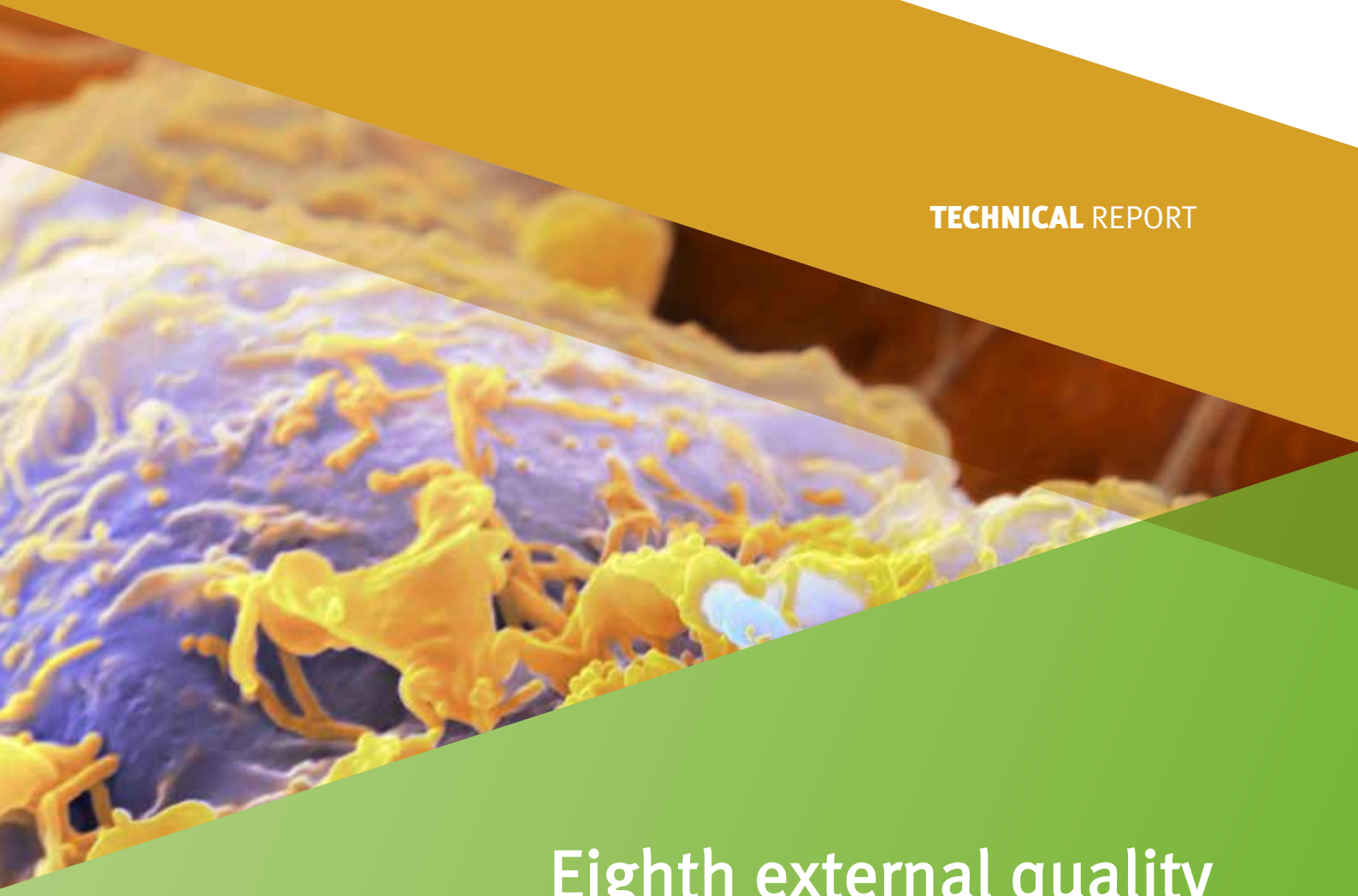
ecdc
EUROPEAN CENTRE FOR DISEASE PREVENTION AND CONTROL

TECHNICAL REPORT



Eighth external quality assessment scheme for *Listeria monocytogenes* typing

www.ecdc.europa.eu

ECDC TECHNICAL REPORT

Eighth 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. Eighth external quality assessment scheme for *Listeria monocytogenes* typing. Stockholm: ECDC; 2022.

Stockholm, August 2022

ISBN 978-92-9498-579-8
doi: 10.2900/733975
Catalogue number TQ-05-22-242-EN-N

Contents

Abbreviations iv
Executive summary 1
1. Introduction 2
1.1 Background 2
1.2 Surveillance of listeriosis 2
1.3 Objectives 3
2. Study design 4
2.1 Organisation 4
2.2 Selection of test isolates/genomes 4
2.3 Distribution of isolates/genomes 5
2.4 Testing 5
2.5 Data analysis 5
3. Results 6
3.1 Participation 6
3.2 Serotyping 6
3.3 Molecular typing-based cluster analysis 9
4. Discussion 18
4.1 Serotyping/-grouping 18
4.2 Molecular typing-based cluster analysis 18
5. Conclusions 20
6. Recommendations 21
6.1 Laboratories 21
6.2 ECDC and FWD-Net 21
6.3 EQA provider 21
References 22
Annex 1. List of participants 23
Annex 2. Participation overview EQA-7 and 8 24
Annex 3. Serotyping result scores 25
Annex 4. Reported cluster of closely-related isolates based on PFGE-derived data 26
Annex 5. Reported sequencing details 27
Annex 6. EQA provider cluster analysis based on WGS-derived data 28
Annex 7. Reported cluster of closely-related isolates based on WGS-derived data 29
Annex 8. Reported SNP distance and allelic differences 30
Annex 9. Additional reported QC parameters 31
Annex 10. Calculated qualitative/quantitative parameters 32
Annex 11. Accessing genomes 37
Annex 12. EQA-8 laboratory questionnaire 41

Figures

Figure 1. Participant scores for conventional serotyping of 11 test isolates 6
Figure 2. Correct conventional serotyping of three repeat isolates through EQA-1-8 (for labs participating in EQA-8) 7
Figure 3. Correct conventional serotyping of three repeat isolates through EQA-1-8 (for labs participating in EQA-8) 7
Figure 4. Correct molecular serogrouping of three repeat isolates from EQA-1-8 (for labs participating in EQA-8) 8
Figure 5. Average score of 11 test isolates 8
Figure 6. Reported SNP/allelic differences for each test isolate to selected cluster representative isolate 12
Figure 7. Minimum spanning tree of core genome multilocus sequence typing participant FASTQ files 13
Figure 8. Participant allele difference from reference result (EQA provider) for each test isolate 14

Tables

Table 1. Serotype/groups and sequence type of test isolates/genomes 4
Table 1a. Number/percentage of laboratories submitting results for each method 6
Table 2. Results of cluster analysis based on PFGE-derived data 9
Table 3. Results of cluster analysis based on WGS-derived data 10
Table 4. Results of SNP-based cluster analys 10
Table 5. Results of allele-based cluster analysis 11
Table 6. Summary of selected QC parameters reported by participants 15
Table 7. Results of raw reads submitted by participants and evaluated by EQA provider QC pipeline 16
Table 8. Results of the participants' assessment of the EQA provided genomes 17

Abbreviations

AD	Allelic differences
BN	BioNumerics
cgMLST	Core genome multilocus sequence type
EFSA	European Food Safety Authority
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 eighth round of the external quality assessment (EQA-8) 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-8 contains serotyping, and molecular typing-based cluster analysis.

Human listeriosis is a relatively rare but serious foodborne disease with a European Union (EU) notification rate of 0.46 cases per 100 000 population in 2019 [3]. The number of human listeriosis cases in the EU has increased from 2008 to 2016. During the period 2017−2019, the level of reported cases was stable.

Since 2007, ECDC has been responsible for EU-wide surveillance of listeriosis, including facilitating the detection and investigation of food-borne outbreaks. Surveillance data, including certain basic typing parameters, are reported by European Union/European Economic Area (EU/EEA) countries to the European Surveillance System (TESSy). Since 2012, the EQA scheme has covered molecular typing methods used for EU-wide surveillance.

The effective molecular typing-enhanced surveillance relies on the capacity of NPHRLs in the FWD-Net to produce comparable typing results. ECDC has opened up the possibility for countries to submit WGS data for *L. monocytogenes* to TESSy for to be use in EU-wide surveillance and cross-sectoral comparison. The previous EQA schemes (EQA-1 to EQA-4) included quality assessment of the pulsed-field gel electrophoresis (PFGE) performed for *L. monocytogenes*. Since EQA-5, the PFGE part has been modified to only assess the ability to identify a cluster using the PFGE method and the quality assessment part has been excluded.

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 Europe and to represent a broad range of clinically relevant types of 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 of three laboratories (18%) compared to EQA-5, but the same level of participation as EQA-6 and EQA-7. The majority of participants (11/17; 65%) completed the full EQA scheme.

In total, 13 (76%) laboratories participated in the serotyping part. Molecular serogrouping results were provided by 11 of 13 (85%) participants. Two participants performed both conventional serotyping and molecular serogrouping. The performance of molecular serogrouping was high, with 99% correct results. For the conventional method, 75% of the participants correctly serotyped all test isolates. One participant mistyped eight 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 the EQA-8, 15 (88%) performed molecular typing-based cluster analysis using a method of their choice. The idea of the cluster analysis part of the EQA was to assess the NPHRLs' ability to identify a cluster that was genetically closely related - i.e. to correctly categorise the cluster test isolates regardless of the method used, instead of the ability 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 using less discriminatory methods (e.g. PFGE). Neither of the two participants using PFGE identified the cluster correctly. Fourteen 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 86% (12/14) used core genome multilocus sequence typing (cgMLST), while only 14% (2/14) using single nucleotide polymorphism (SNP). The most widely used scheme for the EQAs was the Ruppitsch (cgMLST), however the Pasteur scheme (cgMLST) was still used by 33% (5/15) of the participants in EQA-8.

In EQA-5 to EQA-8, 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 than SNP analysis. However, the conclusion was not as obvious in EQA-6, EQA-7 or EQA-8, since only a few SNP analyses were reported in these schemes compared with six SNP analyses in EQA-5.

In EQA-8, the EQA provider introduced an additional part to the molecular typing-based cluster analysis: an assessment of five genomes provided for the EQA. The idea was to mimic an urgent outbreak situation, where sequence data may have been produced in other laboratories and the available sequences would have to be addressed despite possibly being poor quality. Almost all participants successfully identified the two high-quality genomes as either a cluster isolate (Fasta file 93%) or a non-cluster isolate (93%). Both poor-quality genomes were identified by 93% of the participants. Only 50% of the participants identified the genome with 15% *L. monocytogenes* ST1 contamination.

1. Introduction

1.1 Background

ECDC is an EU agency with a mandate to operate dedicated surveillance networks. The agency's mission is to identify, assess and communicate current and emerging threats to human health from communicable diseases. ECDC 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 extends such cooperation and supports the implementation of quality assurance schemes [1].

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

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 won the new round of tenders (2017–2020) for all three pathogens. Since 2017, the EQA scheme for *L. monocytogenes* 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-8.

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. Between 2008 and 2016, the number of human listeriosis cases increased in the EU. Since 2016, the number of confirmed human listeriosis cases has stabilised: 2 509 cases in 2016, 2 480 in 2017, 2 549 in 2018 and 2 621 in 2019, corresponding to a notification rate of 0.46−0.48 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 Member States 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 for pathogen characterisation in the surveillance of foodborne infections. Since 2012, ECDC has enhanced EU surveillance by incorporating molecular typing data through isolate-based reporting 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;
- aid the study of a particular pathogen's characteristics and behaviour in a community of hosts.

Molecular typing-enhanced surveillance gives users access to EU-wide molecular typing data for the pathogens included. 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 collaboration.

1.3 Objectives

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

1.3.1 Serotyping

The EQA-8 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-8 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 the opportunity to assess extra genomes, determine whether they were part of the defined cluster and describe their observations and considerations. 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-8 was funded by ECDC and arranged by SSI following ISO/IEC 17043:2010 [4]. EQA-8 included serotyping and molecular typing-based cluster analysis and was carried out during the period August–December 2020.

Invitations were emailed to ECDC contact points in the FWD-Net (26 countries nominated laboratories to participate in the EQA rounds from 2017–2020) by 18 May 2020, with a deadline to respond by 8 June 2020. In addition, an invitation was sent to the EU candidate country Turkey. As many EU countries were still struggling with the COVID-19 pandemic during the early summer of 2020 and laboratory staff were handling COVID-19 samples, the EQA shipment and deadlines were postponed to autumn 2020.

Eighteen public health national reference laboratories in EU/EEA and EU candidate countries accepted the invitation to participate and 17 submitted results (Annex 1). In Annex 2, details of participation in EQA-7 and EQA-8 are listed to give an overview of the trend in the number of participants. The EQA test isolates were sent to participants on 2 September 2020. Participants were asked to submit their results by 1 November 2020 using the online form (Annex 12). If WGS was performed, submission of the raw reads (FASTQ files) was requested. The EQA submission protocol was distributed by email and was available online.

2.2 Selection of test isolates/genomes

Seventeen candidate isolates were analysed using the methods set out in the EQA (serotyping and WGS) before and after re-culturing. All candidate isolates remained stable using these methods and a final selection of 17 test isolates; 13 test isolates and two sets of technical duplicates (same isolate culture twice) was made.

Seventeen *L. monocytogenes* test isolates were selected to fulfil 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-8; and
- include two sets of technical duplicates (Annex 8).

Eleven test isolates for serotyping were selected to cover different serotypes/-groups (1/2a/IIa, 1/2b/IIb, 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) (ST9 and ST580). Five of the serotyping isolates were also included in the set for cluster analysis.

Table 1. Serotype/groups and sequence type of test isolates/genomes

Method	Number of test isolates	Serogroup/ST	Annex
Only serotyping	Six isolates Sero1-Sero6	1/2a/IIa x 1 1/2b/IIb x 2 1/2c/IIc x 1 4b/IVb x 2	3
Both serotyping and cluster analysis	Five isolates (Sero7/REF1-Sero11/REF5)	1/2a/IIc x 5 / ST9 x 4 and ST580 x 1)	3, 6 and 8
Only cluster analysis	Six isolates REF6-REF11 Five genomes REF2*, REF7/REF9*, REF8*, REF11* and REF12	ST9 x 6 ST9 x 5 (modified genomes: contaminated with 15% *Listeria monocytogenes* ST1 and two with reduced coverage)	6, 8 and 11

Technical duplicate sets were REF2/REF6 and REF7/REF9 (Annex 3 and Annex 8)
'Repeat isolates' included in EQA-1 to 8 (Sero3, Sero4 and Sero5). Sero4 was a different isolate to that used in previous years, although it was the same serotype/group.
**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 EQA-1 to 8: Sero3 (1/2c-IIc), Sero4 (1/2a –IIa) and Sero5 (4b - IVb).

Based on the WGS-derived data, the cluster of closely-related isolates consisted of five *L. monocytogenes* ST9 isolates (including the technical duplicate sets REF2/REF6 and REF7/REF9). Characteristics of all the *L. monocytogenes* test isolates are listed in Annexes 4–12. For the additional genomes, three were altered; two with reduced coverage and one mixed with 15% *L. monocytogenes* ST1. The last two were genomes with acceptable quality reads, one provided as a Fasta file.

2.3 Distribution of isolates/genomes

All 17 test isolates were blinded and shipped on 2 September 2020. The protocol for the EQA exercise and a letter stating the unique isolate IDs were included in the packages which were distributed individually to the participants by email on 2 September 2020 as an additional precaution. Twelve participants received their dispatched isolates within one day, five within two days and one received them sixteen days after shipment. The packages were shipped from SSI, labelled 'UN3373 Biological Substance'. No participants reported damage to the shipment or errors in the unique isolate IDs.

On 22 September 2020, 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). The laboratories had the possibility to submit the 7-gene Multi Locus Sequence Types (ST) of isolates in the cluster analysis and were also asked to report the number of loci in the used allelic scheme for cluster analysis and/or the name of the used SNP pipeline.

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

2.5 Data analysis

The submitted serotyping and cluster analysis results, as well as the raw reads, were imported to a dedicated *Listeria* EQA-8 BioNumerics (BN) database. Due to the COVID pandemic, the EQA provider allowed one participant to submit data 14 days after the deadline and an extra participant was allowed to send additional data 45 days after the deadline, however the data was excluded as it only represented a partial analysis.

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 taken from WGS-derived data. Cluster analysis, based on WGS-derived data, was obtained from allele-based analysis (cgMLST [6] and SNP analysis (NASP, [7]). The correct number of closely-related *L. monocytogenes* isolates by WGS was five ST9 isolates: REF2, REF6, REF7, REF8 and REF9 (REF2/REF6 and REF7/REF9 were technical duplicate sets). The EQA provider found at most two allele differences or seven SNPs between any two isolates in the cluster. The remaining six of the cluster test isolates were additional five ST9s and one ST580.

The participant's descriptions of the EQA-provider's manipulated genomes are listed in Annex 11. This analysis was not commented on in the individual reports, but will be summarised in this report.

Individual evaluation reports were distributed to participants in December 2020 and certificates of attendance in March 2021. If WGS data were used, the evaluation report included a quality assessment made by the EQA provider's in-house quality control pipeline (e.g. coverage, N50, sequence length, and number of contigs). The evaluation report did not include an evaluation based on quality thresholds.

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 that signed up, 17 completed and submitted their results. The majority of the participants (71%, 12/17) completed the full EQA scheme. In total, 13 (76%) participants participated in serotyping and 15 (88%) in cluster analysis. Conventional serotyping results were provided by 31% (4/13) of the participants and two of these laboratories also performed molecular serogrouping. Molecular serogrouping results were provided by 11 (85%) participants.

Most participants (87%: 13/15) reported cluster analysis using only WGS-derived data, while one (7%) reported using only PFGE data and one (7%) submitted cluster data based on both PFGE and WGS (Table 1a).

Table 1a. 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	2	9	2	13	1	13	1	15
Percentage of participants	15	69	15	76*	7	87	7	88*

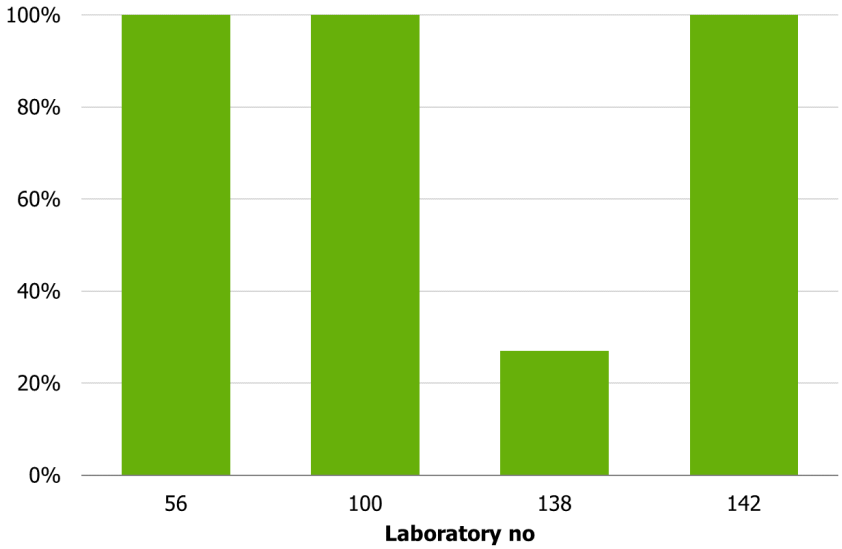
Twelve of the 17 participants (71%) completed both parts (serotyping and cluster analysis) of the EQA.
** Percentage of total number of participating laboratories (17).*

3.2 Serotyping

3.2.1 Conventional serotyping

Four participants performed conventional serotyping of *L. monocytogenes* (Figure 1). Performance was high (82%) as three of the participants correctly serotyped all 11 test isolates. One laboratory (138) had issues with multiple isolates (different serotypes), and only serotyped three of the 11 isolates correctly.

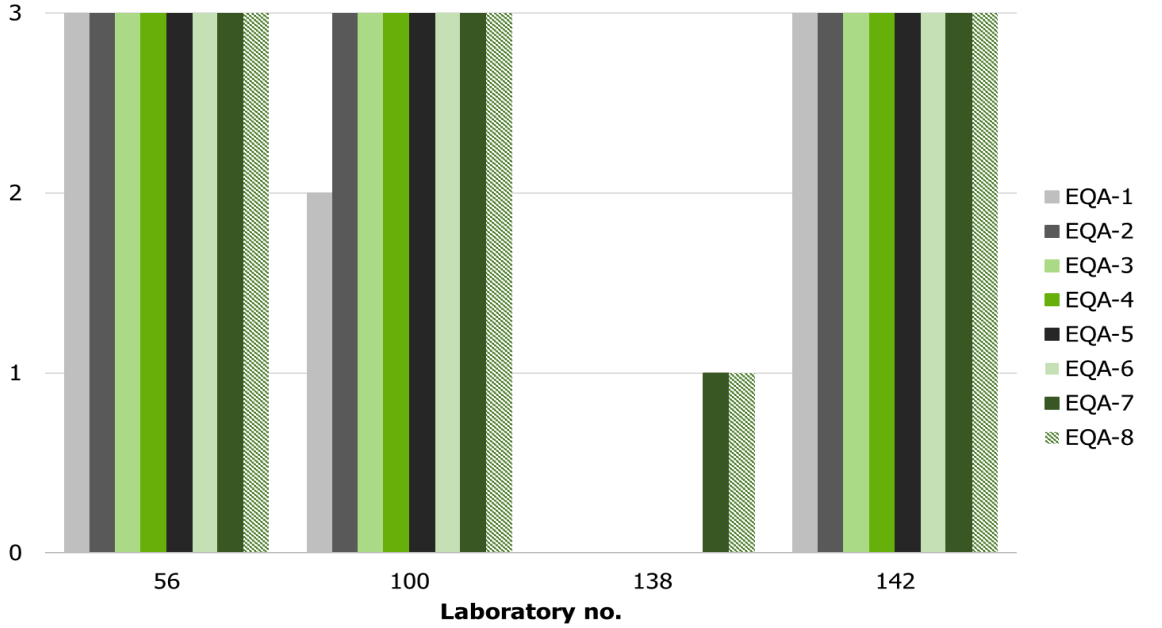
Figure 1. Participant scores for conventional serotyping of 11 test isolates



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 in conventional serotyping of the three 'repeat isolates' from EQA-1 to EQA-8. Only laboratories participating in EQA-8 are shown. The reproducibility of conventional serotyping results of the repeat isolates shows stability and high performance for four of the participants taking part every year (laboratories 56, 100 and 142). However, laboratory 138, participating for the second time, did not report serotyping results correctly for all the repeat isolates in either EQA-7 or EQA-8.

Figure 2. Correct conventional serotyping of three repeat isolates through EQA-1 to 8 for laboratories participating in EQA-8

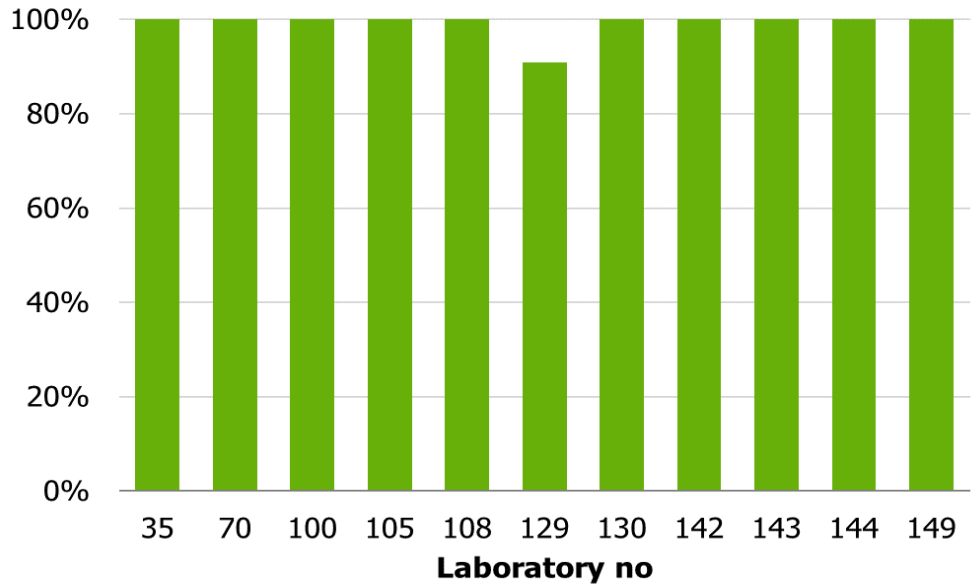


Arbitrary numbers represent the participating laboratories. Bars represent the number of correctly assigned serotypes for the three repeat isolates (Sero3, 4 and 5).

3.2.2 Molecular serogrouping

Eleven participants performed molecular serogrouping of *L. monocytogenes* in EQA-8 (Figure 3). In EQA-5, two new laboratories (96 and 130) participated in molecular serogrouping, but only laboratory 130 continued to participate. Molecular serogrouping was carried out in accordance with guidelines by Doumith et al. [5] and nomenclature from Doumith et al. [8] was used. Ten (91%) participants were able to correctly serogroup all 11 EQA test isolates. Four of the 11 participants reported using WGS-based analysis (*in silico* PCR) for molecular serogrouping. In addition to the results shown below an additional three laboratories (49, 56 and 141) submitted molecular serogroup results for the isolates Sero7–Sero11 as they also were a part of the cluster test set. However, the partial results were excluded from this report.

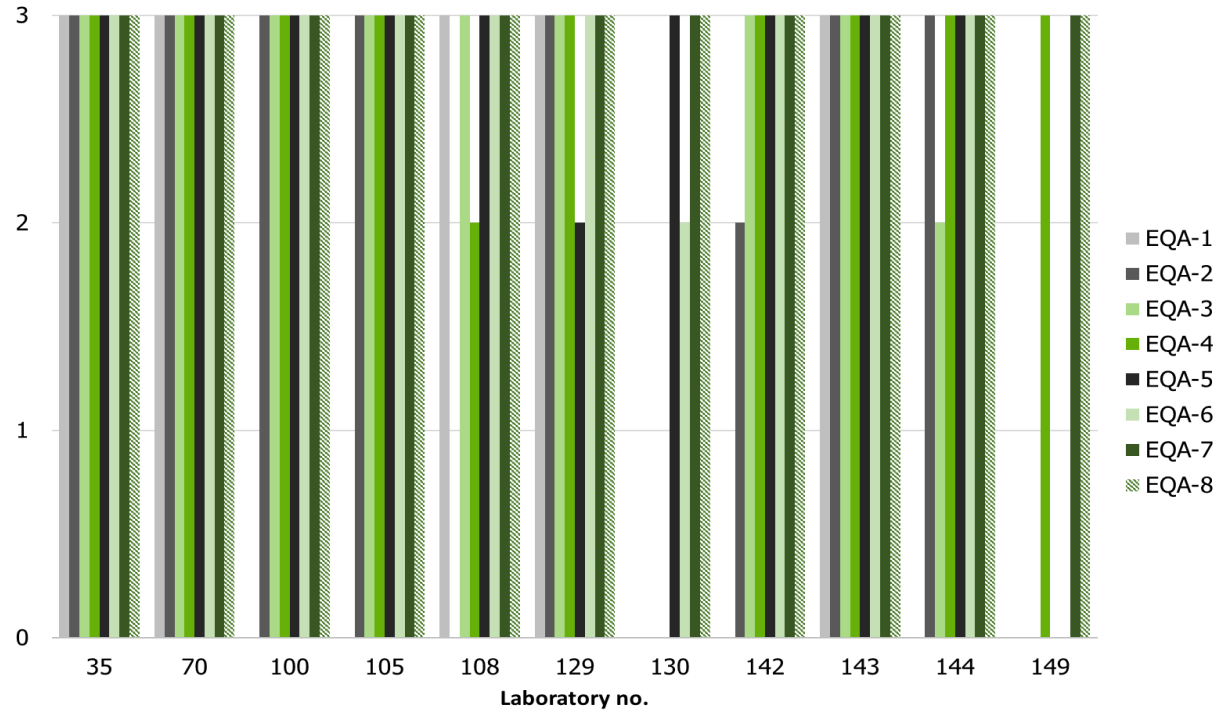
Figure 3. Participant scores for molecular serogrouping of 11 *L. monocytogenes* test isolates



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 in molecular serogrouping when assessing the three repeat isolates during the eight EQAs. Of the 11 laboratories that participated in EQA-8, 6/11 (55%) correctly serogrouped all three repeat isolates in all the EQA rounds they participated in.

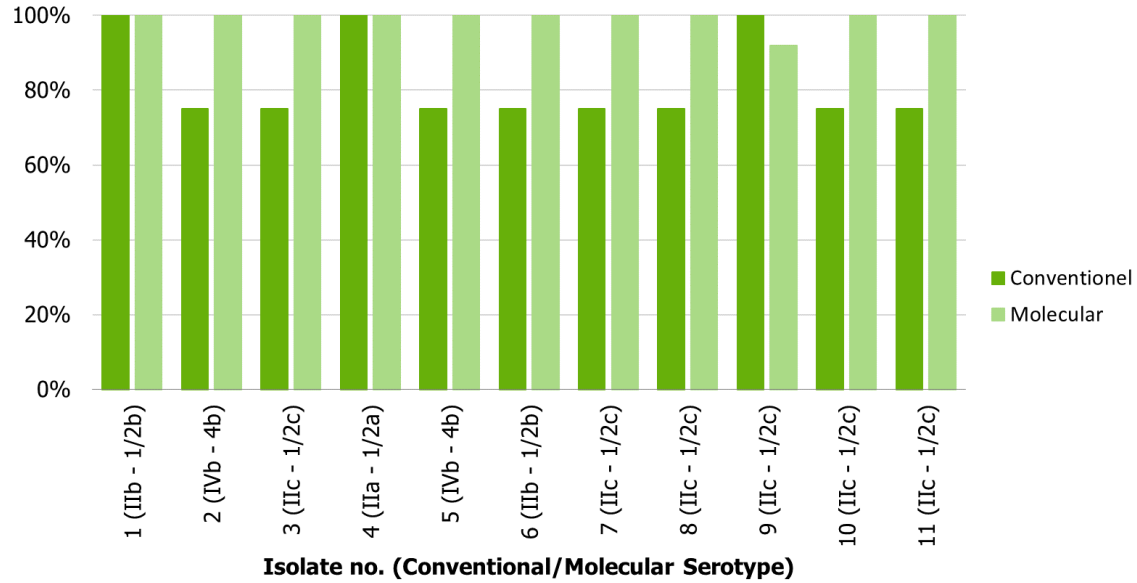
Figure 4. Correct molecular serogrouping of three repeat isolates from EQA-1 to 8 (for laboratories participating in EQA-8)



Arbitrary numbers represent the participating laboratories. Bars represent the number of correctly assigned serogroups for the three repeat isolates (Sero3, 4 and 5).

Figure 5 shows the reported error distributed per isolate. No isolate had more than one laboratory reporting an error. All the errors seen in the conventional serotyping were reported by one laboratory 138 and laboratory 129 reported one error for isolate sero9 in the molecular serotyping. No more than one error was seen per isolate.

Figure 5. Average score of 11 test isolates



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

3.3 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 was pre-categorised by the EQA provider. The expected cluster of closely related *Listeria monocytogenes* ST9 isolates contained five 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 ST9 isolates: REF2, REF6, REF7, REF8, and REF9 (REF2/REF6 and REF7/REF9 were technical duplicates). The EQA provider found at most two allele differences or seven SNPs between any two isolates in the cluster. The rest of the cluster test isolates were additional five ST9s and one ST580 (Annexes 6, 8).

3.3.1 PFGE-derived data

Two (2/17, 12%) 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 using a less discriminatory method. Both participants identified the five isolates as belonging to the correct WGS cluster and an additional one (REF4) or two isolates (REF4 and REF5). One additional laboratory (130) also submitted PFGE data of only five isolates which were in both the serotyping set and the cluster analysis set: Sero7-Sero11=REF1-REF5. The participant did not 'order' the cluster test isolates when signing up for the EQA. Partial data is not accepted and therefore not included in the report.

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

Isolates		Laboratory ID	
Isolate number	**ST**	**138**	**142**
REF1	ST9	No	No
REF2‡#1	ST9	Yes	Yes
REF3	ST580	No	No
REF4	ST9	Yes	Yes
REF5	ST9	Yes	No
REF6‡#1	ST9	Yes	Yes
REF7‡#2	ST9	Yes	Yes
REF8‡	ST9	Yes	Yes
REF9‡#2	ST9	Yes	Yes
REF10	ST9	No	No
REF11	ST9	No	No
Correct cluster identified*		**No**	**No**
Included the five WGS cluster isolates		**Yes**	**Yes**

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

3.3.2 WGS-derived data

Reported results from participants

Fourteen participants (14/17, 82%) performed cluster analysis using WGS-derived data. All laboratories reported using their own laboratory for sequencing. Different sequencing platforms were listed among the participants: one MiniSeq, seven MiSeq, four NextSeq, one Ion GeneStudio S5 System and one Ion Torrent. All reported using commercial kits for library preparation. Of the 14 participants, eleven (79%) used Illumina's Nextera kit. One participant listed less time for shearing and volume changes than that stated in the manufactory protocol (Annex 5).

Performance was very 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	49	56	70	100	105	108	129	135	141	142	144	149
REF1	ST9	No	No	No	No	No	No	No	No	No	No	No	No	No	No
REF2‡#1	ST9	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
REF3	ST580	No	No	No	No	No	No	No	No	No	No	No	No	No	No
REF4	ST9	No	No	No	No	No	No	No	No	No	No	No	No	No	No
REF5	ST9	No	No	No	No	No	No	No	No	No	No	No	No	No	No
REF6‡#1	ST9	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
REF7‡#2	ST9	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
REF8‡	ST9	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
REF9‡#2	ST9	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
REF10	ST9	No	No	No	No	No	No	No	No	No	No	No	No	No	No
REF11	ST9	No	No	No	No	No	No	No	No	No	No	No	No	No	No
Main analysis		Allele	Allele	Allele	Allele	Allele	SNP	Allele	SNP	Allele	Allele	Allele	Allele	Allele	Allele
Additional analysis 1		SNP		Allele											
Additional analysis 2															
Cluster- identified		Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

‡: closely-related isolates (in grey)
#: technical duplicate isolates (in bold)
ST: 7 multilocus sequence type
Allele: allele-based analysis (cgMLST)
SNP: single-nucleotide polymorphism (Annex 7).

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 49).

Of the 14 participants, two (14%) used SNP as the main analysis, one used an in-house pipeline and one a published pipeline. One used a reference-based approach with an ST35 isolate as reference. The other laboratory used an assembly-based approach, with CLC for both read mapper and variant caller, and the other used Burrows-Wheeler Aligner (BWA) as the read mapper.

Tables 4 and 5 show the overview of the submitted data. Each laboratory reported SNP distances/allelic differences by isolate (see Annex 9).

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 (ST9)	BWA	GATK	0–7	48-299 (109)
19*	NASP	Reference based	ST9 and isolate ID 1114	BWA	GATK	0-8	34-178 (95)
100	SNV cgMLST SeqSphere	Reference based	NC_003210 (ST35)	BWA	-	0-3	10-65 (54)
108	In-house pipeline	Assembly based	-	CLC Assembly Cell v.5.2	CLC Assembly Cell v.5.2	0–12	212-2674 (1156)

¤: reported distance to ST9 isolates (non-ST9)
**: additional analysis*
For detailed data see Annex 8.

Twelve of the 14 participants used allele-based analysis as the main analysis for cluster detection (Table 5). Seven (58%) only used an assembly-based allele calling method, three used both mapping and assembly-based allele calling and two used only a mapping based allele calling method (Table 5). All 12 reported using cgMLST, seven (58%) used cgMLST Ruppitsch (1701 loci) [9], four cgMLST Pasteur (1748 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	12–59 (46)
19	BioNumerics	Assembly based and mapping based	SPAdes	Applied Math (cgMLST/Pasteur)	1748	0-2	13-60 (45)
35	SeqPhere	Only assembly based	Spades, Velvet	Ruppitsch (cgMLST)	1701	0–3	11-58 (45)
49	BioNumerics	Assembly based and mapping based	SPAdes	Applied Math (cgMLST/Pasteur)	1748	0-2	13-61 (46)
49*	BioNumerics	Assembly based and mapping based	SPAdes	4804	4804	0-7	26-131 (105)
56	MentaLiST	Only mapping based	-	Pasteur (cgMLST)	1748	0–3	10-50 (24)
70	SeqPhere	Assembly based and mapping based	SKESA	Ruppitsch (cgMLST)	1701	0–3	11-59 (46)
105	SeqPhere	Only assembly based	SPAdes	Ruppitsch (cgMLST)	1701	0–3	11-59 (46)
129	SeqPhere	Only assembly based	Velvet	§ The Ridom SeqSphere+ software's Target Definer cgMLST.	1503	0–5	11-47 (36)
135	SeqPhere	Only mapping based	-	Ruppitsch (cgMLST)	1701	0–3	11-59 (46)
141	SeqPhere	Only assembly based	SPAdes 3.11.1	Ruppitsch (cgMLST)	1701	0–4	11-60 (48)
142	BIGSdb-Lm	Only assembly based	SPAdes	Pasteur (cgMLST)	1748	0–3	13-61 (48)
144	SeqPhere	Only assembly based	Velvet	Ruppitsch (cgMLST)	1701	0–3	11-59 (46)
149	SeqPhere	Only assembly based	Velvet	Ruppitsch (cgMLST)	1701	0–4	12-60 (48)

§: modified from submitted information
¤: reported distance to ST9 isolates (non-ST9)
**: additional analysis*
(Annex 8).

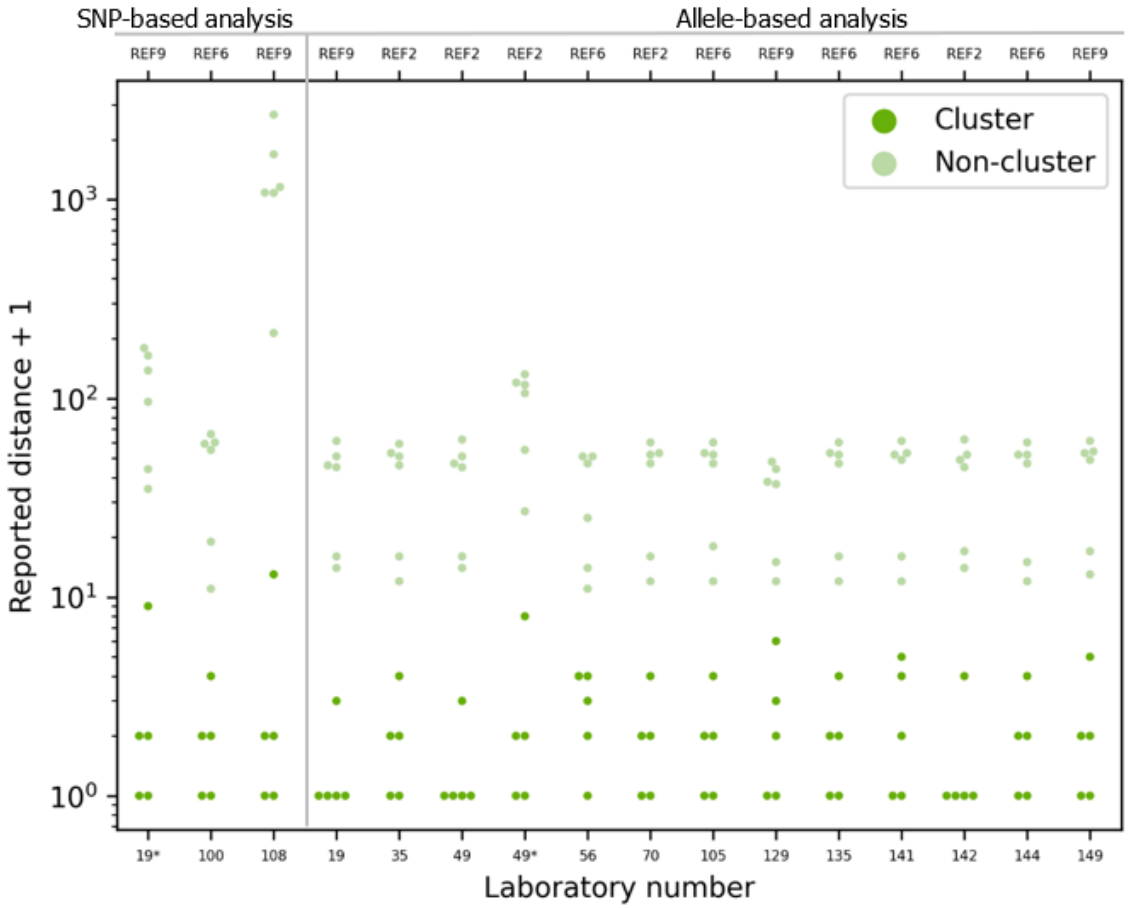
Laboratories 100 and 108 were the only two participants to perform SNP analysis as the main analysis and identify the correct cluster of closely-related isolates. Laboratory 19 performed SNP analysis as an additional analysis and identified the correct cluster of closely-related isolates by cgMLST (main analysis). The reported SNP distances ranged from 0−3 to 0−12 within the cluster. For the test isolates outside the cluster, the laboratories reported an SNP distance to the selected cluster isolates at 10-2674. A clear separation was obtained by all the participants, however, the smallest distance was reported by laboratory 100, using a single-nucleotide variant based on a cgMLST (SNV) approach.

All 12 (100%) laboratories reported the correct cluster and up to five allele differences within the correct cluster (Table 5/Figure 6). Two laboratories reported 0−2 allele differences, seven laboratories reported 0−3 allele differences, two laboratories reported 0−4 allelic differences, and one laboratory 0−5 allelic differences inside the cluster. Ion Torrent data was used by laboratory 56 and was comparable to the results of other platforms.

Most of the laboratories (9/12) selected the technical duplicates pair (REF2/REF6) as the cluster representative in the main analysis (Figure67).

Five of the test isolates (REF1, REF4, REF5, REF10 and REF11) were also ST9, but not predefined by the EQA-provider as part of the cluster. Based on cgMLST, the twelve laboratories reported allele differences to the selected cluster isolate at 10–61 for this group of isolates. One test isolate (REF3) was ST580 and allele differences were reported to the selected cluster isolate at 24-48. (Annex 8).

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



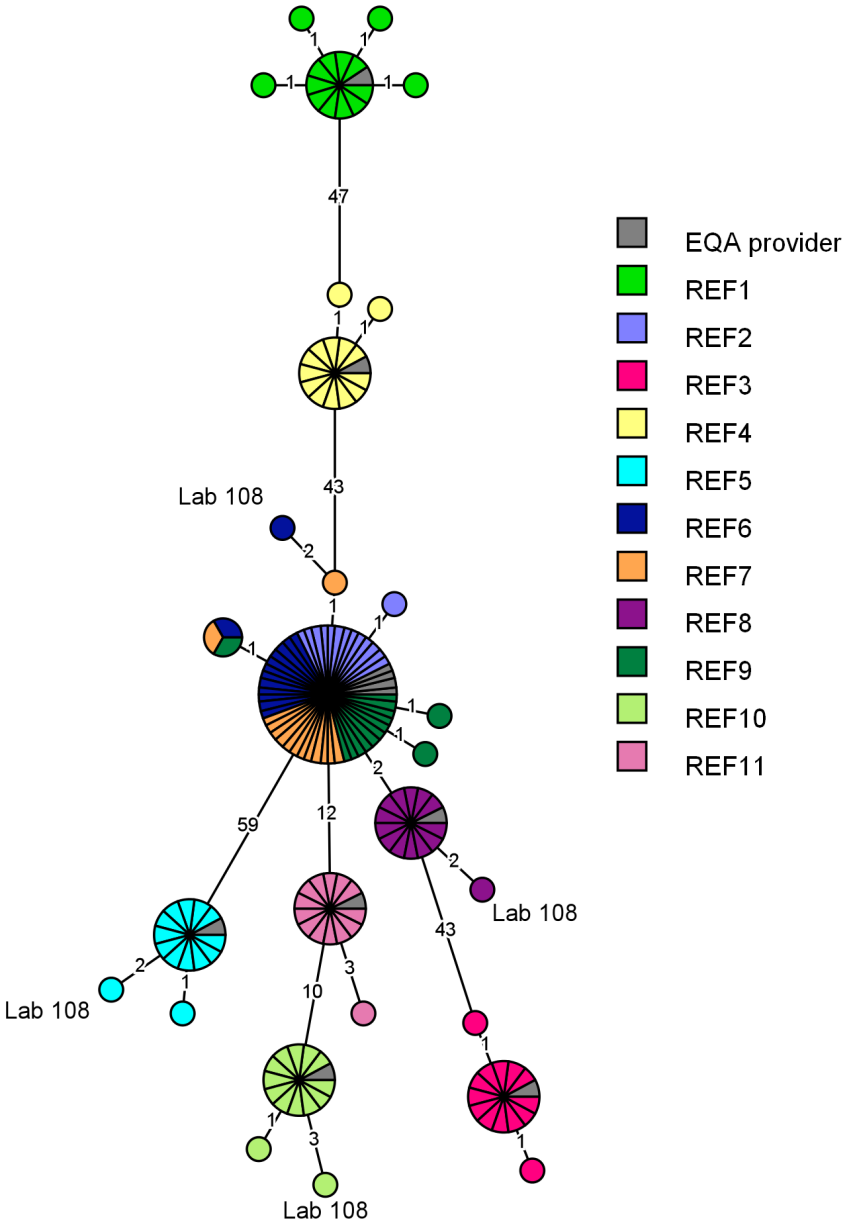
SNP: single nucleotide polymorphism
Selected cluster representative marked as REF.
Dark green: reported cluster of closely related isolates
Light green: not reported as part of cluster.

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 [10].

The overall cgMLST analysis, shown in the minimum spanning tree (MST), based on submitted raw reads (FASTQ files) from 14 laboratories, reveals clear clustering of the results for each test isolate (Figure 7). Four of the data notes from laboratory 108 are separated with 2–3 AD from the other results.

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



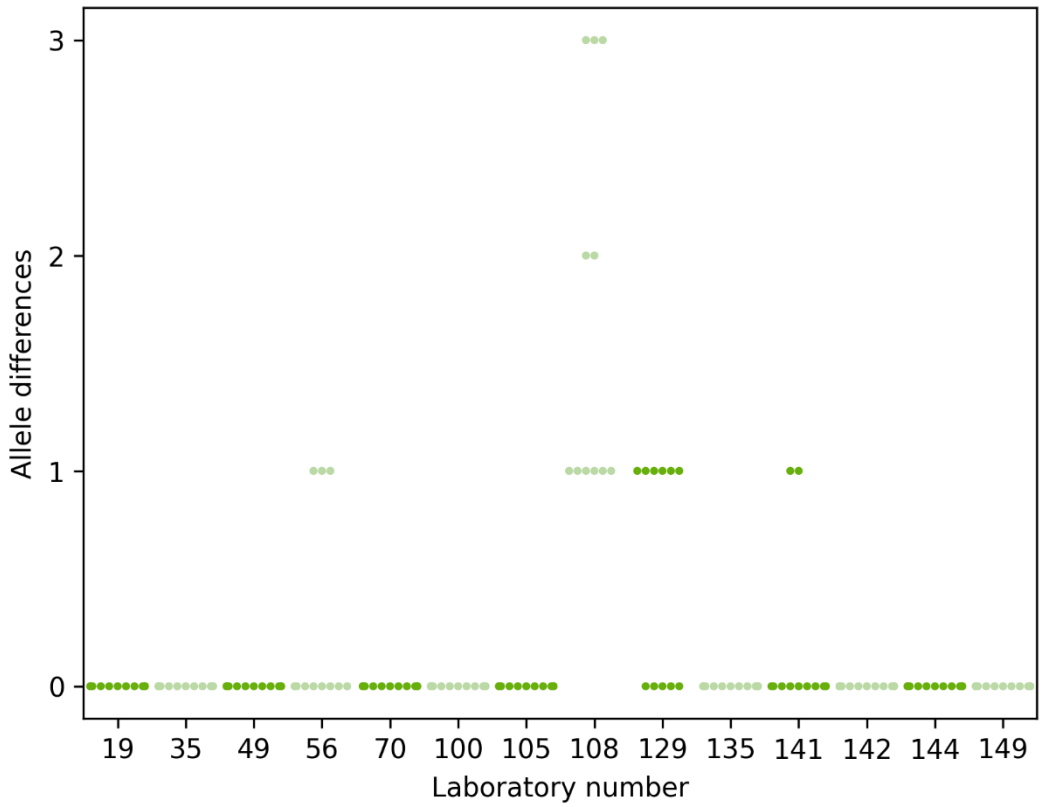
Minimum spanning tree (MST) in log scale of core genome multi-locus sequence typing (cgMLST) [6] based on submitted raw reads (FASTQ files). Each of the REF1–11 test isolates have a different colour. REF results from the EQA-provider are in grey.

The allele differences in Figure 8 do not exactly match those illustrated in the individual reports and consequently those in Figures 9, as all are based on the same data. This discrepancy is caused by loci being dropped if they do not pass quality control for all isolates in the analysis. Joint analysis therefore contains fewer loci.

For each laboratory, cgMLST was performed on the raw reads submitted (FASTQ files), applying an 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. As seen in Figure 8 and 9, all laboratories have small differences to the reference isolates. Laboratory 108 had most sequences different to those of the EQA-provider.

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 132 of 154 results (86%), no difference was identified. As seen in Figure 8, for 17 results (11%), a difference of one allele from the REF isolate was calculated and a difference of 2–3 alleles was seen for five results (3%). Results from Laboratory 108 showed allele difference for all 11 isolates. The difference is for the results of Ion Torrent data analysed in BioNumerics, however Laboratory 56 also used Ion Torrent data although it had a smaller number of differences.

The laboratories responded to QC parameters used to evaluate their data separately. Both coverage and confirmation of the genus were the most widely used QC parameters, with 93% of the laboratories using them (Table 6). Participants used different thresholds of coverage ranging from 10–50 x coverage. The laboratories reported the different programmes used for the contamination check of the genus. The number of good cgMLST loci was used as a QC parameter by 86% of the participants, with thresholds ranging from 89–98%. Q score and genomic size were used by 43% and 71% of the participants respectively. Additional QC parameters are listed in Annex 9 by some of the participants.

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 other genus	Min. 50x	-	2.8 - 3.1Mb	Min. 95% core % max.15 loci with multiple consensus. No actual threshold employed on regular basis for either.
35	SeqSphere	>30-fold	-	-	>95%
49	-	<20 = Fail 20-29 = Warn >=30 = Pass	>=30 = Pass	2.7 - 3.2MB = Pass	Core genome %
56	No mismatches in the alignment with the seven housekeeping genes	50	22	2.8 Mb	-
70	Mach in SeqSphere	50x If less, the number of good targets should be > 95%.	-	Length of contigs assembled < reference genome + 10%	cgMLST alleles found and called > 95%
100	KmerFinder 3.2 CGE	40x	-	2.9 Mb	Percentage of good targets < 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 four nucleotides. Sequences with a length <70 bases were also removed.	<=3.3Mb	>=95%
108	Assembled genomes aligned against *L. monocytogenes*	20x	-	+/- 20% of RefSeq reference strain	-
129	Presence of prfA gene (LIP) and occasionally Kraken.	>29	-	-	>89
135	Assessing the genome completeness with reference markers for Listeria using the open source tool CheckM	>10	>30	2.7-3.23 Mb	>90%
141	J Species, cge tools, SeqSphere	30x	-	~2.8-3.0 Mb	Min. 98% good targets
142	Kraken: ID check + contamination check (warning >1%, failure >5%)	Min. 30x	Min. 28.0	-	Warning <95% calls Failure <90% calls
144	https://cge.cbs.dtu.dk/services/SpeciesFinder/	50x	>30	2.8-3	>98%
149	KRAKEN	-	-	-	>90% good targets
Percentage of laboratories using QC parameter	**93%**	**93%**	**43%**	**71%**	**86%**

For each laboratory, the raw reads submitted (FASTQ files) were evaluated by the EQA provider's in-house quality control pipeline [10]. Table 7 shows the QC parameters and range of QC values per laboratory. For the full QC evaluation of all isolates, see Annex 10.

According to the QC parameters, sequencing quality was uniformly good. Overall, coverage was high. One laboratory (70) had a *Burkholderia dolosa* contamination in one of the eleven isolates (Annex 10).

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

		Laboratory ID													
Parameters	Ranges*	19	35¤	49	56	70	100	105	108	129	135	141	142	144	149
Detected species	{Lm}	*Lm*	*Lm*	*Lm*	*Lm*	*Lm, Bd*	*Lm*	*Lm*	*Lm*	*Lm*	*Lm*	*Lm*	*Lm*	*Lm*	*Lm*
% Species 1		91.9-95.5	65.3-97.2	96.7-98.8	94.2-98.4	91.5-97.4	95.9-99.2	96.0-97.0	89.0-98.7	95.3-98.5	94.0-98.1	90.7-96.5	96.4-98.8	94.5-98.5	85.9-94.9
% Species 2	{<5%}	0.1-0.2	0.0-0.5	0.0-0.1	0.0-0.1	0.2-6.5	0.0-0.4	0.0-0.0	0.0-0.3	0.0 - 0.1	0.3-1.4	0.0-0.3	0.1-0.3	0.0-0.1	0.4-4.9
Unclassified reads (%)		4.3-7.9	2.7-34.2	1.1-3.0	1.5-5.6	1.6-6.4	0.9-3.9	2.9-3.9	1.3-10.7	1.4-4.6	1.3-5.5	3.4-8.8	0.9-3.1	1.5-5.4	3.4-10.7
Length at >25 x min. coverage (Mbp)	{>28 ∧ <31}	2.9-3.0	2.9-3.0	2.9-3.0	2.8-3.0	2.9-3.0	0.0-3.0	2.9-3.0	2.5-2.8	2.7-3.0	2.9-3.0	2.9-3.0	2.9-3.0	2.9-3.0	2.9-3.0
Length [1-25] x min. coverage (kbp)	{<250}	0.0-5.0	0.0-53	0.0	0.0-0.0	0.0-0.0	0.0-2783	0.0	0.0-6.0	0.0-271	0.0	0.9-119	0.0-11	0.0	0.0-0.6
No. of contigs at 25 x min. coverage	{>0}	23-48	28-41	12-28	#250-885	19-31	6-32	15-24	#2253-3522	17-101	13-18	39-150	12-50	14-21	16-28
No. of contigs [1-25] x min. coverage	{<1000}	0-4	0-15	0	#0	0	0-557	0	#0-23	0-61	0	2-58	0-11	0	0-1
Average coverage	{>50}	76-134	77-209	59-174	200-296	35-74	15-187	97-433	58-92	28-207	154-460	31-81	48-86	57-94	81-237
No. of reads (x 1000)		1641-2860	1735-4388	622-2072	2240-3455	789-1635	193-2489	1958-8859	597-971	604-4230	3117-9238	435-955	581-1073	1135-1896	944-5164
Average read length		139-142	129-144	232-287	259-279	151-151	223-238	148-150	266-293	141-149	151-151	217-276	237-247	147-149	151-270
Average insert size		235-260	0-251	242-448	NA	163-271	274-334	330-443	NA	315-421	336-415	250-378	361-576	317-383	248-343
N50 (kbp)		124-361	192-476	477-581	#5-24	435-581	18-581	476-564	#1-2	34-435	477-564	30-191	147-605	476-582	477-581

**: indicative QC range*
Lm: L. monocytogenes
Bd: Burkholderia dolosa
NA: not available
#: QC values unreliable due to assembly issues for Ion Torrent data
¤: single end reads (Annex 10).

Assessment of the EQA provided genomes

All participants were invited to perform an additional assessment of the cluster analysis. Five genomes were provided with a request to check if they were related to the cluster, thereby mimicking an urgent outbreak situation where it is impossible to rerun the sequence and the sequences must be assessed despite poor quality, etc.

The five genomes provided had to be individually assessed and compared with the data already produced in the cluster analysis. The participants had to determine whether the genomes were part of the defined cluster.

The participants were instructed to describe their observations and considerations leading to the decision. The EQA-provider had manipulated some of the raw reads: the five genomes were a mix of cluster isolates with contamination (one isolate), reduced coverage (two isolates), one which was an assembly (Fasta file), and a non-cluster isolate with good quality (Table 8).

For genome 1, (a cluster isolate with altered coverage (15x)), 93% (13/14) correctly observed poor quality of the genome and nine laboratories applied average coverage, low N50, below threshold of cgMLST targets or low number of contigs to disregard the genome and did not perform a cluster analysis. Of the five laboratories that performed the cluster analysis three correctly suggested that the isolate was a cluster isolate and two suggested it was a non-cluster isolate.

For genome 2, (a cluster isolate, good quality, assembly Fasta file), 93% (13/14) of the laboratories were able to use the Fasta file in their analysis, and all 13 correctly identified genome 2 as part of a cluster of closely- related isolates. One laboratory could not use the Fasta file in its analysis.

For genome 3, (a cluster isolate (REF11) with altered coverage (10x)), 93% (13/14) correctly observed poor quality of the genome, 11 applied average coverage, low N50, below threshold of cgMLST targets or low number of contigs to disregard the genome and did not perform a cluster analysis. Only one of the 14 laboratories did not describe any detection of low quality and reported the genome as a non-cluster isolate. In addition, two other laboratories also performed cluster analysis and suggested that the genome was a non-cluster isolate.

For genome 4, (a non-cluster isolate, good quality of reads), 100% (all 14 laboratories) performed the cluster analysis and seven specified the genome was of good quality. Of the 14 laboratories 93% (13/14) correctly identified genome 4 as a non-cluster isolate.

For genome 5, (a cluster isolate, good quality but contaminated with approximately 15% *L. monocytogenes* ST1), 50% (7/14) correctly described contamination present in genome 5. The suspicion of contamination was for six of the seven, based on the size of the genome (as Kraken or other programmes did not identify an additional genome). Seven of the 14 laboratories correctly reported genome 5 as a cluster isolate, four of them without identifying the contamination. A further four laboratories analysed genome 5 and reported the genome as a non-cluster isolate, three of them did not identify the contamination.

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 (REF2) with altered coverage (reduced to 15x)	Poor quality was observed	13	1	0
		Suggested to be a cluster isolate	3	2	9
2	A cluster isolate (REF8), good quality of reads assembled with SKESA to a FASTA file. Two allelic difference to the REF2 in the cluster.	Fasta file was analysed	13	0	1
		Suggested to be a cluster isolate	13	0	1
3	A cluster isolate (REF11) with altered coverage (reduced to 10x).	Poor quality was observed	13	1	0
		Suggested to be a cluster isolate	0	3	11
4	A non-cluster isolate (REF12), good quality of reads, with 58 allelic difference to the cluster isolate (REF2).	Quality accepted	14	0	0
		Suggested to be a cluster isolate	1	13	0
5	A cluster isolate (REF7/REF9) mixed with a *L. monocytogenes* ST1, (approx. 15%) same species contamination.	Contamination was observed	7	7	0
		Suggested to be a cluster isolate.	7	4	3

Raw data available in Annex 11.

4. Discussion

4.1 Serotyping/-grouping

Thirteen (76%) laboratories participated in the serotyping part of EQA-8 and of these 11 (85%) 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-8, highlighting the transition towards the use of molecular serogrouping. Comparing the conventional serotyping results from EQA-1 to 8, the results showed stable and high performances among most of the participants (EQA-1 to 7: 94%; 87%; 91%; 97%; 98%; 100%; 89%; 82%). One laboratory had multiple issues with all the different serotypes in both EQA-7 and EQA-8.

4.1.2 Molecular serogrouping

Since EQA-2, the number of participants in the molecular serogrouping has ranged between 13 and 17 participants. From EQA-7 to EQA-8, the number of participants decreased from 13 to 11. From EQA-6 to EQA-8, three laboratories reported the use of in silico PCR (WGS) serogrouping. However, an additional three laboratories performed WGS only on the isolates overlapping between the serotyping and cluster isolates Sero7-Sero11. The results were correct but are not included in this report as the data is only partial. In addition, the cluster part shows that most of the participants still use PCR for serogrouping, despite being able to perform WGS. This is probably due to cost, as some participants have suggested reducing the number of isolates for the serotyping/grouping part. With regard to molecular serogrouping, the performance was very good in EQA-8, with a score of 99% correct. Over the years from EQA-1 to EQA-7, the general performance among the participating laboratories has been high: 98%; 94%; 94%; 94%; 99%; 97%; 100%; 99%.

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

4.2 Molecular typing-based cluster analysis

In EQA-5 to EQA-8, PFGE was no longer an independent part, but was added as a possible method of choice for cluster identification. The EQA scheme is contemporary with the development of surveillance methods used by NPHRLs in the EU/EEA. The adjustment of the EQA appears to be well accepted by the countries, as 16 of the 17 laboratories (94%) participated in the cluster analysis. Only one laboratory 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, two (13%) 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 using a less discriminatory method. None of the participants only identified the five cluster isolates (defined by WGS), including one or two more isolates instead. Both laboratories included REF4 and one laboratory also included REF5. Laboratory 138 only performed PFGE and Laboratory 142 also performed cluster analysis on WGS-derived data (with the 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 and one in EQA-8, while 82% (14/17) submitted analysis based on WGS-derived data.

4.2.2 WGS-derived data

Fourteen of the 17 laboratories (82%) performed cluster analysis using WGS-derived data. Overall, there has been increased participation since the cluster analysis part was introduced. In EQA-5, 12 laboratories participated in WGS-based cluster analysis. In EQA-6, one laboratory stopped and a new one started. In EQA-7, all participants from EQA-6 continued and one new laboratory started. In EQA-8, again one new laboratory started using WGS. For the first time since EQA-5, all laboratories reported that the sequencing was done at their own premises. The majority (12/14) also reported using an Illumina platform. All reported using commercial kits for library preparation. Performance was very high, with all laboratories (100%) correctly identifying the cluster of closely-related isolates.

Twelve laboratories (86%) reported using an allele-based method as the main analysis and two laboratories (14%) reported using SNP analysis. Compared to EQA-6 (75%) and EQA-7 (85%), this is a very small percentage increase in the use of allele-based analyses as the main analysis. During the EQAs, both Laboratory 56 and 105 changed the main analysis from SNP to allele-based analysis, and laboratory 100 changed from allele based to SNP (SNV approach).

During EQA-5 to EQA-8, the use of the Ruppitsch cgMLST scheme was the most dominant, however in EQA-8 a small decrease was observed, with seven (58%) of the twelve laboratories using the scheme for the main analysis. The ADs reported were very comparable, despite using different schemes with different numbers of loci. The laboratories using the Pasteur scheme with 1 748 loci reported 0−2/0−3 AD and the laboratories using the Ruppitsch scheme reported 0−3/0−4 AD. Laboratory 129 was the only laboratory with a higher number of AD (0−5) and they used an ad hoc scheme of 1 503 loci. Neither the analysis (assembly and/or mapping based) or the specific assembler seemed to influence the results in this EQA.

The EQA provider's analysis of the raw data submitted showed a very high concordance, when using the standardised cgMLST/Enterobase analysis, as 100% had under three AD differences between the sequences of the laboratories and the EQA provider's sequences, Figure 9). As also shown in previous reports, the analysis by the EQA provider (using the standardised cgMLST/Pasture analysis) of the data from Laboratory 108 showed AD for all the test isolates. This laboratory provided Ion Torrent data for which the EQA provider's analysis is not optimised, making correct assembly difficult (also seen in the previous EQAs for Laboratory 108, however not for Laboratory 56 which also provided Ion Torrent data). Therefore, the observed AD may be method artefacts, however the use of Ion Torrent data can complicate the communication and investigation of multi country outbreaks when only using the allelic method.

The laboratories performing SNP analysis reported a clear separation of the cluster and the non-cluster isolates, despite some variation in the distance outside the cluster being observed (Figure 7). In particular, Laboratory 100 reported a smaller distance between the cluster isolate (REF 8) and non-cluster isolate REF 11, probably explained by the approach of SNV cgMLST Seqphere.

As seen in previous EQAs and in the publication by Henri et al. 2017 [11], the two approaches for analysing WGS-derived data (allele- and SNP-based analysis) show comparable results. This year however, some variation was seen in the SNP analysis, mainly in the results outside the cluster.

The main QC parameters reported in EQA-8 were a threshold of coverage and the checking of genus/species confirmation. Both an assessment of the genome size and the number of cgMLST allele called were essential for the end use of the data.

In the assessment of the additional genomes provided as part of the EQA, most of the participants successfully identified the poor quality in the two modified genomes with reduced coverage. The contamination of the genome with a different *L. monocytogenes* (ST1) was only detected by 50% (7/14), showing that such contamination is challenging to identify using an allele-based method if only looking at the number of good cgMLST loci. In addition, 93% of the participants correctly concluded that genome 2 (Fasta file) was part of the cluster and that the non-cluster genome 4 was not part of the cluster of closely-related isolates. The same laboratory did not detect the low coverage in genome 1 (x15) and genome 3 (x10) or the contamination in genome 5 and they reported the non-cluster isolate (with good quality of reads) as a cluster isolate for genome 4. Unfortunately, although noting the contamination or the low coverage (genome 1 and genome 3), most of the participants did not proceed to assess whether the data could give some useful information. Most responded that they would need to rerun the sequence, and therefore answered no to the question 'Is this genome a part of the cluster?' which the EQA-provider recoded as 'not analysed'.

5. Conclusions

Seventeen laboratories participated in the EQA-8 scheme, with 13 (76%) performing serotyping and 15 (88%) cluster identification.

Most laboratories (69%, 9/13) performed only molecular serogrouping, while 15% (two) performed molecular serogrouping in combination with conventional serotyping and only 15% (2) performed conventional serotyping alone. In general, the trend towards substituting conventional serotyping with molecular was observed through the eight EQAs. The average quality of conventional serotyping (82%) was just below the range of the previous EQAs. The performance for molecular serogrouping was good, achieving 99% in EQA-8. The general conclusion is that serogrouping with molecular typing achieves the best performance. The switch from conventional serotyping to molecular serogrouping has reached a level where molecular serogrouping can be seen as the best practice in NPHRLs across the EU/EEA.

Two laboratories used PFGE for cluster analysis, one of them using only 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 using less discriminatory methods, such as PFGE. Both participants included more isolates than expected in the cluster for WGS.

Fourteen laboratories performed cluster analysis using WGS-derived data. The performance was very good, with all (100%) of the participants correctly identifying the cluster of closely-related isolates.

In the WGS, an allele-based method was preferred, as 86% (12/14) used cgMLST, compared to 14% (2/14) using SNP as the main method reported for cluster analysis. In general, the reported cgMLST results were at a comparable level of allelic difference (0–5) within the cluster isolates, despite being analysed using different schemes. Similar results, with a slightly larger variation, were obtained by the two laboratories that reported the correct cluster using SNP analysis (0–12 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 good-quality genomes, most participants were able to correctly conclude whether 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 whether the data would suggest the genome was a part of the cluster. Nevertheless, 93% identified the low coverage genomes.

The current EQA scheme for *L. monocytogenes* typing is the eighth EQA organised for laboratories in the 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. For two years, the public health institutes have had the possibility to submit WGS variables for *L. monocytogenes* to TESSy to be used for EU-wide surveillance and cross-sectoral comparison. In addition, public health institutes can also send DNA samples to ECDC's contractor for WGS analysis.

6. Recommendations

6.1 Laboratories

It is recommended that laboratories use EQA-provided data and isolates to validate their analysis methods if incorrect results (e.g. EQA) are obtained or if implementing new methods and procedures.

When laboratories re-name 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 such as EQA schemes, expert exchange visits and workshops.

ECDC will investigate the possibility of doing a validation study of the most commonly used sequencing technologies at NPHRLs.

6.3 EQA provider

The evaluation of the genome sequences provided was a success, with almost all of the participants performing the analysis and identifying the modifications introduced by the EQA provider. It is suggested that this part can be expanded further in future EQAs.

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 from: 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 from: 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 2019 Zoonoses Report. EFSA J. 2021;19(2):640 Available from: https://efsa.onlinelibrary.wiley.com/doi/epdf/10.2903/j.efsa.2021.6406
4. International Organization for Standardization (ISO). ISO/IEC 17043:2010 – Conformity assessment -- General requirements for proficiency testing. Geneva: ISO; 2015. Available from: 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 from: 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 from: http://www.microbiologyresearch.org/content/journal/mgen/10.1099/mgen.0.000074
9. 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.
10. Statens Serum Institut. Serum QC "Bifrost" [Internet, software package]. Copenhagen: SSI; 2020 [cited 24 January 2020]. Available from: https://github.com/ssi-dk/bifrost
11. Henri C, Leekitcharoenphon P, Carleton HA, Radomski N, Kaas RS, el al. An Assessment of Different Genomic Approaches for Inferring Phylogeny of Listeria monocytogenes. Front Microbiol 2017 Nov 29;8:2351. doi: 10.3389/fmicb.2017.02351. eCollection 2017.

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*	Sciensano
Denmark	Diagnostic and Typing of Gastrointestinal Bacteria	Statens Serum Institut
Finland	Expert Microbiology Unit	Finnish Institute for Health and Welfare
France	National Reference Centre and WHO Collaborating Centre for *Listeria*	Institut Pasteur
Germany	Consultant Laboratory for *Listeria monocytogenes*	Robert Koch Institute
Greece	National Reference Centre for *Salmonella Shigella*	University of West Attica
Hungary	Food and Waterborne Diseases Laboratory	Division of Microbiological Reference Laboratories
Ireland	NSSLRL	Medical Microbiology Department
Italy	Department of Food Safety, Nutrition and Veterinary Public Health	Istituto Superiore di Sanità
Latvia	Infectology Centre of Latvia, National Microbiology Reference laboratory	Riga East University hospital
Luxembourg	Epidemiology and Microbial Genomics	Laboratoire National de Santé
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	Enheten för laborativ bakterieövervakning, MI-LB	Folkhälsomyndigheten
The Netherlands	Infectious Disease Research, Diagnostics and Laboratory Surveillance (IDS)	RIVM

Annex 2. Participation overview EQA-7 and 8

	2019 to 2020 (EQA-7)					2020 to 2021 (EQA-8)				
	All[#]	Serotyping		Cluster		All[#]	Serotyping		Cluster	
Laboratory		Conven -tional	Molecular	PFGE	WGS		Conven-tional	Molecular	PFGE	WGS
19	X		X		X	X				X
35	X		X		X	X		X		X
49						X				X
56	X	X	X		X	X	X			X
70	X		X		X	X		X		X
100	X	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		
135*	X				X	X				X
138	X	X		X		X	X		X	
141	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					
149	X		X		X	X		X		X
Number of participants	**17**	4	13	3	13	**17**	4	11	2	14

[#]: participating in at least one element.
**: previously Laboratory 77.*

Annex 3. Serotyping result scores

Conventional serotyping

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

Molecular serogrouping

		Laboratory ID										
Isolate	Provider	35	70	100	105	108	129	130	142	143	144	149
Sero1	IIb	IIb	IIb	IIb	IIb	IIb	IIb	IIb	IIb	IIb	IIb	IIb
Sero2	IVb	IVb	IVb	IVb	IVb	IVb	IVb	IVb	IVb	IVb	IVb	IVb
Sero3	IIc	IIc	IIc	IIc	IIc	IIc	IIc	IIc	IIc	IIc	IIc	IIc
Sero4	IIa	IIa	IIa	IIa	IIa	IIa	IIa	IIa	IIa	IIa	IIa	IIa
Sero5	IVb	IVb	IVb	IVb	IVb	IVb	IVb	IVb	IVb	IVb	IVb	IVb
Sero6	IIb	IIb	IIb	IIb	IIb	IIb	IIb	IIb	IIb	IIb	IIb	IIb
Sero7	IIc	IIc	IIc	IIc	IIc	IIc	IIc	IIc	IIc	IIc	IIc	IIc
Sero8	IIc	IIc	IIc	IIc	IIc	IIc	IIc	IIc	IIc	IIc	IIc	IIc
Sero9	IIc	IIc	IIc	IIc	IIc	IIc	IIa	IIc	IIc	IIc	IIc	IIc
Sero10	IIc	IIc	IIc	IIc	IIc	IIc	IIc	IIc	IIc	IIc	IIc	IIc
Sero11	IIc	IIc	IIc	IIc	IIc	IIc	IIc	IIc	IIc	IIc	IIc	IIc
		A	A	A	B	B	B	A	A	A	A	B

A: PCR-based serotyping, B: WGS-based serotyping
Purple: repeat isolates in EQA-1 to 8 (Sero3, Sero4 and Sero5). Sero4 was a different isolate to previous years, but the same serotype/group.
Pink shading: incorrect results.

Annex 4. Reported cluster of closely-related isolates based on PFGE-derived data

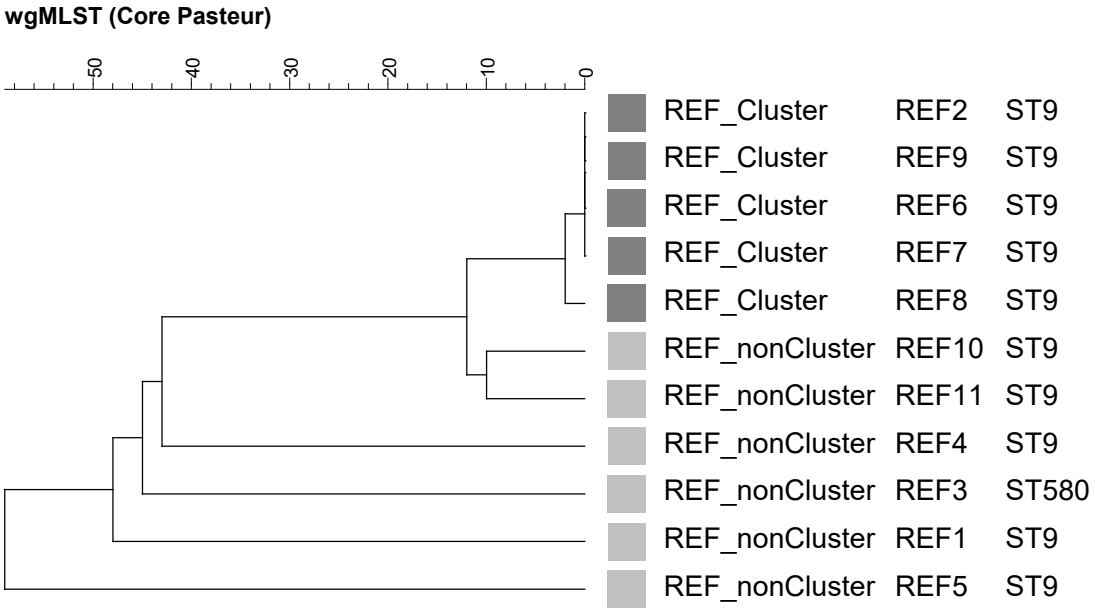
Lab ID	Reported cluster	Corresponding REF isolates	Correct
Provider		REF2, REF6, REF7, REF8 and REF9 (2/6 and 7/9 technical duplicates)	
138	1118, 1308, 1558, 1678, 1005, 1147, 1839	REF8, REF7, REF9, REF6, REF4, REF5, REF2	No
142	1181, 1550, 1711, 1928, 1951, 1999	REF2, REF4, REF6, REF9, REF8, REF7	No

Annex 5. Reported sequencing details

Sequencing performed	Protocol (library prep)	Commercial kit	Sequencing platform
In own laboratory	Commercial kits	NexteraXT	NextSeq
In own laboratory	Commercial kits	Nextera XT DNA Library Prep Kit (Illumina)	MiSeq
In own laboratory	Commercial kits	Illumina DNA Prep	MiSeq
In own laboratory	Commercial kits	NEBNEXT FAST DNA fragmentation and library prep set per Ion Torrent (50rxns) *	Ion Gene Studio S5 System
In own laboratory	Commercial kits	Nextera Flex Illumina	MiniSeq Illumina
In own laboratory	Commercial kits	Illumina Nextera Flex	MiSeq
In own laboratory	Commercial kits	nextera DNA flex library prep (Illumina)	NextSeq
In own laboratory	Commercial kits	Ion Xpress TM Plus Fragment Library Kit for AB, Library Builder TM System	IonTorrent S5XL
In own laboratory	Commercial kits	Nextera XT	MiSeq
In own laboratory	Commercial kits	Illumina DNA prep kit	NextSeq
In own laboratory	Commercial kits	Nextera XT DNA Library Kit, Illumina	MiSeq
In own laboratory	Commercial kits	Nextera	MiSeq
In own laboratory	Commercial kits	Illumina Nextera Flex	MiSeq
In own laboratory	Commercial kits	Kapa HyperPlus (Kapa Biosystems)	NextSeq

**The shearing was carried out for 15 minutes at 25 degrees instead of 20 minutes. This is because a 400bp sequencing protocol was used. The reaction was performed in half the volume suggested in the manufacturer's instructions, starting from 100 ng of DNA.*

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



Single linked dendrogram of core genome multi-locus sequence typing (cgMLST) profiles of Listeria *EQA-8 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 7. Reported cluster of closely-related isolates based on WGS-derived data

Lab ID Provider	Reported cluster	Corresponding to REF isolates REF2, REF6, REF7, REF8 and REF9 (2/6 and 7/9 technical duplicates)	Correct
19	1063, 1114, 1696, 1846, 1823	REF6, REF9, REF7, REF8, REF2	Yes
35	1261, 1424, 1570, 1661, 1938	REF6, REF9, REF7, REF2, REF8	Yes
49	1393, 1639, 1726, 1997, 1608	REF2, REF9, REF6, REF7, REF8	Yes
56	1186, 1310, 1408, 1710, 1975	REF6, REF9, REF7, REF8, REF2	Yes
70	1123, 1269, 1406, 1640, 1879	REF9, REF2, REF7, REF8, REF6	Yes
100	1736, 1959, 1486, 1949, 1609	REF6, REF2, REF7, REF9, REF8	Yes
105	1032, 1561, 1638, 1506, 1944	REF6, REF2, REF8, REF9, REF7	Yes
108	1000, 1357, 1369, 1682, 1437	REF9, REF7, REF2, REF6, REF8	Yes
129	1214, 1367, 1613, 1630, 1673	REF6, REF8, REF9, REF7, REF2	Yes
135	1470, 1693, 1851, 1881, 1143	REF9, REF6, REF7, REF8, REF2	Yes
141	1372, 1572, 1887, 1979, 1416	REF7, REF6, REF9, REF8, REF2	Yes
142	1181, 1711, 1928, 1951, 1999	REF2, REF6, REF9, REF8, REF7	Yes
144	1027, 1249, 1026, 1404, 1886	REF6, REF2, REF7, REF9, REF8	Yes
149	1284, 1374, 1282, 1971, 1200	REF9, REF7, REF2, REF6, REF8	Yes

Annex 8. Reported SNP distance and allelic differences

SNP distances

			Laboratory ID		
Isolate number	ST	Provider	19*	100	108
REF1	ST9	299	163	59	2674
REF2‡#1	ST9	0	1	0	1
REF3	ST580	109	95	54	1156
REF4	ST9	162	137	58	1690
REF5	ST9	234	178	65	1079
REF6‡#1	ST9	0¤	1	0¤	1
REF7‡#2	ST9	1	0	1	0
REF8‡	ST9	7	8	3	12
REF9‡#2	ST9	1	0¤	1	0¤
REF10	ST9	56	43	18	1084
REF11	ST9	48	34	10	212

Allelic distances

			Laboratory ID												
Isolate number	ST	Provider	19	35	49	49*	56	70	105	129	135	141	142	144	149
REF1	ST9	49	50	50	50	119	50	51	51	37	51	51	51	51	52
REF2‡#1	ST9	0	0	0¤	0¤	0¤	1	0¤	0	2	0	0	0¤	0	1
REF3	ST580	46	45	45	46	105	24	46	46	36	46	48	48	46	48
REF4	ST9	44	44	52	44	116	46	52	52	43	52	52	44	51	53
REF5	ST9	59	60	58	61	131	50	59	59	47	59	60	61	59	60
REF6‡#1	ST9	0¤	0	0	0	0	0¤	0	0¤	1	0¤	0¤	0	0¤	1
REF7‡#2	ST9	0	0	1	0	1	2	1	1	0	1	1	0	1	0
REF8‡	ST9	2	2	3	2	7	3	3	3	5	3	3	3	3	4
REF9‡#2	ST9	0	0¤	1	0	1	3	1	1	0¤	1	4	0	1	0¤
REF10	ST9	15	15	15	15	54	13	15	17	14	15	15	16	14	16
REF11	ST9	12	13	11	13	26	10	11	11	11	11	11	13	11	12

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

Annex 9. Additional reported QC parameters

Lab ID	1		2		3		4	
	Parameter	**Threshold**	**Parameter**	**Threshold**	**Parameter**	**Threshold**	**Parameter**	**Threshold**
19	N50	Available from QC analysis but no threshold	Number of contigs	Available from QC analysis but no threshold	Number of unidentified bases (N) or ambigiues sites	Available from QC analysis but no threshold		
35	N50	>20 kb						
49	N50	Pass > 100 000	Nr BAF multiple	Pass <20				
56	MLST	No mismatches						
100	N50	400000	SAV	cluster density, clusters passing filter and Q30 score were all according to Illumina recommend-ations.				
135	Contamination of assembly	CheckM, <4%	Contig number > 500 nucleotides	<300	N50	>10,000	GC%	37.6-38.2%
141	No. of contigs	200 bases (contigs shorter than 200 bases have to be ignored)						
142	GC content of 38%:	Warning >2% deviation failure >4% deviation						
144	N50	Read length						

Annex 10. Calculated qualitative/quantitative parameters

		Laboratory 19										
Quali-/Quantitative	Ranges*	1063	1096	1114	1260	1288	1696	1823	1846	1966	1970	1990
Detected species	{*Lm*}	Lm	Lm	Lm	Lm	Lm	Lm	Lm	Lm	Lm	Lm	Lm
% Species 1		94.7	95.1	95.1	95.0	94.8	95.3	95.3	95.5	94.9	91.9	95.2
% Species 2	{<5%}	0.1	0.1	0.1	0.1	0.2	0.1	0.1	0.1	0.1	0.1	0.2
Unclassified reads (%)		5.0	4.8	4.7	4.8	4.9	4.4	4.5	4.3	4.9	7.9	4.6
Length at >25 x min. coverage (Mbp)	{>28 ∧ <31}	2.9	3.0	2.9	2.9	3.0	2.9	2.9	2.9	3.0	2.9	3.0
Length [1-25] x min. coverage (kbp)	{<250}	0.5	0.3	0.0	4.4	1.6	0.0	5.0	0.0	3.3	4.4	3.7
No. of contigs at 25 x min. coverage	{>0}	39	45	38	45	46	29	31	23	32	48	45
No. of contigs [1-25] x min. coverage	{<1000}	1	1	0	4	2	0	4	0	2	4	3
Average coverage	{>50}	120	91	134	120	76	110	95	111	105	112	106
No. of reads (x 1000)		2 575	1 980	2 860	2 544	1 641	2 315	2 004	2 331	2 281	2 396	2 278
Average read length		139	141	140	140	141	142	142	142	140	140	141
Average insert size		235	253	235	241	260	258	259	254	248	246	244
N50 (kbp)		192	171	192	129	147	264	276	361	242	124	150

		Laboratory 35										
Quali-/Quantitative	Ranges*	1004	1110	1261	1424	1476	1566	1570	1590	1661	1876	1938
Detected species	{Lm}	Lm	Lm	Lm	Lm	Lm	Lm	Lm	Lm	Lm	Lm	Lm
% Species 1		65.3	96.0	94.8	95.6	95.2	94.6	96.1	97.2	94.3	95.9	95.3
% Species 2	{<5%}	0.5	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Unclassified reads (%)		34.2	4.0	5.1	4.3	4.7	5.3	3.9	2.7	5.6	4.1	4.7
Length at >25 x min. coverage (Mbp)	{>28 ∧ <31}	2.9	2.9	2.9	2.9	3.0	2.9	2.9	3.0	2.9	3.0	2.9
Length [1-25] x min. coverage (kbp)	{<250}	0.0	0.0	0.0	0.0	0.0	53.2	0.0	1.9	0.2	0.0	0.0
No. of contigs at 25 x min. coverage	{>0}	39	40	41	32	32	31	29	28	32	35	30
No. of contigs [1-25] x min. coverage	{<1000}	0	0	0	0	0	15	0	1	1	0	0
Average coverage	{>50}	109	111	77	97	127	110	115	209	119	108	124
No. of reads (x 1000)		2 402	2 423	1 735	2 140	2 890	2 544	2 517	4 388	2 736	2 436	2 769
Average read length		134	136	131	133	132	131	136	144	129	135	133
Average insert size		0	0	0	0	0	0	0	251	0	0	0
N50 (kbp)		232	192	223	435	476	462	476	476	476	456	435

		Laboratory 49										
Quali-/Quantitative	Ranges*	1098	1354	1393	1463	1608	1639	1726	1777	1813	1976	1997
Detected species	{Lm}	Lm	Lm	Lm	Lm	Lm	Lm	Lm	Lm	Lm	Lm	Lm
% Species 1		97.7	98.6	98.8	96.7	98.2	98.1	98.5	98.6	98.6	98.4	98.5
% Species 2	{<5%}	0.0	0.0	0.0	0.1	0.0	0.0	0.0	0.1	0.0	0.0	0.0
Unclassified reads (%)		2.1	1.3	1.1	3.0	1.7	1.8	1.3	1.3	1.3	1.6	1.4
Length at >25 x min. coverage (Mbp)	{>28 ∧ <31}	2.9	3.0	2.9	2.9	2.9	2.9	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	0.0	0.0	0.0	0.0
No. of contigs at 25 x min. coverage	{>0}	16	28	19	17	16	12	14	24	18	15	13
No. of contigs [1-25] x min. coverage	{<1000}	0	0	0	0	0	0	0	0	0	0	0
Average coverage	{>50}	78	137	174	132	111	72	88	129	123	77	59
No. of reads (x 1000)		840	1 793	2 072	1 519	1 239	780	912	1 499	1 412	825	622
Average read length		280	232	249	259	269	275	287	260	266	282	282
Average insert size		400	242	269	291	323	425	448	293	312	426	448
N50 (kbp)		509	477	477	581	477	514	514	477	510	514	514

		Laboratory 56										
Quali-/Quantitative	Ranges*	1105	1186	1207	1310	1408	1441	1514	1615	1619	1710	1975
Detected species	{Lm}	Lm	Lm	Lm	Lm	Lm	N/A	Lm	N/A	Lm	N/A	Lm
% Species 1		94.2	98.3	98.4	97.8	98.0	-	98.2	-	97.9	-	98.1
% Species 2	{<5%}	0.1	0.0	0.0	0.0	0.0	-	0.0	-	0.0	-	0.0
Unclassified reads (%)		5.6	1.7	1.5	2.1	2.0	-	1.7	-	2.1	-	1.9
Length at >25 x min. coverage (Mbp)	{>28 ∧ <31}	2.9	2.9	2.9	2.9	2.9	-	2.9	-	3.0	-	2.8
Length [1-25] x min. coverage (kbp)	{<250}	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}	250	602	624	270	562	-	633	-	633	-	885
No. of contigs [1-25] x min. coverage #	{<1000}	0	0	0	0	0	-	0	-	0	-	0
Average coverage	{>50}	296	259	220	200	215	-	227	-	226	-	274
No. of reads (x 1000)		3 455	2 930	2 345	2 240	2 395	-	2 477	-	2 599	-	2 969
Average read length		259	263	279	265	267	-	278	-	264	-	274
Average insert size #		46	61	49	42	59	-	58	-	48	-	48
N50 (kbp) #		24	7	8	17	10	-	9	-	8	-	5

		Laboratory 70										
Quali-/Quantitative	Ranges*	1123	1228	1269	1319	1324	1384	1406	1407	1610	1640	1879
Detected species	{Lm}	Lm	Lm	Lm	Lm	Lm	Lm, Bd	Lm	Lm	Lm	Lm	Lm
% Species 1		94.6	93.4	94.9	91.5	95.7	91.8	95.3	96.1	95.6	97.4	97.3
% Species 2	{<5%}	2.0	2.7	1.8	1.7	1.7	6.5	1.7	1.5	1.8	0.2	0.6
Unclassified reads (%)		3.2	3.5	3.1	6.4	2.4	1.6	2.8	2.3	2.4	2.2	2.0
Length at >25 x min. coverage (Mbp)	{>28 ∧ <31}	2.9	3.0	2.9	2.9	3.0	2.9	2.9	3.0	3.0	2.9	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}	22	31	21	19	25	28	21	22	29	20	19
No. of contigs [1-25] x min. coverage	{<1000}	0	0	0	0	0	0	0	0	0	0	0
Average coverage	{>50}	35	51	67	55	62	46	58	74	44	44	43
No. of reads (x 1000)		789	1 206	1 479	1 185	1 364	1 040	1 268	1 635	991	915	896
Average read length		151	151	151	151	151	151	151	151	151	151	151
Average insert size		209	163	204	231	230	202	217	218	222	251	271
N50 (kbp)		435	477	477	581	477	477	477	477	477	478	477

		Laboratory 100										
Quali-/Quantitative	Ranges*	1104	1421	1484	1486	1553	1609	1666	1671	1736	1949	1959
Detected species	{Lm}	Lm	Lm	Lm	Lm	Lm	Lm	Lm	Lm	Lm	Lm	Lm
% Species 1		98.9	98.8	98.5	99.1	98.6	99.2	95.9	99.0	98.9	98.7	98.8
% Species 2	{<5%}	0.0	0.1	0.3	0.2	0.0	0.1	0.2	0.0	0.1	0.1	0.0
Unclassified reads (%)		1.0	1.1	1.1	0.7	1.3	0.7	3.9	0.9	0.9	1.2	1.1
Length at >25 x min. coverage (Mbp)	{>28 ∧ <31}	3.0	2.9	2.9	2.9	3.0	2.8	2.9	3.0	2.9	2.9	2.9
Length [1-25] x min. coverage (kbp)	{<250}	11.1	88.2	0.0	0.0	1.4	154.2	3.2	2.1	0.0	0.0	4.4
No. of contigs at 25 x min. coverage	{>0}	15	29	18	18	20	32	18	15	15	17	16
No. of contigs [1-25] x min. coverage	{<1000}	5	58	0	0	2	97	1	1	0	0	4
Average coverage	{>50}	104	84	187	70	130	62	136	113	94	160	123
No. of reads (x 1000)		1 332	1 062	2 489	893	1 683	779	1 704	1 450	1 188	2 036	1 537
Average read length		237	238	223	233	234	234	237	235	235	233	238
Average insert size		328	329	274	297	316	306	327	314	319	314	334
N50 (kbp)		477	230	449	477	453	135	581	478	477	477	450

		Laboratory 105										
Quali-/Quantitative	Ranges*	1032	1133	1239	1474	1506	1561	1638	1641	1669	1944	1977
Detected species	{Lm}	Lm	Lm	Lm	Lm	Lm	Lm	Lm	Lm	Lm	Lm	Lm
% Species 1		97.0	96.6	96.4	96.0	96.4	97.0	96.1	96.6	97.0	96.9	96.7
% 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 (%)		2.9	3.4	3.5	3.9	3.5	2.9	3.8	3.3	3.0	3.1	3.2
Length at >25 x min. coverage (Mbp)	{>28 ∧ <31}	2.9	3.0	2.9	2.9	2.9	2.9	2.9	3.0	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	0.0	0.0	0.0
No. of contigs at 25 x min. coverage	{>0}	15	17	17	17	17	16	15	17	24	18	18
No. of contigs [1-25] x min. coverage	{<1000}	0	0	0	0	0	0	0	0	0	0	0
Average coverage	{>50}	165	130	106	150	157	193	97	130	433	274	166
No. of reads (x 1000)		3 272	2 631	2 113	2 988	3 140	3 835	1 958	2 625	8 859	5 490	3 401
Average read length		150	149	149	150	149	149	148	149	148	148	149
Average insert size		402	418	443	407	401	376	432	411	330	347	403
N50 (kbp)		477	476	477	564	477	477	477	514	477	477	490

		Laboratory 108										
Quali-/Quantitative	Ranges*	1000	1085	1088	1351	1357	1369	1437	1458	1682	1738	1740
Detected species	{Lm}	Lm	Lm	Lm	Lm	Lm	Lm	Lm	Lm	Lm	Lm	Lm
% Species 1		98.5	98.5	98.7	98.5	98.5	98.5	98.5	98.4	98.5	98.5	89.0
% Species 2	{<5%}	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.3
Unclassified reads (%)		1.4	1.4	1.3	1.4	1.5	1.4	1.4	1.6	1.5	1.5	10.7
Length at >25 x min. coverage (Mbp)	{>28 ∧ <31}	2.6	2.8	2.6	2.7	2.7	2.6	2.6	2.7	2.5	2.6	2.6
Length [1-25] x min. coverage (kbp)	{<250}	0.9	0.9	6.0	1.3	0.9	2.6	1.5	0.0	5.4	2.5	0.6
No. of contigs at 25 x min. coverage #	{>0}	3 101	2 253	3 116	3 186	2 765	3 069	3 217	2 898	3 522	3 259	3 278
No. of contigs [1-25] x min. coverage #	{<1000}	4	4	23	6	4	10	7	0	17	9	3
Average coverage	{>50}	74	72	62	79	67	62	71	92	58	61	79
No. of reads (x 1000)		766	825	633	833	706	653	728	971	597	644	877
Average read length		288	266	291	291	285	283	293	289	292	290	285
Average insert size #		0	0	0	0	0	0	0	0	0	0	0
N50 (kbp) #		1	2	1	1	2	1	1	1	1	1	1

		Laboratory 129										
Quali-/Quantitative	Ranges*	1141	1161	1214	1306	1335	1367	1613	1630	1673	1781	1795
Detected species	{Lm}	Lm	Lm	Lm	Lm	Lm	Lm	Lm	Lm	Lm	Lm	Lm
% Species 1		97.9	97.6	98.5	97.8	98.5	97.8	98.2	97.0	98.0	95.3	96.3
% Species 2	{<5%}	0.0	0.0	0.1	0.0	0.0	0.1	0.0	0.1	0.0	0.1	0.1
Unclassified reads (%)		2.0	2.3	1.4	2.1	1.5	2.1	1.7	2.8	1.9	4.6	3.6
Length at >25 x min. coverage (Mbp)	{>28 ∧ <31}	3.0	2.9	2.9	3.0	3.0	2.9	2.9	2.9	2.9	2.7	2.9
Length [1-25] x min. coverage (kbp)	{<250}	3.3	0.4	8.1	5.6	7.8	6.2	4.1	0.8	0.0	271.4	3.5
No. of contigs at 25 x min. coverage	{>0}	71	33	67	34	95	74	31	34	17	101	29
No. of contigs [1-25] x min. coverage	{<1000}	5	1	12	4	8	8	3	1	0	61	2
Average coverage	{>50}	207	149	140	130	138	142	150	155	146	28	70
No. of reads (x 1000)		4 230	3 004	2 795	2 659	2 807	2 827	2 976	3 122	2 907	604	1 422
Average read length		147	147	148	149	147	148	149	148	149	141	146
Average insert size		351	348	341	378	315	354	378	384	373	351	421
N50 (kbp)		81	238	96	202	69	76	192	168	435	34	173

		Laboratory 135										
Quali-/Quantitative	**Ranges***	**1015**	**1143**	**1189**	**1251**	**1300**	**1380**	**1470**	**1480**	**1693**	**1851**	**1881**
Detected species	{Lm}	Lm	Lm	Lm	Lm	Lm	Lm	Lm	Lm	Lm	Lm	Lm
% Species 1		94.0	97.6	97.8	98.1	98.1	97.9	97.8	98.0	97.6	97.2	97.8
% Species 2	{<5%}	0.3	0.6	0.3	0.3	0.4	0.4	0.4	0.4	0.5	1.4	0.6
Unclassified reads (%)		5.5	1.6	1.7	1.4	1.4	1.6	1.6	1.5	1.7	1.3	1.5
Length at >25 x min. coverage (Mbp)	{>28 ∧ <31}	2.9	2.9	3.0	3.0	2.9	3.0	2.9	3.0	2.9	2.9	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}	17	15	18	17	18	13	14	18	15	15	16
No. of contigs [1-25] x min. coverage	{<1000}	0	0	0	0	0	0	0	0	0	0	0
Average coverage	{>50}	460	231	206	324	317	351	238	209	203	154	191
No. of reads (x 1000)		9 238	4 687	4 235	6 658	6 395	7 210	4 824	4 272	4 127	3 117	3 874
Average read length		151	151	151	151	151	151	151	151	151	151	151
Average insert size		392	405	384	398	336	387	339	399	396	415	384
N50 (kbp)		564	477	514	515	491	514	514	477	477	477	477

		Laboratory 141										
Quali-/Quantitative	**Ranges***	**1197**	**1372**	**1416**	**1554**	**1572**	**1578**	**1607**	**1887**	**1969**	**1978**	**1979**
Detected species	{Lm}	Lm	Lm	Lm	Lm	Lm	Lm	Lm	Lm	Lm	Lm	Lm
% Species 1		90.7	95.9	95.1	94.7	94.8	96.5	95.9	95.0	95.8	94.9	94.7
% Species 2	{<5%}	0.3	0.1	0.0	0.1	0.1	0.0	0.0	0.1	0.0	0.1	0.1
Unclassified reads (%)		8.8	3.9	4.8	4.9	4.8	3.4	4.0	4.8	4.1	4.9	5.1
Length at >25 x min. coverage (Mbp)	{>28 ∧ <31}	2.9	2.9	2.9	3.0	2.9	3.0	2.9	2.9	3.0	2.9	2.9
Length [1-25] x min. coverage (kbp)	{<250}	4.0	0.9	1.7	2.4	7.9	1.9	118.6	1.2	3.5	3.0	10.3
No. of contigs at 25 x min. coverage	{>0}	42	39	58	61	55	51	150	50	64	51	85
No. of contigs [1-25] x min. coverage	{<1000}	3	2	1	2	6	3	58	2	5	5	10
Average coverage	{>50}	81	74	74	57	53	70	31	47	47	80	59
No. of reads (x 1000)		955	856	870	641	579	833	435	524	648	926	688
Average read length		252	263	257	276	276	259	217	271	220	263	260
Average insert size		314	324	338	378	376	311	250	358	260	338	346
N50 (kbp)		112	191	70	81	92	108	30	98	125	98	61

		Laboratory 142										
Quali-/Quantitative	**Ranges***	**1137**	**1181**	**1280**	**1336**	**1550**	**1650**	**1711**	**1831**	**1928**	**1951**	**1999**
Detected species	{Lm}	Lm	Lm	Lm	Lm	Lm	Lm	Lm	Lm	Lm	Lm	Lm
% Species 1		98.2	97.0	98.4	98.2	97.6	96.4	98.2	98.2	98.8	98.1	98.2
% Species 2	{<5%}	0.2	0.2	0.1	0.2	0.2	0.3	0.1	0.1	0.1	0.2	0.2
Unclassified reads (%)		1.5	2.6	1.3	1.4	2.0	3.1	1.5	1.6	0.9	1.6	1.4
Length at >25 x min. coverage (Mbp)	{>28 ∧ <31}	3.0	2.9	3.0	3.0	3.0	2.9	2.9	2.9	2.9	2.9	2.9
Length [1-25] x min. coverage (kbp)	{<250}	0.0	3.7	0.0	0.0	0.0	11.3	0.0	0.0	4.8	0.0	0.0
No. of contigs at 25 x min. coverage	{>0}	18	20	15	13	18	36	13	15	50	14	12
No. of contigs [1-25] x min. coverage	{<1000}	0	3	0	0	0	11	0	0	7	0	0
Average coverage	{>50}	60	48	67	60	63	65	61	69	86	75	57
No. of reads (x 1000)		745	581	831	752	791	802	736	839	1 073	916	699
Average read length		245	247	243	244	245	243	246	244	237	244	244
Average insert size		484	576	451	453	472	413	504	446	361	446	455
N50 (kbp)		514	415	568	514	515	250	514	516	147	605	565

		Laboratory 144										
Quali-/Quantitative	Ranges*	1026	1027	1034	1094	1163	1249	1331	1404	1549	1774	1886
Detected species	{Lm}	Lm	Lm	Lm	Lm	Lm	Lm	Lm	Lm	Lm	Lm	Lm
% Species 1		98.4	97.9	94.5	98.0	97.7	97.8	98.1	98.5	97.7	98.5	98.0
% Species 2	{<5%}	0.0	0.0	0.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Unclassified reads (%)		1.6	2.0	5.4	2.0	2.2	2.1	1.9	1.5	2.2	1.5	1.9
Length at >25 x min. coverage (Mbp)	{>28 ∧ <31}	2.9	2.9	2.9	3.0	3.0	2.9	2.9	2.9	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	0.0	0.0	0.0	0.0
No. of contigs at 25 x min. coverage	{>0}	16	14	17	21	16	14	19	17	18	18	14
No. of contigs [1-25] x min. coverage	{<1000}	0	0	0	0	0	0	0	0	0	0	0
Average coverage	{>50}	86	91	57	70	64	66	72	79	77	94	59
No. of reads (x 1000)		1 713	1 810	1 135	1 429	1 310	1 323	1 444	1 581	1 565	1 896	1 172
Average read length		149	149	149	148	148	148	147	149	149	149	148
Average insert size		368	358	383	345	368	317	376	340	358	346	355
N50 (kbp)		476	509	582	477	533	478	477	477	477	477	508

		Laboratory 149										
Quali-/Quantitative	Ranges*	1200	1252	1282	1284	1304	1374	1569	1792	1931	1971	1988
Detected species	{Lm}	Lm	Lm	Lm	Lm	Lm	Lm	Lm	Lm	Lm	Lm	Lm
% Species 1		94.1	91.4	93.5	94.6	90.2	93.6	94.9	94.5	94.1	87.1	85.9
% Species 2	{<5%}	2.4	3.5	1.5	0.9	3.1	1.0	0.4	0.6	1.0	4.9	2.8
Unclassified reads (%)		3.4	4.8	4.8	4.3	6.3	5.3	4.6	4.7	4.7	6.8	10.7
Length at >25 x min. coverage (Mbp)	{>28 ∧ <31}	2.9	2.9	2.9	2.9	3.0	2.9	3.0	3.0	3.0	2.9	2.9
Length [1-25] x min. coverage (kbp)	{<250}	0.0	0.0	0.0	0.0	0.6	0.0	0.0	0.0	0.0	0.0	0.0
No. of contigs at 25 x min. coverage	{>0}	16	28	20	23	21	22	28	26	24	26	19
No. of contigs [1-25] x min. coverage	{<1000}	0	0	0	0	1	0	0	0	0	0	0
Average coverage	{>50}	81	101	135	121	124	119	88	97	139	167	237
No. of reads (x 1000)		944	1 330	2 842	2 461	2 729	2 445	1 812	2 017	2 977	3 819	5 164
Average read length		270	249	151	151	151	151	151	151	151	151	151
Average insert size		343	286	255	305	285	305	317	266	248	271	260
N50 (kbp)		477	477	477	477	477	477	477	477	480	477	581

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
Lm: L. monocytogenes
Bd: Burkholderia dolosa

Annex 11. Accessing genomes

	Participant		EQA provider*	
Lab ID	Cluster	Description Genome 1	QC issues	Cluster
EQA provider		A cluster isolate (REF2) with altered coverage (reduced to 15x)	Yes	Yes
19	No	Genome 1 is of poor quality. The read coverage is low, resulting in a low core % and poor genome assembly (low N50, very small genome size). The genome can therefore not be used for comparison and should be re-sequenced for better genome quality.	Yes	-
35	No	Sequencing quality is too low to be accepted. However, the isolate may belong to the cluster and therefore should be re-sequenced.	Yes	-
49	No	The average read coverage (15), length (9728) and % core genes (17%) are far too low to perform a proper analysis on this isolate. In our laboratory we would measure the DNA concentration of the original extract, repeat extract if too low, and re-sequence. Of interest: the 17% of genes that were present were an exact match with those of the representative strain from this cluster so it is possible it would prove related on re-sequencing.	Yes	-
56	No	The coverage (16x) is not sufficient for analysis to re-extract the genome and do sequencing	Yes	-
70	No	Sequencing quality of Genome 1 is not optimal, assembly was done by bwa (using a reference mapping). Coverage is only 15 and allele calling detected 82.7% of targets. Comparison of this genome against the cluster is based on 1 405 targets and generates an AD of 25, which is above the alert distance of 10.	Yes	No
100	No	Genome 1 has too low coverage (15x) and too low % of good targets, it also belongs to a different serogroup (IIa instead of IIc).	Yes	-
105	No	Result of sequence read quality QC FAIL: CGC <90% (44%), depth of coverage 10-15X so the genome 1 should be rejected.	Yes	-
108	Yes	The sample is a part of the cluster and is closely-related to genome 2 and sample 1437, however the sequence data is of lower quality and does not pass the quality parameters in terms of coverage >20x. Our WGS results suggest IIa as serotype, but we believe it is a IIc. In routine analysis we would have re-run this sample.	Yes	Yes
129	No	It is unknown whether genome 1 is part of the cluster. Only 40% of targets were present, and coverage was 15. If, genome 1 could still be included in the comparison (despite poor quality), it would be part of the cluster with 0 allele difference, but all other strains would also be very close to the cluster, with few allele differences. Re-run should really be done. In the meantime, epidemiological investigation should focus on cases with clear results being part of the cluster.	Yes	Yes
135	No	Genome 1 would not be considered for cluster analysis, as the quality is too low. There are a low number of reads present, the contig number and GC % are too high, and the N50 and coverage are too low. In our workflow, this sample would be re-run.	Yes	-
141	No	Too many allelic differences (AD): 65 AD to cluster representative Isolate 1572. Quality of sequence is not good - only 76.5 % good targets, 419 missing values.	Yes	No
142	No	QC not OK: low coverage (10x) low allele calling for cgMLST (65.6%)	Yes	-
144	Yes	Closely related to outbreak cluster.	No	Yes
149	No	The quality of the sequence is too low. We cannot evaluate if genome 1 is part of the cluster or not. We would recommend re-sequencing. If that does not work we will extract DNA and do the library prep and sequencing again. Bad quality since:-40.6% good targets (<90% good targets)-Low coverage-Low N50	Yes	-

	Participant		EQA provider*	
Lab ID	Cluster	Description Genome 2	Fasta file analysed	Cluster
EQA provider		A cluster isolate (REF8), good quality of reads assembled with SKESA to a FASTA file. 2 allelic difference to the REF2 in the cluster.	Yes	Yes
19	Yes	Genome 2 has 2 allele differences (2 AD) to the cluster representative genome and is therefore considered part of the cluster.	Yes	Yes
35	Yes	In 1701 locus cgMLST, genome 2 is indifferent to EQA-sample 1938 and differs only in 3-6 alleles to the remaining 5 isolates (including genome 5) belonging to this cluster.	Yes	Yes
49	Yes	Genome 2 was a Fasta file and quality was satisfactory. There were just 2 cgMLST alleles between genome 2 and the representative isolate so this would be considered part of the cluster.	Yes	Yes
56	No	Our toolset cannot manage Fasta files.	-	-

	Participant		EQA provider*	
Lab ID	**Cluster**	**Description Genome 2**	**Fasta file analysed**	**Cluster**
70	Yes	Genome 2 was provided as a Fasta file of assembly and 99.3% of the targets were identified, which reflects a very good sequencing quality. Genome 2 has an identical core genome to one representative of the detected cluster: strain 1640.	Yes	Yes
100	Yes	The allelic difference with representative isolate is 2. Genome 2 is closely related to the representative isolate. Also sequence type, complex type and serogroups are the same as for representative isolate.	Yes	Yes
105	Yes	Genome 2 PASSES all QCs and is 3 allelic differences to the selected cluster isolate.	Yes	Yes
108	Yes	We do not have the possibility to evaluate the quality parameters for FASTA sequences as for sequencing read files. The genome clusters with 1437 and is closely related to 1000, 1357, 1369, 1682, genome1 and genome5.	Yes	Yes
129	Yes	Because no Fastq reads were available, comparison with Fasta was necessary. Percentage of good targets was acceptable 98%. Two allele differences with strain 5 (index) were found.	Yes	Yes
135	Yes	We observe an allelic distance of 0 to isolate 6 (ID 1881), which is part of the cluster. We observe an allelic distance of 3 to the representative isolate 4 (ID 1693).	Yes	Yes
141	Yes	Only a few allelic differences (AD): 3 AD to cluster representative Isolate 1572, 0 AD to cluster Isolate 1979. Quality of sequence is good - 99.3% good targets, genome size OK (2.9)	Yes	Yes
142	Yes	Only 3 allelic differences with reference sequence 1181	Yes	Yes
144	Yes	Closely related to outbreak cluster	Yes	Yes
149	Yes	Genome 2 has good quality (99.3% good targets) and clusters together with 1200 (0 allele difference), and only two alleles from 1284.	Yes	Yes

	Participant		EQA provider*	
Lab ID	**Cluster**	**Description Genome 3**	**QC Issues**	**Cluster**
EQA provider		A cluster isolate (REF11) with altered coverage (reduced to 10x)	Yes	Yes
19	No	Genome 3 is of poor quality. The read coverage is low, resulting in a low core % and poor genome assembly (low N50, very small genome size). The genome can therefore not be used for comparison and should be resequenced for better genome quality.	Yes	-
35	No	Sequencing quality is too low to be accepted. However, the isolate may belong to the cluster and therefore should be re-sequenced.	Yes	-
49	No	The average read coverage (10), length (10587) and % core genes (7%) are far too low to perform a proper analysis of this isolate. In our laboratory we would measure the DNA concentration of the original extract, repeat extract if too low, and re-sequence.	Yes	-
56	No	The coverage (11x) is not sufficient for analysis, re-extracting the genome and sequencing.	Yes	-
70	No	The sequencing quality of Genome 3 is quite low and assembly was done by using bwa (with a reference mapping). With this approach, 69.7% of good targets were detected associated with a coverage of 10. By comparing this genome 3 to one representative of the cluster, an AD of 45 is found by using the 1185 targets available for the analysis (it is above the cluster alert of 10).	Yes	No
100	No	Genome 3 has too low coverage (10x) and too low % of good targets, it also belongs to a different serogroup (IIa instead of IIc).	Yes	-
105	No	Result of sequence read quality QC FAIL: CGC <90% (28%), depth of coverage 10-15X so genome 3 should be rejected.	Yes	-
108	No	The sequence does not pass our quality parameters and cannot be included in the cluster assessment. The coverage is too low.	Yes	-
129	No	Sequence quality was really poor. Only 23% of targets were present and average coverage was 11. If genome 3 still to be included in the comparison (despite poor quality), it seemed not to be part of the cluster (despite 2 allele difference from index). Genome 3 was closer to another strain, which is not part of the cluster. This strain should be re-run before any conclusions.	Yes	No
135	No	Genome 3 would not be considered for cluster analysis, as the quality is too low. There are a low number of reads present, the contig number and GC % are too high, and the N50 and coverage are too low. In our workflow, this sample would be re-run.	Yes	-
141	No	Too many allelic differences (AD): 99 AD to cluster representative Isolate 1572. Quality of sequence is not good - only 61.6% good targets, 674 missing values.	Yes	-
142	No	QC not OK: low coverage (7x) low allele calling for cgMLST (47.6%)	Yes	-
144	No	Based on allelic distance, does not belong to this outbreak.	No	No
149	No	The quality of the sequence is too low. We cannot evaluate if genome 3 is part of the cluster or not. We would recommend re-sequencing. If that does not work we will extract DNA and do the library prep and sequencing again. Bad quality since:- 23.6% good targets (<90% good targets)-Low coverage-Low N50.	Yes	-

Participant			EQA provider*	
Lab ID	**Cluster**	**Description Genome 4**	**QC Accepted**	**Cluster**
EQA provider		A non cluster isolate (REF12), good quality of reads, 58 allelic difference to the cluster isolate (REF2)	Yes	No
19	No	Genome 4 has 59 allele differences (59 AD) to the cluster representative genome and is therefore not considered part of the cluster.	Yes	Yes
35	No	The allelic distance to the closest related isolate is 61 alleles.	Yes	Yes
49	No	Genome 4 is not part of this cluster as there are 59 cgMLST alleles difference between it and the representative isolate.	Yes	Yes
56	No	Because there are 86 AD to sample 1186	Yes	Yes
70	No	99.8% of good targets are available for comparing this genome to the cluster. An allele distance of 64 excludes genome 4 from the outbreak.	Yes	Yes
100	No	Genome 4 has good coverage (114x). The allelic difference to representative isolate is 70, so it is not part of the cluster. It also belongs to different serogroup (IIa instead of IIc) and it is of a different complex type to the representative isolate.	Yes	Yes
105	No	Genome 4 PASSES all QCs, it is not a part of the cluster because is >4 allelic differences (is 62 ADs to the selected cluster isolate).	Yes	Yes
108	No	Genome 4 is not closely related to any other isolate in the cluster assessment.	Yes	Yes
129	No	Good quality sequence. Allele difference 48 to the index.	Yes	Yes
135	No	Genome 4 is of sufficient quality. Its allelic distance to the representative isolate 4 (ID 1693) of the cluster is 63 and it is therefore not part of the cluster.	Yes	Yes
141	No	Too many allelic differences (AD): 61 AD to cluster representative Isolate 1572 Quality of sequence is good - 99.3% good targets, 31 missing values	Yes	No
142	No	Allelic difference with reference sequence 1181 = 64 AD	Yes	No
144	Yes	Closely related to outbreak cluster	Yes	Yes
149	No	Genome 4 has good quality (99.3% good targets). It has 64 alleles difference to 1284 and is not part of the cluster.	Yes	No

	Participant		EQA provider*	
Lab ID	**Cluster**	**Description Genome 5**	**Contamination**	**Cluster**
EQA provider		A cluster isolate (REF7/REF9) mixed with a *Listeria monocytogenes* ST1, (approx. 15%) same species contamination.	Yes	Yes
19	No	Genome 5 is probably contaminated with other Listeria (no fail in Kraken analysis), here resulting in too large genome size, low N50, many contigs, many unknown bases and ambiguous sites and a high number of loci with multiple consensus. However, the core % is high (99%) and based on the cgMLST comparison, Genome 5 is suspected to be closely related to the outbreak cluster (with 11 AD to the representative outbreak genome), however, the isolate should be re-streaked for pure culture and re-sequenced for conformation.	Yes	Yes
35	Yes	In 1701 locus cgMLST, genome 5 differs only in 2-6 alleles to the remaining five isolates (including genome 5) belonging to this cluster. However, the assembly of genome 5 is of borderline quality (N50 <30 kb, Percentage of good cgMLST targets <95%). The isolate should be re-sequenced.	Yes	Yes
49	No	The length (4.7 MB), number of contigs (4068), N50 size (1743) and number of non-ACGT (3578) and N (182945) suggest a mixed genome. I would propose sub-culturing a single colony of Listeria, extracting DNA and re-sequencing.	Yes	-
56	No	Because there are 182 AD with sample 1186.	No	No
70	Yes	Genome 5 has a coverage of 114 and 98.6% of good targets were detected. The sequencing quality is really good. This genome has an identical core genome than the two following strains included in the cluster: 1123 and 1406.	No	Yes
100	Yes	The allelic difference with representative isolate is 1. Genome 5 is closely related to the representative isolate. Average coverage is good (114x), sequence type, complex type and serogroups are the same as for representative isolate.	No	Yes
105	No	Genome 5 PASSES all QCs, with the exception of the sequence read quality which is WARN (CGC = 93.8%), so genome 5 is accepted for outbreak. Genome 5 is not a part of the cluster because it is >4 allelic differences, although it would be part of the extended cluster because it is 7 AD from two isolates of the cluster (1506 and 1944).	No	No
108	Yes	The genome size is too large, suggesting that the sample is contaminated. The possible contamination is also reflected in the results on the molecular serotype. Reads are found for all the genes, prs, flaA, prfA ORF2110, ORF2819, lmo0737, lmo1118, however the coverage for ORF2110 and ORF2819 is very low. Our WGS analysis suggests IVb but we believe that IIc is the dominant serotype in this sample. The sample is a part of the cluster and closely related to genome2 and sample 1437. In routine analysis we would recalculate and rerun this sample.	Yes	Yes
129	Yes	Good quality sequence. Allele difference 5 to the index.	No	Yes

	Participant		EQA provider*	
Lab ID	**Cluster**	**Description Genome 5**	**Contamination**	**Cluster**
135	No	Genome 5 would not be considered for cluster analysis, as it is contaminated with another species. This results in a too high a contig number and GC %, also the N50 is too low. In our workflow, this sample would be re-run.	Yes	-
141	No	Too many allelic differences (AD): 21 AD to cluster representative Isolate 1572 Quality of sequence is not good - only 55.7% good targets, 772 missing values. Genome size too big for Listeria (4.9) - possibly due to contamination	Yes	No
142	No	QC not OK: data probably origin from a mix of two Listeria samples: Genome size too big: 3.6 MB PCR serogroup alleles are all called, 100 cgMLST alleles have double calls and 173 alleles lack (complete) calling.	Yes	-
144	No	Based on allelic distance it does not belong to this outbreak	No	No
149	Yes	Genome 5 is good quality (93.5% good targets) and clusters with only two alleles from 1284.	No	Yes

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

Annex 12. EQA-8 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 respond to all the questions.

1. Listeria EQA-8 2020-21

Dear participant,

Welcome to the eighth External Quality Assessment (EQA-8) scheme for typing of *Listeria* in 2020–2021. Please note that most of the fields are required to be filled in before the submission can be completed. Any comments can be written at the end of the form. You are always welcome to contact us at 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) and 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 IDs in the following section) – Go to 9
☐ Did not participate in the serotyping part – Go to 14.

9. Serotyping isolate IDs

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	Auto-agglutinable	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 125.

15. Cluster isolates IDs

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 22.

17. Cluster analysis based on PFGE data

18. Please list the ID for the isolates included in the cluster of closely-related isolates detected by PFGE combining *ApaI*- and *AscI*- results

Please use a semicolon (;) to separate the IDs

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. Submitting cluster results

(State one answer only)

☐ Cluster analysis based on WGS data – Go to 23
☐ Do not wish to submit any cluster results based on WGS data – Go to 125.

23. Cluster analysis based on WGS data

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

As a 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 26
☐ Allele based – Go to 33
☐ Other – Go to 25.

25. If another analysis is used, please describe your approach (including assembler, number of loci, variant caller, read mapper or reference ID etc.) – Go to 40.

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

27. Please select the approach used for the SNP analysis

(State one answer only)

☐ Reference based – Go to 28
☐ Assembly based – Go to 31.

28. 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)

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

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

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

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

33. Please select tools used for the allele analysis

(State one answer only)

☐ BioNumerics – Go to 35
☐ SeqPhere – Go to 35
☐ BIGSdb-*Lm* – Go to 35
☐ Other – Go to 34.

34. If another tool is used please enter here:

35. Please indicate allele calling method

(State one answer only)

☐ Assembly based and mapping based – Go to 36
☐ Only assembly based – Go to 36
☐ Only mapping based – Go to 37.

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

37. Please select scheme used for the allele analysis

(State one answer only)

☐ Applied Math (wgMLST) – Go to 39
☐ Applied Math (cgMLST/Pasteur) – Go to 39
☐ Pasteur (cgMLST) – Go to 39
☐ Ruppitsch (cgMLST) – Go to 39
☐ Other – Go to 38.

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

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

40. 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 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 IDs for isolates in the cluster detected with the additional analysis.

41. Please list the IDs for the isolates included in the cluster

Please use semicolon (;) to separate the IDs

42. Select a representative isolate in the cluster

Indicate the isolate ID

43. Results for 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		

44. Analysis of the EQA provided genomes

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

Please evaluate this part as a simulation, mimicking a large outbreak situation in your country.

These genomes (1-5) are very important because they might solve the outbreak.

Each of the provided genomes should be assessed as to whether it could be a part of the cluster defined in the first part.

Explain your assessment of each genome in detail, please do not just suggest rerunning the sequence, but explain what you observe and what you would suggest as the conclusion.

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

45. 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.

46. Explain your assessment of genome 1 in details

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

47. In an outbreak situation, would you consider EQA provided genome 2 as 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.

48. Explain your assessment of genome 2 in details

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

49. 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.

50. Explain your assessment of genome 3 in detail

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

51. 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.

52. Explain your assessment of genome 4 in details

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

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

(State one answer only)

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

54. Explain your assessment of genome 5 in detail

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

55. Would you like to add results performed with an additional analysis of 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 56
☐ No – Go to 95.

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

(State one answer only)

☐ SNP based – Go to 58
☐ Allele based – Go to 65
☐ Other – Go to 57.

57. If another analysis is used please describe your approach: - Go to 72.

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

59. Please select the approach used for the SNP analysis

(State one answer only)

☐ Reference based – Go to 60
☐ Assembly based – Go to 63.

60. 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)

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

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

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

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

65. Please select tool used for the allele analysis

(State one answer only)

☐ BioNumerics – Go to 67
☐ SeqPhere – Go to 67
☐ BIGSdb-*Lm* – Go to 67
☐ Other – Go to 66.

66. If another tool is used please list here:

67. Please indicate allele calling method:

(State one answer only)

☐ Assembly based and mapping based – Go to 68
☐ Only assembly based – Go to 68
☐ Only mapping based – Go to 69.

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

69. Please select scheme used for the allele analysis

(State one answer only)

☐ Applied Math (wgMLST) – Go to 71
☐ Applied Math (cgMLST/Pasteur) – Go to 71
☐ Pasteur (cgMLST) – Go to 71
☐ Ruppitsch (cgMLST) – Go to 71
☐ Other – Go to 70.

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

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

72. Additional analysis on data derived from WGS

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

(indicate the isolate ID)

74. 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		

75. 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 76
☐ No – Go to 95.

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

(State one answer only)

☐ SNP based – Go to 78
☐ Allele based – Go to 85
☐ Other – Go to 77.

77. If another analysis is used please describe your approach: - Go to 92

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

79. Please select the approach used for the SNP analysis

(State one answer only)

☐ Reference based – Go to 80
☐ Assembly based – Go to 83.

80. 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).

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

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

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

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

85. Please select tool used for the allele analysis

(State one answer only)

☐ BioNumerics – Go to 87
☐ SeqPhere – Go to 87
☐ BIGSdb-*Lm* – Go to 87
☐ Other – Go to 86.

86. If another tool is used please enter here:

87. Please indicate allele calling method:

(State one answer only)

☐ Assembly based and mapping based – Go to 88
☐ Only assembly based – Go to 88
☐ Only mapping based – Go to 89.

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

89. Please select scheme used for the allele analysis

(State one answer only)

☐ Applied Math (wgMLST) – Go to 91
☐ Applied Math (cgMLST/Pasteur) – Go to 91
☐ Pasteur (cgMLST) – Go to 91
☐ Ruppitsch (cgMLST) – Go to 91
☐ Other – Go to 90.

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

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

92. Third analysis of data derived from WGS

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

Indicate the isolate ID

94. 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		

95. Additional questions for the WGS part

96. Where was the sequencing performed?

(State one answer only)

☐ In own laboratory
☐ Externally.

97. Protocol used to prepare the library for sequencing

(State one answer only)

☐ Commercial kits – Go to 98
☐ Non-commercial kits – Go to 100.

98. Please indicate name of commercial kit:

99. If relevant please list deviation from commercial kit briefly in a few bullet points: - Go to 101

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

101. The sequencing platform used

(State one answer only)

☐ Ion Torrent PGM - Go to 103
☐ Ion Torrent Proton - Go to 103
☐ Genome Sequencer Junior System (454) - Go to 103
☐ Genome Sequencer FLX System (454) - Go to 103
☐ Genome Sequencer FLX+ System (454) - Go to 103
☐ PacBio RS - Go to 103
☐ PacBio RS II - Go to 103
☐ HiScanSQ - Go to 103
☐ HiSeq 1000 - Go to 103
☐ HiSeq 1500 - Go to 103
☐ HiSeq 2000 - Go to 103
☐ HiSeq 2500 - Go to 103
☐ HiSeq 4000 - Go to 103
☐ Genome Analyzer Iix - Go to 103
☐ MiSeq - Go to 103
☐ MiSeq Dx - Go to 103
☐ MiSeq FGx - Go to 103
☐ ABI SOLiD - Go to 103
☐ NextSeq - Go to 103
☐ MinION (ONT) - Go to 103
☐ Other - Go to 102.

102. If another platform is used please list here:

103. 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 those most frequently reported by the participants in the Listeria EQA-5 and EQA-7 scheme.

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

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

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

☐ Yes
☐ No – Go to 106.

105. Procedure used to evaluate confirmation of genus:

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

(State one answer only)

☐ Yes
☐ No – Go to 108.

107. Procedure or threshold used for coverage:

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

(State one answer only)

☐ Yes
☐ No – Go to 110.

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

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

(State one answer only)

☐ Yes
☐ No – Go to 112.

111. Procedure or threshold used for genome size:

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

(State one answer only)

☐ Yes
☐ No – Go to 114.

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

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

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

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

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

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

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

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

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

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

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

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

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

125. Comment(s):

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

126. Thank you for your participation

Thank you for filling out the submission form for the Listeria EQA-8.

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

We highly recommend documenting 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.

PDF TQ-05-22-242-EN-N

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

Publications Office of the European Union

PDF ISBN 978-92-9498-545-3