <31}	*2.9*	*3.0*	*3.0*	*3.0*	*3.0*	*2.9*	*3.0*	*2.9*	*3.1*	*2.9*	*2.9*
Length [1-25] x min. coverage (kbp)	{<250}	*1.0*	*0.0*	*0.0*	*0.0*	*0.0*	*0.0*	*0.9*	*3.0*	*0.0*	*0.0*	*27.6*
No. of contigs at 25 x min. coverage	{>0}	*30*	*22*	*21*	*25*	*30*	*14*	*28*	*39*	*21*	*23*	*51*
No. of contigs [1-25] x min. coverage	{<1000}	*2*	*0*	*0*	*0*	*0*	*0*	*2*	*5*	*0*	*0*	*21*
Average coverage	{>50}	*175*	*135*	*180*	*200*	*161*	*151*	*179*	*185*	*184*	*174*	*99*
No. of reads (x 1000)		*5151*	*4111*	*5466*	*6063*	*4943*	*4404*	*5441*	*5368*	*5652*	*5139*	*3028*
Average read length		*100*	*99*	*99*	*98*	*98*	*100*	*99*	*99*	*100*	*99*	*98*
Average insert size		*297*	*282*	*318*	*203*	*220*	*311*	*311*	*276*	*303*	*290*	*244*
N50 (kbp)		*214*	*439*	*348*	*425*	*347*	*425*	*201*	*173*	*439*	*294*	*99*

		Laboratory 149										
Quali-/Quantitative	Ranges*	1003	1122	1372	1432	1938	1987	2027	2366	2582	2685	2733
Detected species	{*Lm*}	*Lm*	*Lm*	*Lm*	*Lm*	*Lm*	*Lm*	*Lm*	*Lm*	*Lm*	*Lm*	*Lm*
% Species 1		*89.7*	*89.5*	*92.0*	*88.9*	*88.4*	*90.4*	*91.7*	*95.8*	*92.5*	*94.1*	*94.6*
% Species 2	{<5%}	*6.4*	*7.2*	*2.7*	*5.3*	*3.9*	*5.6*	*4.3*	*1.3*	*4.6*	*3.2*	*2.9*
Unclassified reads (%)		*2.7*	*2.7*	*4.1*	*4.5*	*3.5*	*2.4*	*2.7*	*1.6*	*2.1*	*2.0*	*1.8*
Length at >25 x min. coverage (Mbp)	{>28 ∧ <31}	*3.0*	*3.0*	*3.0*	*2.9*	*2.9*	*3.0*	*3.0*	*2.9*	*3.0*	*3.1*	*2.9*
Length [1-25] x min. coverage (kbp)	{<250}	*0.4*	*0.0*	*0.0*	*0.0*	*0.0*	*0.7*	*0.0*	*0.0*	*0.0*	*0.0*	*0.0*
No. of contigs at 25 x min. coverage	{>0}	*23*	*21*	*21*	*18*	*16*	*22*	*18*	*19*	*22*	*20*	*31*
No. of contigs [1-25] x min. coverage	{<1000}	*1*	*0*	*0*	*0*	*0*	*2*	*0*	*0*	*0*	*0*	*0*
Average coverage	{>50}	*103*	*91*	*118*	*75*	*87*	*78*	*49*	*120*	*137*	*177*	*196*
No. of reads (x 1000)		*1442*	*1299*	*1499*	*953*	*1006*	*1027*	*590*	*1287*	*1967*	*2469*	*2537*
Average read length		*262*	*258*	*266*	*273*	*285*	*270*	*284*	*284*	*246*	*245*	*249*
Average insert size		*315*	*331*	*361*	*386*	*457*	*325*	*467*	*375*	*290*	*288*	*291*
N50 (kbp)		*440*	*425*	*440*	*580*	*580*	*440*	*439*	*540*	*440*	*439*	*298*

Quality assessment made by the EQA provider in-house quality control pipeline.
**: indicative QC ranges*
NA: not available
#: QC values unreliable due to assembly issues for Ion Torrent data
¤: single end reads
Lm: L. monocytogenes

Annex 12. Accessing Genomes

	Participant		EQA provider*	
Lab ID	**Cluster**	**Description Genome 1**	**Contamination**	**Cluster**
EQA provider	Yes	Cluster isolate (REF5/REF11) mixed with a *L. innocua* (approx. 15%)	Yes	Yes
19	No	Genome 1 is discarded from the analysis because of poor quality sequence. The size of the genome is almost double the size than expected, the N50 value is low, the number of assembled contigs are very high and the number of unidentified bases (N) are also high. Also, the number of loci with multiple consensus is high and the corepercent is low.	Yes	-
35	No	Even though the List_genome1 forms a cluster with 0 allelic differences with List_genome3 after 1701 locus cgMLST, I would not consider this result as a valid result, as 215 out of 1701 loci in the List_genome1 sample were not called in the List_genome1 sample (the percentage of good cgMLST targets for List_genome1 is 87.5% only, and thus below the QC cut-off in SeqSphere that is needed to assign a CT value).	Yes	No
56	No	At the moment, it has not been possible to perform the analysis.	-	-
70	Yes	Using mapping-based assembly in Seqsphere (BWA), 1390/1701 (81.7%) of alleles were called which is unusually low for this high coverage (50x). This could be due to mixture of 2 closely related isolates. We would normally re-isolate and re-sequence. But one of the twos isolates at least seems very close to representative isolate. As it is, the isolate has an allele difference of 2 to representative isolate 2 (1690). In a crisis situation, it cannot be excluded that this sample/isolate forms part of the cluster.	Yes	Yes
100	No	Genome 1 has an average coverage of 35X and a percentage of good targets of 84,9%. For these reasons, the results are not satisfactory enough to give a conclusion.	Yes	-
105	No	The result for several QCs were 'WARN' or 'FAIL': Assembly size >3.3Mb ('WARN')- CGM > 1 ('FAIL')- CGC: 90.8% ('WARN').	Yes	-
108	Yes	It clusters within ST8 and the isolates 1504,1406,1943 and 1995, it also clusters with genome 3. It is also closely related to isolate 2207. Cluster analysis was made within ST8. Epidemiological criteria is also included considering clusters on daily basis. We also saw that 'genome 1' was very large, if it was our sample we would consider it to be contaminated.	Yes	Yes
129	No	Genome 1 did not pass our QC. The percentage of good targets was 88.1 %. Also, the contig count of 900 suggested poor quality. Resequencing would have been requested. (If genome 1 is analysed despite of poor quality, it differed by 8 alleles from isolate 1215 in cgMLST, and by 15 alleles if accessory genome was used in addition to the core.)	Yes	No
135	No	Genome 1 was excluded because of missing values in more than 10% of distance columns.	Yes	-
141	No	63 alleles difference low coverage (av. cov. assembeled 12x), only 83,1 % good targets we would recommend to repeat sequencing of this sample genome size very high (5.2 Mbases for *Listeria* ~2.8-3.0 Mbases).	Yes	-
142	No	Contamination with *L. innocua*< 95% of cgMLST alleles called.	Yes	-
146	No	We were unable to process this genome through our pipeline as it was mixed with an *L. innocua* isolate and so failed our quality checks. If we had the culture we would have purified and re-sequenced.	Yes	-
149	No	87.4% good targets (< 95% good targets) Low coverage. Low N50Re-sequencing is necessary. Decision on inclusion into the cluster cannot be made with the current sequence.	Yes	-

	Participant		EQA Provider*	
Lab ID	**Cluster**	**Description Genome 2**	**Quality issues**	**Cluster**
EQA provider	No	A cluster isolate (REF8) with altered coverage (reduced to 10x).	Yes	Yes
19	No	Genome 2 is discarded from the analysis because of poor quality sequence. The genome size is too small, the N50 value is low and the average coverage is low. The corepercent is too low and almost no consensus is seen for any loci. Also the species analysis identified a high number of unidentified reads.	Yes	-
35	No	Genome 2 did not pass QC cut-off (20.8% good alleles only).	Yes	-
56	No	the coverage (11X) is not sufficient for the analysis.	Yes	-
70	No	Using mapping-based assembly in Seqsphere (BWA), 996/1701 (58.6%) alleles called. This isolate has an allele difference of 23 against representative isolate 2 (1690).	Yes	No
100	No	Genome 2 has an average coverage of 11X and a percentage of good targets of 21,1%. For these reasons, the results are not satisfactory enough to give a conclusion.	Yes	-
105	No	The result for several QCs were 'WARN' or 'FAIL': Assembly size: 1.3MbCGM > 1 ('FAIL'), CGC< 90% ('FAIL').	Yes	-
108	No	Genome 2 is another ST ST1419, and it doesn't cluster with any of the isolates in the cluster.	No	No
129	No	Genome 2 did not pass our QC. The percentage of good targets was only 22.3 % and average coverage was 11. Also, the contig count of 3579 suggested poor quality. Resequencing would have been requested.	Yes	-
135	No	Genome 2 was excluded because of missing values in more than 10% of distance columns.	Yes	-
141	No	44 alleles difference low coverage (av. cov. assembled 5x), only 49,8 % good targets we would recommend to repeat sequencing of this sample.	Yes	-
142	No	Low coverage<95% of cgMLST alleles called.	Yes	-
146	No	We were unable to process this genome through our pipeline as it had too low of a sequence yield. We require a yield of least 90Mb to pass our quality checks.	Yes	-
149	No	20.9% good targets (< 95% good targets) Low coverage Low N50Re-sequencing is necessary. Decision on inclusion into the cluster cannot be made with the current sequence.	Yes	-

	Participant		EQA provider*	
Lab ID	**Cluster**	**Description Genome 3**	**QC accepted**	**Cluster**
EQA provider	Yes	A cluster isolate assembled with SKESA to a FASTA file. Zero allelic difference to the cluster (REF5).	Yes	Yes
19	Yes	Genome 3 cluster with the outbreak isolates and has zero allele differences to the representative outbreak isolate 1490, and is therefore considered as part of the outbreak.	Yes	Yes
35	No	Allelic difference between the two genomes that passed the QC cut-off (genome 3 and genome 4) is 29 and thus above the cut-off used for cluster detection.	Yes	No
56	No	At the moment, our toolset can't manage Fasta files.	No	-
70	Yes	AD of 0 against representative isolate 2 (1690).	Yes	Yes
100	Yes	Genome 3 has a percentage of good targets of 100%. The allelic difference with the representative isolate (Isolate 1 - 1065) is 0. Genome 3 is closely related to the representative isolate.	Yes	Yes
105	Yes	Genome 3 assembly matches with representative isolate (isolate ID 1112) (0 Allelic differences).	Yes	Yes
108	Yes	It clusters within ST8 and the isolates 1504, 1406, 1943 and 1995, it also clusters with genome 1. It is also closely related to isolate 2207.	Yes	Yes
129	Yes	Genome 3 had 0 allele differences to one of the isolates included in the cluster. Percentage of good targets was 99.8 (pass) and contig count (27) suggested good sequence quality.	Yes	Yes
135	Yes	Genome 3 is identical to isolate 1 and 4 which are part of the cluster.	Yes	Yes
141	Yes	0 alleles difference from selected representative isolate 1309 (ST8, CT2994) good quality of the sequence, 100 % good targets identical CT (ST8, CT2994) to cluster isolate 1309.	Yes	Yes
142	Yes	ST80 cgMLST allele differences.	Yes	Yes
146	No	We were unable to process this genome through our pipeline as it was provided in fasta format and we require the raw reads as fastq files. Typically, we would ask for the fastqs to be sent in this scenario.	No	-
149	Yes	100% good targetsST8, CT29940 AD from 1003KRAKEN: *Listeria*.	Yes	Yes

	Participant		EQA provider*	
Lab ID	**Cluster**	**Description Genome 4**	**QC accepted**	**Cluster**
EQA provider	No	A non-cluster isolate. 28 allelic difference to the cluster isolate (REF5).	Yes	No
19	No	Genome 4 differ with 28 allele differences to the representative outbreak isolate and is therefore not considered a part of the outbreak.	(Yes)	No
35	No	Allelic difference between the two genomes that passed the QC cut-off (genome 4 and genome 3) is 29 and thus above the cut-off used for cluster detection.	Yes	No
56	No	Because the allele difference within the cluster is over 25.	(Yes)	No
70	No	AD of 29 against representative isolate 2 (1690).	(Yes)	No
100	No	Genome 4 has an average coverage of 77X and a percentage of good targets of 99,9%. The allelic difference with the representative isolate (Isolate 1 - 1065) is 37, because of that the Genome 4 is not closely related to the representative isolate.	Yes	No
105	No	Genome 4 assembly is 29 allelic differences.	(Yes)	No
108	No	Genome 4 belongs to ST8, but it differs a lot from the cluster.	(Yes)	No
129	No	Genome 4 differed from isolate 1215 by 66 alleles. QC was ok for this genome.	(Yes)	No
135	No	Genome 4 has a distance of 37 which falls out of our cut-off of 10 that we use to determine a cluster.	(Yes)	No
141	No	29 alleles difference from selected representative isolate 1309 (ST8, CT2994) good quality of the sequence, 100 % good targets different CT (ST8, CT8350) to cluster isolate 1309.	Yes	No
142	No	ST829 cgMLST allele differences (>7 AD).	(Yes)	No
146	No	We processed this genome through our pipeline and found it to belong to CC8, however the closest isolate (1778) was 650 SNPs distant and so we would not consider it to be part of the same cluster.	Yes	No
149	No	100% good targetsST8, CT835029 AD from 1003KRAKEN *Listeria*.	Yes	No

(Yes): participant performed a cluster analysis, therefore the EQA provider assume the quality was accepted by the participant, despite no details.

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

Annex 13. EQA-7 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. *Listeria* EQA-7 2019

Dear participant,

Welcome to the seventh External Quality Assessment (EQA-7) scheme for typing of *Listeria* in 2019–2020. Please note that most of the fields are required to be filled in before the submission can be completed. Any comments can be written at the end of the form. You are always welcome to contact us at list.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
☐ Denmark
☐ Finland
☐ France
☐ Germany
☐ Greece
☐ Hungary
☐ Italy
☐ Latvia
☐ Luxembourg
☐ Norway
☐ Slovenia
☐ Spain
☐ Sweden
☐ 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. Serotyping of *Listeria*

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 serotyping part – Go to 14

9. Serotyping isolate ID's

Please enter the isolate ID (4 digits)

Listeria

Isolate 1 ___
Isolate 2 ___
Isolate 3 ___
Isolate 4 ___
Isolate 5 ___
Isolate 6 ___
Isolate 7 ___
Isolate 8 ___
Isolate 9 ___
Isolate 10 ___
Isolate 11 ___

10. Submitting results – Serotyping of *Listeria*

(State one answer only)

☐ Both molecular and conventional serotyping – Go to 11
☐ Molecular serotyping – Go to 11
☐ Conventional serotyping – Go to 13

11. Method used for molecular serotyping of *Listeria*

(State one answer only)

☐ PCR-based
☐ WGS-based

12. Results for serotyping *Listeria* – molecular serotyping

Please select the serotype

(State only one answer per question)

Isolate	Molecular serotype					
	IIa	IIb	IIc	IVb	L	Un-typeable
Isolate 1						
Isolate 2						
Isolate 3						
Isolate 4						
Isolate 5						
Isolate 6						
Isolate 7						
Isolate 8						
Isolate 9						
Isolate 10						
Isolate 11						

13. Results for serotyping *Listeria* – conventional serotyping

Please select the serotype

(State only one answer per question)

Isolate		Conventional serotype												
Isolate 1	1/2a	1/2b	1/2c	3a	3b	3c	4a	4ab	4b	4c	4d	4e	7	Autoagglutinable	Un-typeable
Isolate 2															
Isolate 3															
Isolate 4															
Isolate 5															
Isolate 6															
Isolate 7															
Isolate 8															
Isolate 9															
Isolate 10															
Isolate 11															

14. Submitting cluster results

(State one answer only)

☐ Cluster analyses based on PFGE and/or WGS – Go to 15
☐ Did not participate in the cluster part – Go to 124

15. Cluster isolates ID's

Please enter the cluster isolate ID (4 digits)

Isolate	Cluster isolate ID
Isolate 1	
Isolate 2	
Isolate 3	
Isolate 4	
Isolate 5	
Isolate 6	
Isolate 7	
Isolate 8	
Isolate 9	
Isolate 10	
Isolate 11	

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 23

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 combining *ApaI*- and *AscI*- results:

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

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

Indicate the isolate ID

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

21. *AscI* – Total number of bands (>33kb) in the selected representative cluster isolate

22. Results for cluster analysis – PFGE (*ApaI* and *AscI*)

Please use 9999 for not analysed

	ApaI – Total number of bands (>33kb)	*ApaI* – Number of bands with same/shared position as the profile of the selected cluster isolate (>33kb)	*AscI* – Total number of bands (>33kb)	*AscI* – Number of bands with same/shared position as the profile of the selected cluster isolate (>33kb)
Isolate 1				
Isolate 2				
Isolate 3				
Isolate 4				
Isolate 5				
Isolate 6				
Isolate 7				
Isolate 8				
Isolate 9				
Isolate 10				
Isolate 11				

23. Submitting cluster results

(State one answer only)

☐ Cluster analysis based on WGS data – Go to 24
☐ Do not wish to submit any cluster results based on WGS data – Go to 124

24. Cluster analysis based on WGS data

25. 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 27
☐ Allele-based – Go to 34
☐ Other – Go to 26

26. If another analysis is used, please describe your approach: – Go to 41

27. Please report the used SNP-pipeline (reference if publicly available or in-house pipeline)

28. Please select the approach used for the SNP analysis

(State one answer only)

☐ Reference-based – Go to 29
☐ Assembly-based – Go to 32

29. Reference genome used

Please indicate Multi-locus Sequence Type (e.g. ST8) and isolate ID (e.g. one of the isolates from the current cluster, ID of a public reference strain or an in-house isolate)

30. Please indicate the read mapper used (e.g. BWA, Bowtie2)

31. Please indicate the variant caller used (e.g. SAMtools, GATK)

32. Please indicate the assembler used (e.g. SPAdes, Velvet)

33. Please specify the variant caller used (e.g. NUCMER)

34. Please select tools used for the allele analysis

(State one answer only)

☐ BioNumerics – Go to 36
☐ SeqPhere – Go to 36
☐ BIGSdb-*Lm* – Go to 36
☐ Other – Go to 35

35. If another tool is used please enter here:

36. Please indicate allele calling method

(State one answer only)

☐ Assembly-based and mapping-based – Go to 37
☐ Only assembly-based – Go to 37
☐ Only mapping-based – Go to 38

37. Please indicate the assembler used (e.g. SPAdes, Velvet)

38. Please select scheme used for the allele analysis

(State one answer only)

☐ Applied Math (wgMLST) – Go to 40
☐ Applied Math (cgMLST/Pasteur) – Go to 40
☐ Pasteur (cgMLST) – Go to 40
☐ Ruppitsch (cgMLST) – Go to 40
☐ Other – Go to 39

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

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

41. 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.

42. Please list the ID's for the isolates included in the cluster

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

43. Select a representative isolate in the cluster

Indicate the isolate ID

44. Results for cluster analysis (e.g. SNP or allelebased)

Please use 9999 for not analysed.

Isolate	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		
Isolate 11		

45. Analysis of the EQA provided genomes

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

46. 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

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

48. 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

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

50. 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

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

52. 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

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

54. 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 55
☐ No – Go to 94

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

(State one answer only)

☐ SNP-based – Go to 57
☐ Allele-based – Go to 64
☐ Other – Go to 56

56. If another analysis is used please describe your approach: - Go to 71

57. Please report the used SNP pipeline (reference if publicly available or in-house pipeline)

58. Please select the approach used for the SNP analysis

(State one answer only)

☐ Reference based – Go to 59
☐ Assembly based – Go to 62

59. Reference genome used

Please indicate Multi-locus Sequence Type (e.g. ST8) and isolate ID (e.g. one of the isolates from the current cluster, ID of a public reference strain or an in-house isolate)

60. Please indicate the read mapper used (e.g. BWA, Bowtie2)

61. Please indicate the variant caller used (e.g. SAMtools, GATK)

62. Please indicate the assembler used (e.g. SPAdes, Velvet)

63. Please specify the variant caller used (e.g. NUCMER)

64. Please select tool used for the allele analysis

(State one answer only)

☐ BioNumerics – Go to 66
☐ SeqPhere – Go to 66
☐ BIGSdb-*Lm* – Go to 66
☐ Other – Go to 65

65. If another tool is used please list here:

66. Please indicate allele calling method:

(State one answer only)

☐ Assembly based and mapping based – Go to 67
☐ Only assembly based – Go to 67
☐ Only mapping based – Go to 68

67. Please indicate the assembler used (e.g. SPAdes, Velvet)

68. Please select scheme used for the allele analysis

(State one answer only)

☐ Applied Math (wgMLST) – Go to 70
☐ Applied Math (cgMLST/Pasteur) – Go to 70
☐ Pasteur (cgMLST) – Go to 70
☐ Ruppitsch (cgMLST) – Go to 70
☐ Other – Go to 69

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

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

71. Additional analysis on data derived from WGS

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

(indicate the isolate ID)

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

Please use 9999 for not analysed

Isolate	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		
Isolate 11		

74. 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 75
☐ No – Go to 94

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

(State one answer only)

☐ SNP based – Go to 77
☐ Allele-based – Go to 84
☐ Other – Go to 76

76. If another analysis is used please describe your approach: - Go to 91

77. Please report the used SNP-pipeline (reference if publicly available or in-house pipeline)

78. Please select the approach used for the SNP analysis

(State one answer only)

☐ Reference based – Go to 79
☐ Assembly based – Go to 82

79. Reference genome used

Please indicate Multi-locus Sequence Type (e.g. ST8) and isolate ID (e.g. one of the isolates from the current cluster, ID of a public reference strain or an in-house isolate).

80. Please indicate the read mapper used (e.g. BWA, Bowtie2)

81. Please indicate the variant caller used (e.g. SAMtools, GATK)

82. Please indicate the assembler used (e.g. SPAdes, Velvet)

83. Please specify the variant caller used (e.g. NUCMER)

84. Please select tool used for the allele analysis

(State one answer only)

☐ BioNumerics – Go to 86
☐ SeqPhere – Go to 86
☐ BIGSdb-*Lm* – Go to 86
☐ Other – Go to 85

85. If another tool is used please enter here:

86. Please indicate allele calling method:

(State one answer only)

☐ Assembly based and mapping based – Go to 87
☐ Only assembly based – Go to 87
☐ Only mapping based – Go to 88

87. Please indicate the assembler used (e.g. SPAdes, Velvet)

88. Please select scheme used for the allele analysis

(State one answer only)

☐ Applied Math (wgMLST) – Go to 90
☐ Applied Math (cgMLST/Pasteur) – Go to 90
☐ Pasteur (cgMLST) – Go to 90
☐ Ruppitsch (cgMLST) – Go to 90
☐ Other – Go to 89

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

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

91. Third analysis on data derived from WGS

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

Indicate the isolate ID

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

Please use 9999 for not analysed

Isolate	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		
Isolate 11		

94. Additional questions to the WGS part

95. Where was the sequencing performed?

(State one answer only)

☐ In own laboratory
☐ Externally

96. Protocol used to prepare the library for sequencing:

(State one answer only)

☐ Commercial kits – Go to 97
☐ Non-commercial kits – Go to 99

97. Please indicate name of commercial kit:

98. If relevant please list deviation from commercial kit shortly in few bullets: - Go to 100

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

100. The sequencing platform used

(State one answer only)

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

101. If another platform is used please list here:

102. 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 Listeria EQA-5 and EQA-6 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 evaluated the current criteria.

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

☐ Yes
☐ No – Go to 105

104. Procedure used to evaluate confirmation of genus:

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

(State one answer only)

☐ Yes
☐ No – Go to 107

106. Procedure or threshold used for coverage:

107. Did you use Q score (Phred) to evaluate quality of sequence data?

(State one answer only)

☐ Yes
☐ No – Go to 109

108. Threshold or procedure used to evaluate Q score (Phred):

109. Did you use genome size to evaluate the quality of sequence data?

(State one answer only)

☐ Yes
☐ No – Go to 111

110. Procedure or threshold used for genome size:

111. Did you evaluate the number of good cgMLST loci?

(State one answer only)

☐ Yes
☐ No – Go to 113

112. Threshold or procedure used to evaluate the number of good cgMLST loci:

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

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

114. Other criteria used to evaluate the quality of sequence data – additional criteria 1:

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

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

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

118. Other criteria used to evaluate the quality of sequence data – additional criteria 3:

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

120. Other criteria used to evaluate the quality of sequence data – additional criteria 4:

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

122. Other criteria used to evaluate the quality of sequence data – additional criteria 5:

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

124. Comment(s):

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

125. Thank you for your participation

Thank you for filling out the submission form for the *Listeria* EQA-7.

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

We highly recommend to document this submission form by printing it. You will find the print option after pressing the 'Options' button.

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

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

European Centre for Disease Prevention and Control (ECDC)

Gustav III:s Boulevard 40, 16973 Solna, Sweden

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

An agency of the European Union
www.europa.eu

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

Contact us
publications@ecdc.europa.eu

Follow us on Twitter
@ECDC_EU

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

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

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

Publications Office
of the European Union