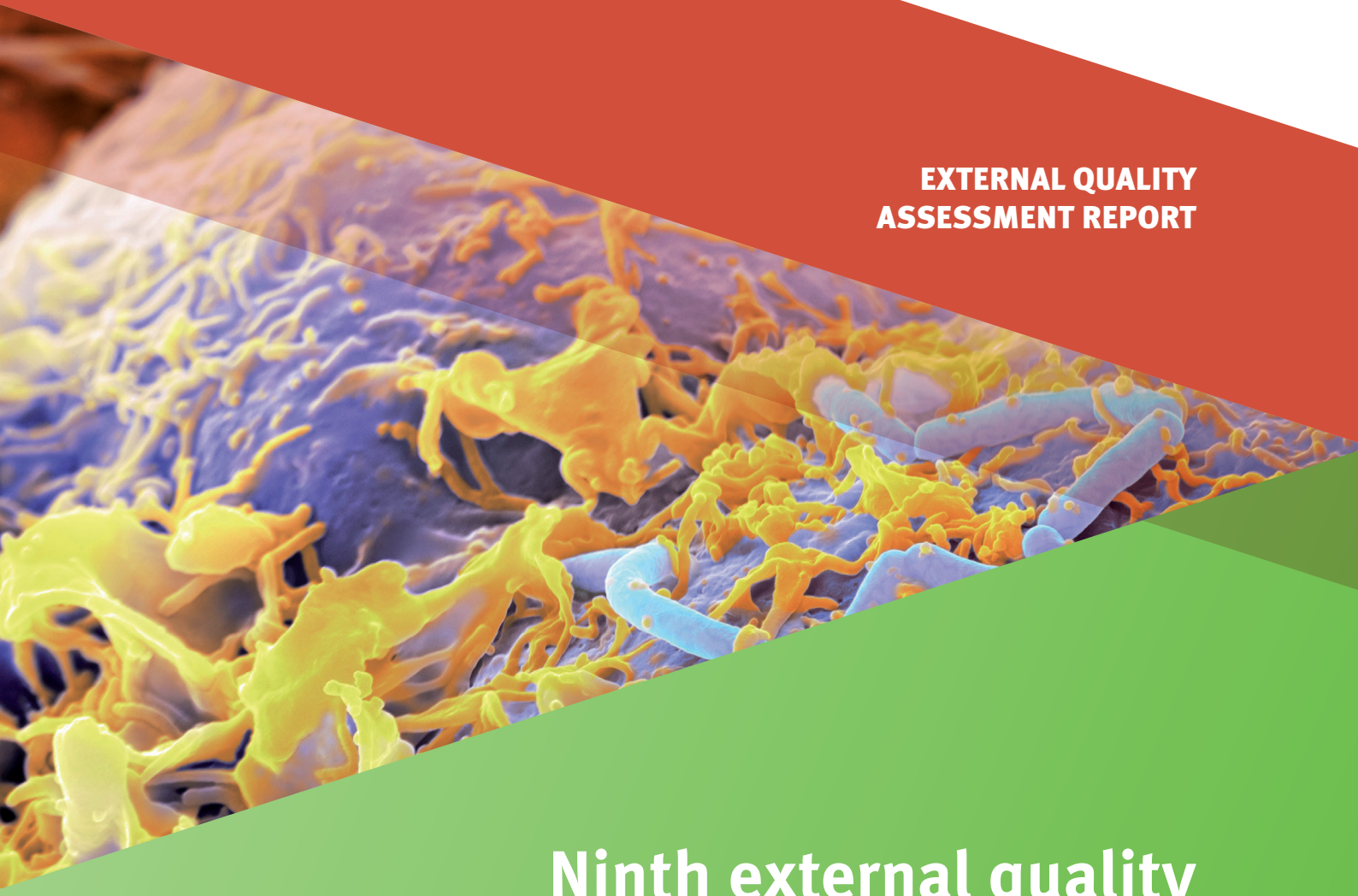
EXTERNAL QUALITY ASSESSMENT REPORT



Ninth external quality assessment scheme for *Listeria monocytogenes* typing

ECDC TECHNICAL REPORT

Ninth 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, Anne Sophie M. Uldall, 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. Ninth external quality assessment scheme for *Listeria monocytogenes* typing. Stockholm: ECDC; 2023.

Stockholm, November 2023.

ISBN 978-92-9498-667-2
doi: 10.2900/938968
Catalogue number TQ-02-23-192-EN-N

Contents

Figures

Tables

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	Quality 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 ninth round of the external quality assessment (EQA-9) 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 series of framework contracts with ECDC. EQA-9 involves 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 0.49 cases per 100 000 population in 2021 [3]. Between 2017 and 2019, the number of human listeriosis cases in the EU increased, while 2020 saw the lowest number of human cases ever reported, due to the impact of the COVID-19 pandemic and the withdrawal of the United Kingdom from the EU. However, the number of cases did not return to the pre-pandemic level in 2021.

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 objective of this EQA is to assess the quality and comparability of typing data reported by NPHRLs participating in FWD-Net. Test strains for the EQA were selected to cover strains that are currently pertinent for public health in Europe and to represent a broad range of clinically relevant types of invasive listeriosis. Seven test strains were selected for serotyping/grouping and molecular typing-based cluster. An additional ten sequences were included for the molecular typing-based cluster analysis. Twenty-one laboratories signed up and 20 completed the exercise, with one laboratory unable to setup the WGS for *L. monocytogenes* before the deadline for submission. This represented an increase of three laboratories compared to EQA-7 and EQA-8. The majority of participants (15/20; 80%) completed the full EQA scheme.

In total, 18 (90%) laboratories participated in the serotyping part. Molecular serogrouping results were provided by 17 of 18 (94%) participants. Four participants performed both conventional serotyping and molecular serogrouping and one only performed the conventional serotyping. The average performance for molecular serogrouping was high, with 96% correct results. For the conventional method, 80% of the participants correctly serotyped all seven test strains. One participant mistyped five of the seven strains in the conventional serotyping, but achieved 100% correct results in the molecular serogrouping, which they performed for the first time. Since the first EQA in 2012, the trend has been towards replacing conventional serotyping with molecular serogrouping, with strong performance; however, in EQA-9 three participants using WGS incorrectly determined the serogroup for one of the 'repeat' strains.

Of the 20 laboratories participating in the EQA-9, 17 (85%) 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 strains regardless of the method used, rather than being able to follow a specific procedure.

The cluster of closely-related strains 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. pulsed-field gel electrophoresis (PFGE)). Neither of the two participants that used PFGE identified the cluster correctly and they could not include the sequences provided in their analysis. Sixteen laboratories performed cluster analysis using WGS-derived data. In general, performance was strong, with 81% of the participants correctly identifying the cluster of closely-related strains when assessing good quality data (QC-status A).

The modified sequence (strain17, cluster strain with coverage reduced to approximately 94%) did interfere with the analysis for some participants, who incorrectly found that this strain was genetically too distant to be part of the cluster. This EQA clearly shows that the effect of differences between the approaches and schemes used by the participants is amplified when assessing data of non-perfect quality. Each laboratory should test and find standard Quality Control (QC) thresholds that work for their approach and scheme. The laboratories assigned a different QC-status for some of the modified strains, and the results affected the analysis and conclusions.

All participants were able to recognise the modified sequence with very reduced coverage; however, some participants could not recognise the sequence with contamination and the sequence with a mix of two different sequences types (STs). The preferred analysis method was the allele-based core genome multilocus sequence typing (cgMLST) used by 88% (14/16), while only 13% (2/16) used single nucleotide polymorphism (SNP). The most widely used scheme for the EQAs was the Ruppitsch (cgMLST) (11/15), while the Pasteur scheme (cgMLST) was used by 31% (5/16) of the participants in EQA-9. The conclusion from this EQA-9 was that cgMLST has a higher consistency than SNP analysis as the results were much more comparable.

The current EQA scheme for *L. monocytogenes* typing is the ninth EQA organised for NPHRLs 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 five years, the public health institutes have had the possibility to routinely submit WGS variables for *L. monocytogenes* to TESSy to be used for EU-wide WGS-enhanced listeriosis surveillance. In addition to the WGS data submitted by the Member States for the listeriosis cluster detection, ECDC also collects and analyses sequence data during multi-country outbreak investigations that are primarily initiated by Member State public health institutes through EpiPulse.

A feedback survey was sent to assess the *Listeria* EQA scheme; the questionnaire contained questions related to accreditation and information on the individual report; 13/20 responded. The usefulness of the QC evaluation of participant-sequenced data and the helpfulness of including low quality data was appreciated by 100% of respondents.

1. Introduction

1.1 Background

ECDC is an EU agency whose mission is to identify, assess and communicate current and emerging threats to human health from communicable diseases. ECDC's founding regulation outlines its mandate as fostering the development of sufficient capacity within EU/EEA dedicated surveillance networks for diagnosis, detection, identification and characterisation of infectious agents that may threaten public health. Under this mandate ECDC 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 quality assessment purposes.

ECDC has outsourced the organisation of EQA schemes for EU/EEA countries in the disease networks. The aim of EQAs is 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 to 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');
- to assess the effects of analytical procedures (method principle, instruments, reagents, calibration);
- to support method development;
- to evaluate individual laboratory performance;
- to identify problem areas;
- to provide continuing education;
- to 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 2021, SSI won the new round of tenders (2022–2025) for *Listeria* and STEC. The *Listeria* EQA covers serotyping and molecular typing-based cluster analysis of. This report presents the results of the *Listeria* EQA-9.

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. From 2017 to 2019, the number of human listeriosis cases increased slightly in the EU (2 474–2 621). However, in 2020, the number of confirmed human listeriosis cases decreased (1 887) and in 2021 the number of cases (2 183) in the EU did not return to pre-pandemic levels. Nevertheless, during the period 2017–2021, the notification rate was stable (0.47–0.49 cases per 100 000 population) [3].

One of ECDC's key objectives 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 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;
- to facilitate the detection and investigation of transmission chains and relatedness of strains across EU/EEA countries and contribute to global investigations;
- to detect the emergence of new and/or evolving pathogenic strains;
- to support investigations to trace the source of an outbreak and identify new risk factors;
- to 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 strains characterised by molecular typing at the national level are part of a multinational cluster that may require cross-border response collaboration.

1.3 Objectives of the EQA-9 on *Listeria*

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

1.3.1 Serotyping

The EQA-9 scheme assessed serotype determination by either conventional antigen-based typing of somatic 'O' antigens and flagellar 'H' antigens, or PCR-based/WGS molecular serogrouping.

1.3.2 Molecular typing-based cluster analysis

The objective of *L. monocytogenes* EQA-9 was to assess the ability to detect a cluster of closely related strains. Laboratories could perform analysis using PFGE and/or derived data from WGS. The cluster analysis was to be conducted on the seven test strains and ten additional test strains (provided genomic sequences). Some of the sequences provided were modified to have QC issues.

2. Study design

2.1 Organisation

Listeria EQA-9 was funded by ECDC and arranged by SSI following ISO/IEC 17043:2010 [4]. EQA-9 included serotyping and molecular typing-based cluster analysis and was carried out during the period July–December 2022.

Invitations were emailed by ECDC to the contact points in the FWD-Net (30 countries) on 7 April 2022, with a deadline to respond by 25 April 2022. In addition, invitations were sent to the EU candidate countries Albania, Bosnia and Herzegovina, Kosovo[1], Montenegro, North Macedonia, Serbia and Türkiye.

Twenty-one NPHRLs in EU/EEA and EU candidate countries accepted the invitation to participate and 20 submitted results (Annex 1). In Annex 2, details of participation in EQA-8 and EQA-9 are listed to give an overview of the number of participants. The EQA test strains were sent to participants on 17 May 2022. Participants were asked to submit their results by 30 August 2022 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 strains/sequences

Ten candidate strains were analysed using the methods set out in the EQA (serotyping and WGS) before and after re-culturing. All candidate strains remained stable using these methods and a final selection of seven test strains; five test strains and a set of technical duplicates (twice from the same culture) were made. In addition, ten sequences (representing the genomes of 10 additional test strains) were provided for the participants to include in the cluster analysis, and four of the sequences were modified by the EQA provider to have various QC issues.

Seventeen *L. monocytogenes* test strains/sequences were selected to fulfil the following criteria:

- cover a broad range of the commonly-reported, clinically-relevant strains of invasive listeriosis in Europe;
- include genetically closely related strains;
- remain stable during the preliminary test period at the organising laboratory;
- include three 'repeat strains' from EQA-1 to EQA-9; and
- include a set of technical duplicates in the serotyping/grouping.

The seven test strains for serotyping were selected to cover different serotypes/-groups (1/2a/IIa, 1/2b/IIb, 1/2c/IIc, and 4b/IVb). This year all test strains and sequences provided had to be assessed in the cluster analysis. The test strains and the sequences provided had different 7-gene Multi-Locus Sequence Types (ST) (ST2, ST3, ST7, ST9 and ST1504) and varied in relatedness.

To follow the development of each laboratory's performance (the reproducibility), three strains of different serotypes/groups were included in EQA-1 to 9: Strain1 (4b - IVb), Strain2 (1/2a –IIa) and strain3 (1/2c-IIc) Based on the WGS-derived data, the selected cluster of closely-related strains consisted of six *L. monocytogenes* ST7 strains (including the technical duplicate set strain2/strain6 and provided sequence strain9). Characteristics of all the *L. monocytogenes* test strains are listed in Table 1 and in Annexes 3, 7, 8, and 11. The EQA provider found at most three allele differences or eight SNPs between any two strains in the cluster (Annex 8). The EQA provider's cluster analysis of WGS-derived data was based on an allele-based (cgMLST [6]) and SNP analysis (NASP [7]). Those participants using PFGE as a cluster method could only evaluate the seven test strains from the package and only two belonged to the cluster of closely related strains based on WGS. The sequences provided represented 5 ST7 strains, one ST1504 strain and four modified sequences; one ST7 mixed with 9% *L. innocua,* one was a mix of sequences of *L. monocytogenes* ST1 and ST224. The last two were genomes with massively reduced coverage and slightly reduced coverage, respectively (Table 1).

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

Table 1. Characteristics of test strains and sequences

Method	Serotyping		Cluster analysis			
Number of strains/sequences	7 strains		7 strains/10 sequences			
Annex	3		4, 6–8			
Strain number				ST	QC-status	Part of the pre-defined Cluster
Strain1∞	Strains for serotyping analysis	4b / IVb	Strains/sequences for cluster analysis	2	-	
Strain2#∞		1/2a / IIa		7	-	Yes
Strain3∞		1/2c / IIc		9	-	
Strain4		1/2a / IIa		7	-	
Strain5		1/2a / IIa		7	-	
Strain6#		1/2a / IIa		7	-	Yes
Strain7		1/2b / IIb		3	-	
Strain8^	-	-		-	B/C	NA
Strain9#	-	IIa		7	A	Yes
Strain10	-	IIa		7	A	
Strain11	-	IIa		1504	A	
Strain12^	-	-		-	C	NA
Strain13	-	IIa		7	A	
Strain14	-	IIa		7	A	Yes
Strain15^	-	-		-	C	NA
Strain16	-	IIa		7	A	Yes
Strain17^	-	IIa		7	B	Yes

#:Technical triplicates were strain2/strain6/strain9-sequence (Annex 3, 6 and 8)
∞: 'Repeat strains' included in EQA-1 to 9 (strain1, strain2 and strain3). Strain2 was a different strain to that used in previous years, although it was the same serotype/group.
^: Modified sequences:
Strain8, A non-cluster sequence contaminated with approx. 9% L. innocua.
Strain12, A non-cluster sequence with massively reduced coverage and removal of genes.
Strain15, Two non-cluster sequences of ST1 and ST224 combined.
Strain17, A cluster sequence (strain14) with reduced core percent from 100% to 94%.
ST: sequence type.
NA: Not applicable, quality too poor to include in the cluster analysis.

2.3 Distribution of strains and sequences

The seven test strains were blinded and shipped on 17 May 2022. The protocol for the EQA exercise and a letter stating the unique strain IDs were included in the packages which were distributed individually to the participants by email on 17 May 2022 as an additional precaution. The packages were shipped from SSI, labelled 'UN3373 Biological Substance'. Fourteen participants received their dispatched strains within two days, five within three days, one received them six days after shipment and another seven days after shipment. No participants reported damage to the shipment or errors in the unique strain IDs.

On 9 June 2022, instructions regarding the procedure for submitting results were emailed to the participants. This included the link to the online site for uploading sequences, downloading the additional test strain 8-17 (FASTQ genomic sequences) and the empty submission form.

2.4 Testing

In the serotyping part, seven *L. monocytogenes* strains 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]) or in silico PCR by using WGS data. The results of serotyping/-grouping were submitted in the online form.

In the cluster analysis part, participants could choose to perform using PFGE (*Apa*I and *Asc*I profiles) and/or WGS-derived data. The participants were instructed to report the IDs of the strains included in the cluster of closely-related strains by method. Laboratories performing WGS could use their own analysis pipeline for cluster analysis – e.g. single nucleotide polymorphism analysis (SNP-based) or whole/core genome multilocus sequence typing (wgMLST)/cgMLST (allele-based). The participants were asked to report the number of loci in the allelic scheme used for cluster analysis and/or the name of the SNP pipeline used.

The participants were asked to report the strains identified as a cluster of closely related strains 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. The results were reported as SNP distance or allelic difference between each test strain and a strain (strain9) selected by the EQA provider. In addition, the 7-gene Multi Locus Sequence Types (ST) and the serotype of strains in the cluster analysis could be submitted.

In addition, each participant needed to assess the QC quality of the provided sequences (four manipulated by the EQA provider). The three possible QC categories were; A: acceptable quality, B: quality only acceptable for outbreak situations (less good quality) and C: unacceptable quality - strain not analysed. The participants were instructed to describe their QC observations and considerations leading to the QC-status decision. The EQA-provider had modified four sequences (strain8, strain12, strain15 and strain17) (see Table 6, Annex 11).

The laboratories uploaded the raw reads (FASTQ files) for further analysis by the EQA provider.

2.5 Data analysis

The submitted serotyping and cluster analysis results, as well as the raw reads, were imported to a dedicated *Listeria* EQA-9 BioNumerics (BN) database. The EQA provider contacted three laboratories as they did not submit the raw reads by the deadline. One additional laboratory was contacted as one of the sequences was uploaded with incomplete data.

Serotyping results were evaluated according to the percentage of correct results, generating a score from 0–100%.

Molecular typing-based cluster analysis was evaluated as correct or incorrect identification of the expected cluster of closely-related strains, according to a pre-defined categorisation by the organiser based on WGS-derived data. This categorisation was obtained by allele-based analysis (cgMLST [6] and SNP analysis (NASP, [7])). The correct number of closely-related *L. monocytogenes* strains/sequences by WGS was six ST7 strains: strain2, strain6, strain9, strain14, strain16 and strain17. The provider included sequences for strain17 with the QC-status B in the analysis. If only assessing 'Acceptable quality' data (QC-status A), strain17 was excluded from the analysis. Strain2/6/9 were from the same culture and were sent to participants as two strains and a provided sequence. The coverage of strain17 was slightly reduced by the EQA provider. The EQA provider found at most three allele differences or eight SNPs (Annex 8). The EQA provider's cluster analysis of WGS-derived data showed three allele differences or eight SNPs between any two strains in the cluster. Allele-based (cgMLST [6]) and SNP analysis (NASP [7]). The participants using PFGE only evaluated test strains 1-7 and only two belonged to the cluster of closely related strains based on WGS.

The participant's descriptions and the QC-status of the EQA-provider's modified sequences are listed in Annex 11.

Individual evaluation reports were distributed to participants in November 2022 and certificates of attendance in February 2023. If WGS data was 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 contained comments on the QC-status of the submitted sequences.

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 21 participants that signed up, 20 completed and submitted their results. The majority of the participants (75%, 15/20) completed the full EQA scheme. In total, 18 (90%) participants participated in serotyping and 17 (85%) in cluster analysis. Conventional serotyping results were provided by 28% (5/18) of the participants and four of these laboratories also performed molecular serogrouping. Molecular serogrouping results were provided by 17 (94%) participants.

Most participants (88%: 15/17) reported cluster analysis using only WGS-derived data, while one (6%) reported using only PFGE data and one (6%) submitted cluster data based on both PFGE and WGS (Table 2).

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

	Serotyping				Cluster analysis			
	Conventional only	Molecular only	Both	Total	PFGE only	WGS only	Both	Total
Number of participants	1	13	4	18	1	15	1	17
Percentage of participants	6	72	22	90*	6	88	6	85*

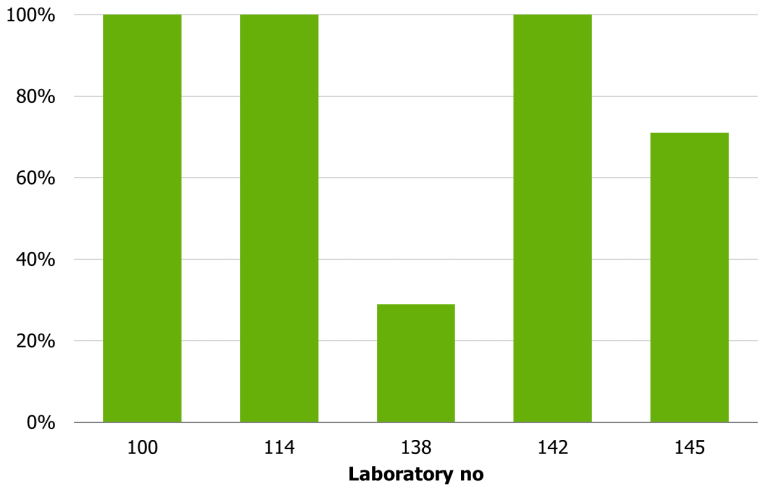
Fifteen of the 20 participants (75%) completed both parts (serotyping and cluster analysis) of the EQA.
** Percentage of the total number of participating laboratories (20).*

3.2 Serotyping

3.2.1 Conventional serotyping

Five participants performed conventional serotyping of *L. monocytogenes* (Figure 1). Performance was strong (80% correct) as three of the participants correctly serotyped all seven test strains, including one laboratory (114) that had not participated since EQA-4. One laboratory (138) once again had major issues as it was only able to correctly serotype two of the seven strains. Laboratory 145 had two incorrect serotypes, reported 1/2a instead of 1/2c for strain3 and for strain7 reported 4b instead of 1/2b.

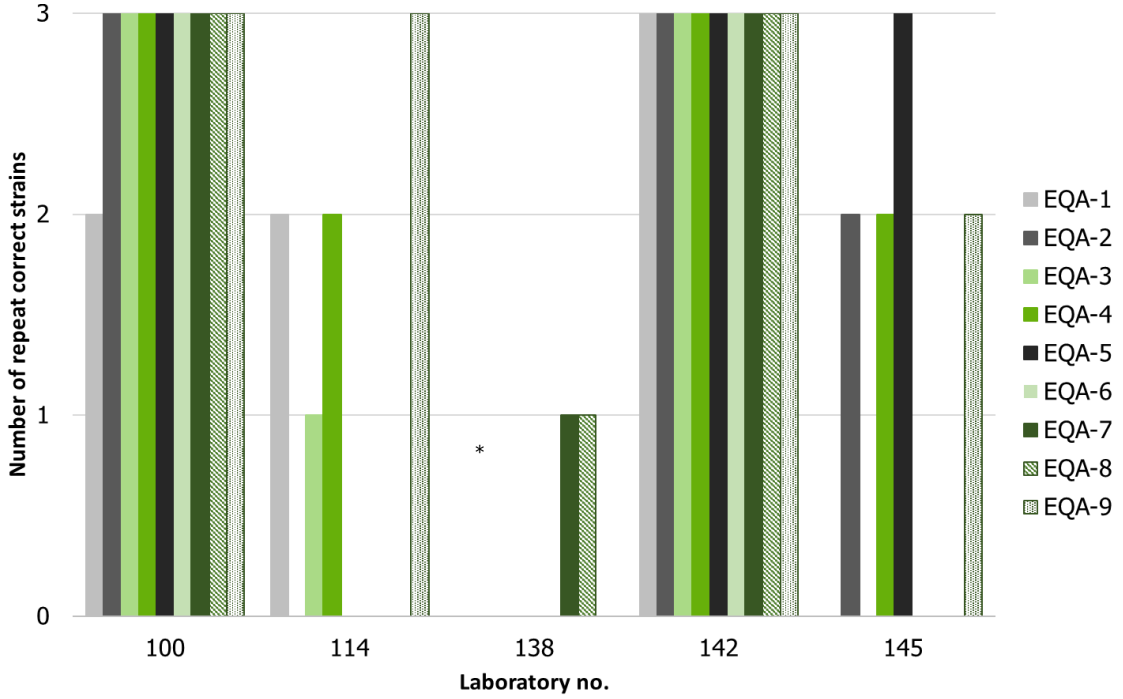
Figure 1. Participant scores for conventional serotyping of seven test strains



Arbitrary numbers represent participating laboratories. Bars represent the percentage of correctly assigned serotypes for the seven test strains (strain1-7).

Figure 2 and Table 3 shows the reproducibility of the individual participants' performances in conventional serotyping of the three 'repeat strains' from EQA-1 to EQA-9. Only laboratories participating in EQA-9 are shown. The reproducibility of conventional serotyping results for the repeat strains shows stability and strong performance for two of the laboratories participating every year (laboratories 100 and 142). However, laboratory 114 also showed strong performance despite not having participated in EQAs for some years. Laboratory 138, participating for the third time, did not report serotyping results correctly for any of the repeat strains in EQA-9 and for only one in EQA-7 and EQA-8. Both in EQA-2 and EQA-9, laboratory 145 reported the `repeat strain3´ incorrectly as 1/2b and 1/2a respectively, even though this strain (1/2c) has been the same throughout all the EQAs. In EQA-5 laboratory, 145 reported the correct serotype for repeat strain3.

Figure 2. Correct conventional serotyping of three repeat strains through EQA-1 to 9 for laboratories participating in EQA-9



Arbitrary numbers represent the participating laboratories. Bars represent the number of correctly assigned serotypes for the three repeat strains (strain1, strain2 and strain3).
**: Laboratory did not correctly identify any of the three repeated strains.*

Table 3. Correctly assigned conventional serotypes for three repeat strains through EQA-1 to 9 for laboratories participating in EQA-9

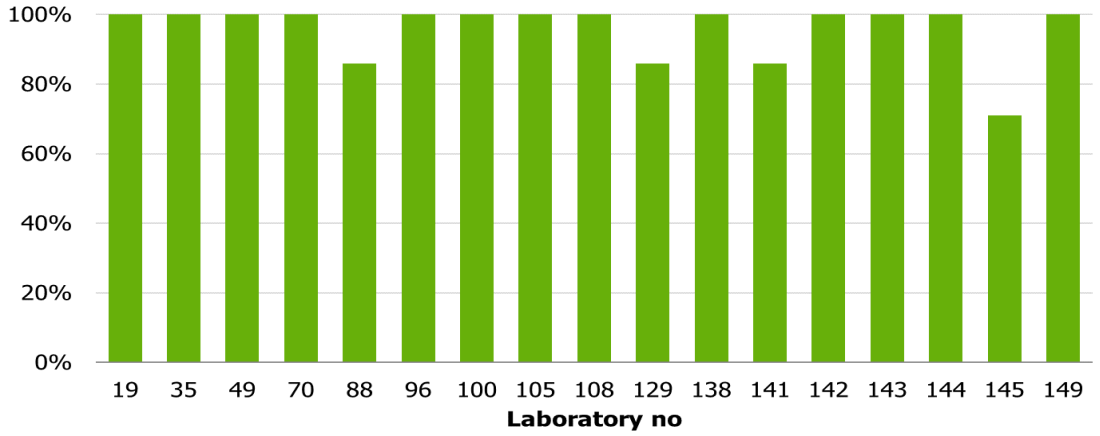
	Laboratory ID				
EQA round	**100**	**114**	**138**	**142**	**145**
EQA-1	2	2		3	
EQA-2	3			3	2
EQA-3	3	1		3	
EQA-4	3	2		3	2
EQA-5	3			3	3
EQA-6	3			3	
EQA-7	3		1	3	
EQA-8	3		1	3	
EQA-9	3	3	0	3	2

Arbitrary numbers represent the participating laboratories.
Number of correctly assigned serotypes for the three repeat strains (strain1, strain2 and strain3)
Empty field: did not participate in that round.

3.2.2 Molecular serogrouping

Seventeen participants performed molecular serogrouping of *L. monocytogenes* in EQA-9 (Figure 3). Molecular serogrouping was carried out in accordance with guidelines by Doumith et al. [5] and nomenclature from Doumith et al. [8] was used. Thirteen (77%) participants were able to correctly serogroup all seven EQA test strains, the average performance was 96%. Ten of the 17 participants reported using WGS-based analysis (in silico PCR) for molecular serogrouping. The majority (4/5) of errors were in strain3, with four different laboratories reporting IIa instead of IIc, three of which (88, 129 and 141) were using WGS-based data (in silico PCR). Only Laboratory 145 submitted serogrouping results using PCR.

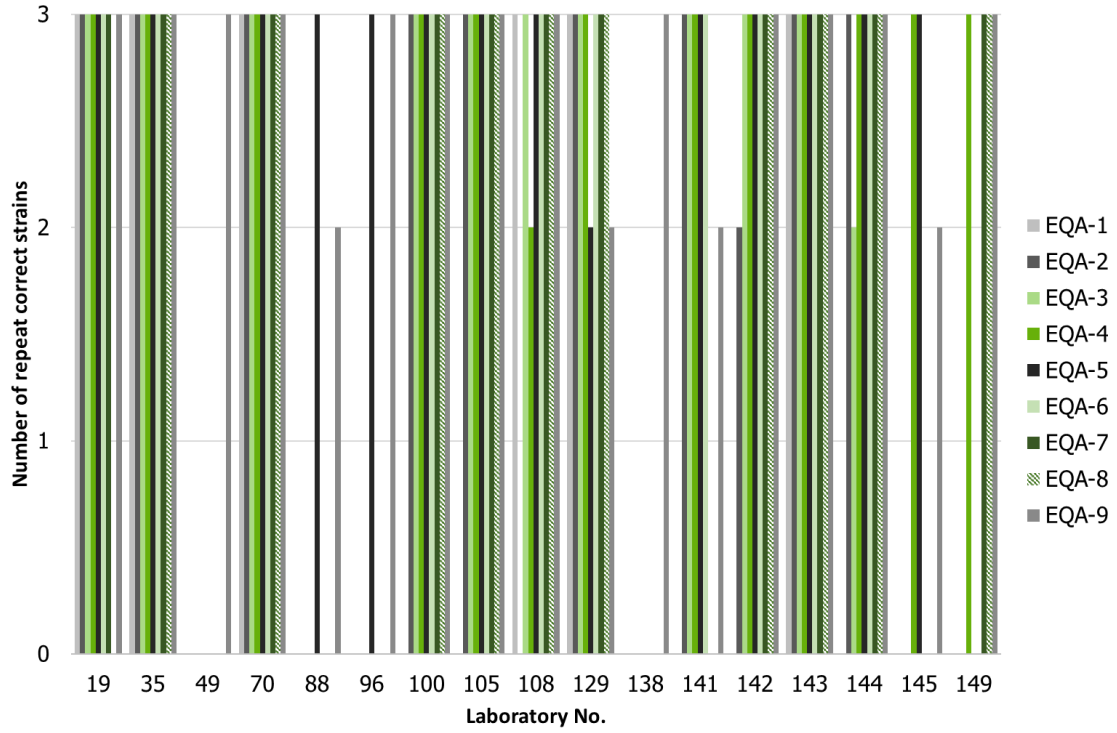
Figure 3. Participant scores for molecular serogrouping of seven *L. monocytogenes* test strains



Arbitrary numbers represent the participating laboratories. Bars represent the percentage of correctly assigned serogroups for the seven test strains (strain1-7).

Figure 4 and Table 4 shows the individual reproducibility of participants' performances in molecular serogrouping when assessing the three repeat strains during the nine EQAs. Of the 17 laboratories that participated in EQA-9, 9/17 (53%) correctly serogrouped all three repeat strains in all the EQA rounds they participated in. In EQA-9 more errors were seen in the repeat strain3 than in previous years. The errors were submitted by four laboratories (88, 129, 141 and 145), three of which had not previously had any errors in the repeat strains.

Figure 4. Correct molecular serogrouping of three repeat strains from EQA-1 to 9 for laboratories participating in EQA-9



Arbitrary numbers represent the participating laboratories. Bars represent the number of correctly assigned serotypes for the three repeat strains (strain1, strain2 and strain3).

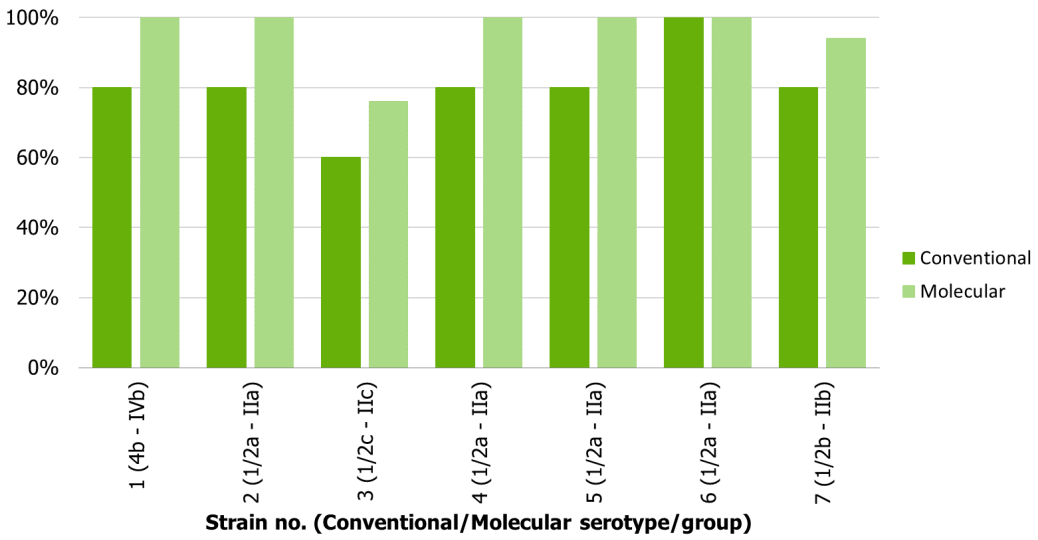
Table 4. Correctly assigned molecular serotypes for three repeat strains through EQA-1 to 9 for laboratories participating in EQA-9

	Laboratory ID																
EQA round	19	35	49	70	88	96	100	105	108	129	138	141	142	143	144	145	149
EQA-1	3	3			3					3	3					3	
EQA-2	3	3		3	3			3	3		3			3	2	3	3
EQA-3	3	3		3	3			3	3	3	3			3	3	3	2
EQA-4	3	3		3	3	0		3	3	2	3			3	3	3	3
EQA-5	3	3		3	3	3	3	3	3	3	2	3		3	3	3	3
EQA-6	3	3		3	3			3	3	3	3	2		3	3	3	3
EQA-7	3	3		3	3			3	3	3	3	3			3	3	3
EQA-8		3		0	3			3	3	3	3	3			3	3	3
EQA-9	3	3	3		3	2	3	3	3	3	2		3	2	3	3	3

Arbitrary numbers represent the participating laboratories.
Number of correctly assigned serotypes for the three repeat strains (strain1, strain2 and strain3)
Empty field: no participation in that round

Figure 5 shows the reported error distributed per strains. Six participants reported an error for strain3, two (138 and 145) using the conventional method (reporting 1/2a or 3c) and four (88, 129, 141 and 145) using molecular method (WGS/PCR) all reporting IIa instead of IIc. The additional errors seen in the conventional serotyping were primarily reported by one laboratory 138 and the additional error (IVb instead of IIb) in the molecular method for strain7 was reported by laboratory 145, which also was reported incorrectly in the conventional serotyping (4b instead of 1/2b).

Figure 5. Average score of 7 test strains



Bars represent the percentage of correctly assigned serotypes/-groups by the participants
Conventional serotyping N=5, Molecular serogrouping N=17.

3.3 Molecular typing-based cluster analysis

Participants were tested on their ability to correctly identify the cluster of closely related strains defined by pre-categorisation from the EQA provider among the 7 test strains and 10 provided sequences.

3.3.1 PFGE-derived data

Two (2/20, 10%) participants performed cluster analysis using PFGE-derived data. The cluster categorisation was based on WGS data including the provided sequences, therefore, as expected, the correct cluster delineation was difficult to obtain using a less discriminatory method. Both participants identified the two strains among the seven test strains as belonging to a cluster, however the laboratories reported that additional one or two strains, respectively, belonged the cluster.

Table 5 shows the overview of the strains that each participant included or excluded in their cluster identification.

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

Strains		Laboratory ID	
Strain number	**ST**	**114**	**142**
Strain1	ST2	No	No
Strain2‡#	ST7	Yes	Yes
Strain3	ST9	No	No
Strain4	ST7	Yes	No
Strain5	ST7	Yes	Yes
Strain6‡#	ST7	Yes	Yes
Strain7	ST3	No	No
Included the two strains from the WGS cluster*		**Yes**	**Yes**
Included additional strains in the cluster		**Yes**	**Yes**

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

3.3.2 WGS-derived data

3.3.2.1 Details reported on equipment and method

Sixteen participants (16/17, 94%) performed cluster analysis using WGS-derived data. One laboratory reported using an external laboratory for sequencing, while 94% used their own laboratory. Different sequencing platforms were listed by the participants: one MiniSeq, six MiSeq, seven NextSeq, one Ion GeneStudio S5 System and one Ion Torrent. All reported using commercial kits for library preparation. Of the 16 participants, twelve (75%) used Illumina's Nextera kit. Two participants had modified the manufacturer's' protocol by changing volumes (Annex 5).

3.3.2.2. Assessment of the QC status of the sequences provided

The participants were instructed to describe their QC observations and considerations leading to the QC status decision and following cluster analysis for the sequences provided (strain8-17). The three levels of QC status were A: acceptable quality, B: quality only acceptable for outbreak situations (less good quality) and C: unacceptable quality - strain not analysed. The EQA-provider had modified four sequences (strain8, strain12, strain15 and strain17). Table 6, Annex 11.

All the sequences provided without modification were reported as acceptable quality (QC status A) by the participants.

For strain8 (a non-cluster sequence contaminated with approximately 9% *L. innocua*) 69% of the participants (11/16) correctly observed the contamination of the sequence and reported a QC status of B or C. Five participants reported the sequence to be of acceptable quality (QC status A). However, one of these five participants did describe the contamination and another reported that it might be a new ST belonging to CC7 as they were unable to get the ST. Six participants included the strain/sequence in their cluster analysis.

For strain12 (A non-cluster sequence with massively reduced coverage and removal of genes) all participants (16/16) correctly identified the poor quality of the sequence, and excluded the sequence from the cluster analysis.

For strain15 (two non-cluster sequences of ST1 and ST224 combined into one FASTQ file), 75% of the participants (12/16) correctly observed the mix STs and an enlarged genome size and reported a QC status C. Two participants (Labs 56 and 108) did not identify the enlarged size (reported QC status A). Two participants (70 and 105) determined that the QC was not optimal (they reported a warning of only 93 cgMLST% or a high number of contigs), but accepted the sequences for outbreak detection (QC status B).

For strain17 (a cluster sequence (strain14) with reduced core percent from 100% to 94%), 44% of the participants (7/16) reported the sequence to be of acceptable QC status A and 44% (7/16) reported the sequence as only acceptable for outbreak investigation. The last two (Labs 105 and 144) reported the sequence to be unacceptable for analysis (QC status C) and both participants reported that the percentage of good cgMLST targets was below their threshold of 95%. These two participants therefore did not analyse the sequence in the cluster analysis.

Table 6. Results of participants' QC assessment of the modified sequences provided for the EQA

Sequences	Characteristics	Provider	A	B	C
Strain8	A non-cluster sequence contaminated with approx. *9% L. innocua*	B/C	5	1	10
Strain12	A non-cluster sequence with massively reduced coverage and removal of genes	C	0	0	16
Strain15	Two non-cluster sequences of ST1 and ST224 combined	C	2	2	12
Strain17	A cluster sequence (strain14) with reduced core percent from 100% to 94%	B	7	7	2

Raw data available in Annex 11.
QC status: acceptable quality (A), quality only acceptable for outbreak situations (less good quality) (B) and unacceptable quality - strain not analysed (C).

3.3.2.3. Cluster analysis results

Each participant had to use their own produced sequences and the sequences provided (after assessment of QC status) in the cluster analysis and report which strains/sequences were part of the cluster of closely-related strains, thereby mimicking an urgent outbreak situation, where it is impossible to re-run the sequencing and sequences sometimes have to be assessed despite being poor quality.

In general, the performance was strong in cluster analysis with WGS-derived data, however for the modified sequence of strain17, the results were divided (see Table 6/Table 7). Thirteen participants (81%) correctly identified the cluster of closely-related strains without the modified strain with QC status B (strain2, strain6, strain9, strain14 and strain16). Only seven participants included the modified strain17 in the cluster of closely-related strains (Table 7). Four of the seven laboratories that included this strain in the analysis obtained too high an AD difference in their analysis to include strain17 in the cluster. The reduced coverage of strain17 (94cgMLST core percent in BioNumerics/Pasture) led to two laboratories excluding it from the analysis due to poor quality (Lab 105 and 144).

Laboratories were instructed to report the data analysis used for cluster identification using a sequence provided (strain9) as reference for reporting SNP distance or allelic differences. Laboratories could report results from one main and one-to-two additional analyses, but the detected cluster had to be based on results from the main analysis. Only two participants reported additional analyses (Laboratories 19 and 100).

Table 7. Results of cluster analysis based on WGS-derived data

		Laboratory ID															
Strain No.	ST	19	35	49	56	70	88	100	105	108	129	135	138	141	142	144	149
Strain1	2	No	No	No	No	No	No	No	No	No	No	No	No	No	No	No	No
Strain2#‡	7	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Strain3	9	No	No	No	No	No	No	No	No	No	No	No	No	No	No	No	No
Strain4	7	No	No	No	No	No	No	No	No	No	No	No	Yes	No	No	No	No
Strain5	7	No	No	No	No	No	No	No	No	No	No	No	Yes	No	No	No	No
Strain6#‡	7	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Strain**7**	3	No	No	No	No	No	No	No	No	No	No	No	No	No	No	No	No
Strain8^	-	No	No	No	No	No	No	No	No	No	No	No	No	No	No	No	No
Strain9#‡	7	Yes	Yes	Yes	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Strain10	7	No	No	No	No	No	No	No	No	No	No	No	Yes	No	No	No	No
Strain11	1504	No	No	No	No	No	No	No	No	No	No	No	Yes	No	No	No	No
Strain12^	-	No	No	No	No	No	No	No	No	No	No	No	No	No	No	No	No
Strain13	7	No	No	No	No	No	No	No	No	No	No	No	Yes	No	No	No	No
Strain14‡	7	Yes	Yes	Yes	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Strain15^	-	No	No	No	No	No	No	No	No	No	No	No	No	No	No	No	No
Strain16‡	7	Yes	Yes	Yes	No	Yes	Yes	Yes	Yes	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Strain17‡^	7	Yes	No	Yes	No	Yes	No	Yes	No	No	Yes	No	Yes	Yes	No	No	Yes
Main analysis		Allele	Allele	Allele	Allele	Allele	Allele	Allele	Allele	SNP	Allele	Allele	SNP	Allele	Allele	Allele	Allele
1. additional		SNP						SNV									
Cluster (without QC B)*		Yes	Yes	Yes	No	Yes	Yes	Yes	Yes	No	Yes	Yes	No	Yes	Yes	Yes	Yes
Cluster identified		Yes	No	Yes	No	Yes	No	Yes	No	No	Yes	No	No	Yes	No	No	Yes

‡: closely-related strains (in grey)
#: technical triplicates (strains or sequence) (in bold)
ST: 7 multilocus sequence type
Allele: allele-based analysis (cgMLST) (Annex 7)
SNP: single-nucleotide polymorphism (Annex 7)
SNV: single-nucleotide variant
^: modified sequences
**: Cluster (without the modified strain17 QC-status B)*

Of the 16 participants, two (13%) used SNP as the main analysis, one used an in-house pipeline and one a published pipeline. Both used a reference-based approach with strain9 (provided) as reference. One used CLC for both read mapper and variant caller, and the other used Burrows-Wheeler Aligner (BWA) as the read mapper and GATK as the variant caller.

Tables 8 and 9 show the overview of the submitted data. Each laboratory reported SNP distances/allelic differences by strain (see Annex 8).

Table 8. Results of SNP-based cluster analysis

Laboratory	SNP-based analysis								
	SNP pipeline	Approach	Reference	Read mapper	Variant caller	Identified pre-defined cluster	Difference reported within cluster (SNP)	Identified cluster without strain17	Included strain17 (difference to strain17 in SNP)
Provider	NASP [8]	Rb	strain9	BWA	GATK	Yes	0-8	Yes	Yes (6)
19*	NASP	Rb	strain9	BWA	GATK	Yes	0-8	Yes	(Yes/6)
100*	§ SNV analysis in SeqSphere based on read mapping assembly	Rb				Yes	0-4	Yes	(Yes/3)
108	In-house pipeline	Rb	strain9	CLC assembly cell	CLC assembly cell	No	0-128	No	(No/163)
138	cfsan-snp-pipeline	Rb	strain9	Bowtie2	GATK	No	0-280	No	(No/21)

**: additional analysis.*
§: modified from submitted information.
Rb: Reference based.
For detailed data see Annex 8.

Fourteen of the 16 participants used allele-based analysis as the main analysis for cluster detection (Table 9). Eleven (79%) only used an assembly-based allele calling method, two used both mapping and assembly-based allele calling and one used only a mapping-based allele calling method (Table 9). All 14 reported using cgMLST, eight (57%) used cgMLST Ruppitsch (1701 loci) [9], five cgMLST Pasteur (1748 loci) and one an in-house cgMLST scheme with only 1503 loci.

Table 9. Results of allele-based cluster analysis

Laboratory	Allele-based analysis								
	Approach	Allelic calling method	Assembler	Scheme	No. of loci	Identified pre-defined cluster	Difference reported within cluster (AD)	Identified cluster without strain17	Included strain17 (difference to strain17 in AD)
Provider	BioNumerics	*A&M*	SPAdes	Applied Math (cgMLST/Pasteur)	1748	Yes	0-3	Yes	(Yes/1)
19	BioNumerics	*A&M*	SPAdes	Applied Math (cgMLST/Pasteur)	1748	Yes	0-3	Yes	(Yes/1)
35	SeqPhere	*OAB*	Velvet	Ruppitsch (cgMLST)	1701	No	0-9	Yes	(No/31)
49	BioNumerics	*OAB*	SKesa	Applied Math (cgMLST/Pasteur)	1748	Yes	0-3	Yes	(Yes/1)
56	MentaList	*OMB*	-	Pasteur (cgMLST)	1748	No	128-129	No	(No/1015)
70	SeqPhere	*A&M*	SKESA	Ruppitsch (cgMLST)	1701	Yes	0-4	Yes	(Yes/2)
88	§INNUca4.2.2 chewBBACA 2.8.5	*OAB*	SPAdes INNUca4.2.2	Pasteur (cgMLST)	1748	No	0-4	Yes	(No/12)
100	SeqPhere	*OAB*	SKESA	Ruppitsch (cgMLST)	1701	Yes	0-3	Yes	(Yes/2)
105	SeqPhere	*OAB*	SPAdes	Ruppitsch (cgMLST)	1701	No	0-5	Yes	(excluded strain17 QC)
129	SeqPhere	*OAB*	Velvet	§Ridom SeqSphere+	1503	Yes	0-3	Yes	(Yes/3)
135	SeqPhere	*OAB*	SPAdes	Ruppitsch (cgMLST)	1701	No	0-3	Yes	(No/13)
141	SeqPhere	*OAB*	Spades 3.15.2	Ruppitsch (cgMLST)	1701	Yes	0-6	Yes	(Yes/6)
142	BIGSdb-*Lm*	*OAB*	SPAdes	Pasteur (cgMLST)	1748	No	0-5	Yes	(No/14)
144	SeqPhere	*OAB*	Velvet	Ruppitsch (cgMLST)	1701	No	0-4	Yes	(excluded strain17 QC)
149	SeqPhere	*OAB*	Velvet	Ruppitsch (cgMLST)	1701	Yes	0-4	Yes	(Yes/4)

§: modified from submitted information
A&M: Assembly- and mapping-based
For more details see Annex 8.
OAB: Only assembly based
OMB: Only mapping based

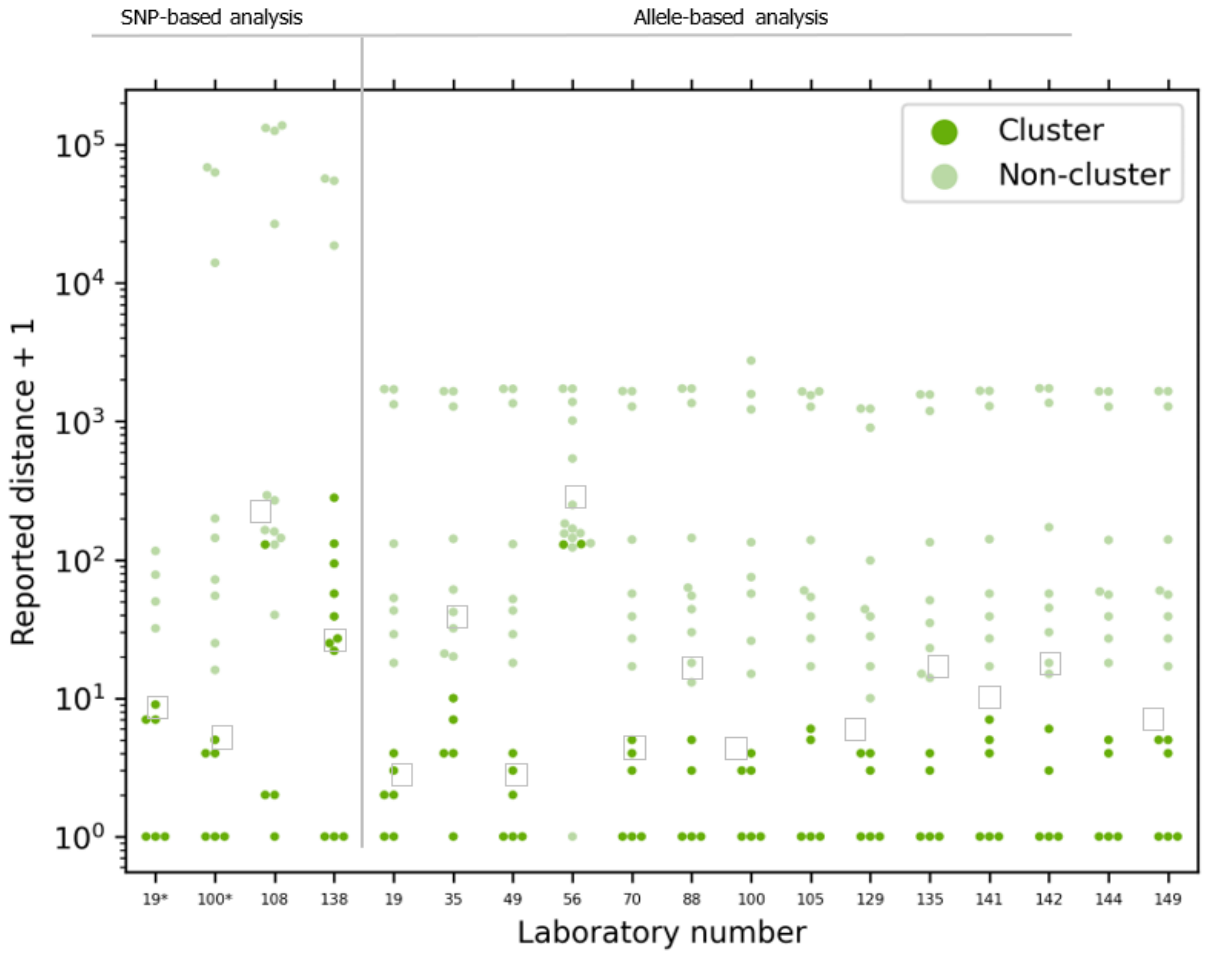
Laboratories 108 and 138, which were the only two participants that performed SNP analysis as the main analysis, did not identify the correct cluster of closely-related strains with or without strain17 (sequence with slightly reduced coverage). One of them was a new laboratory (138) participating for the first time in the cluster analysis using WGS-based data. This laboratory included all ST7 in the reported cluster. The other laboratory (108) only reported strain2, strain6, strain9 and strain14 as the cluster and excluded strain16 and strain17 (143−163 SNPs). Laboratories 19 and 100 performed SNP analysis as an additional analysis and identified the correct cluster of closely-related strains by cgMLST (main analysis). The reported SNP distances ranged from 0−8 within the cluster including strain17 for the additional analysis and a clear separation was obtained. However, the smallest distance was reported by laboratory 100, using a single-nucleotide variant based on a cgMLST (SNV) approach.

The laboratories using allele-based method can be divided into four groups (Table 7/Figure 8). Group 1 (7/14) which correctly identified the pre-defined cluster of six strains, with a max 0-6 AD within the cluster and included strain17. Group 2 (2/14) excluded strain17 based on their QC analysis of the sequences provided, and therefore did not report strain17 as part of the cluster. Group 3 (4/14) which did include strain17 in their cluster analysis, but obtained high AD (12-31 AD) between the reference strain (strain9) and strain17, meaning that they excluded strain17 from the reported cluster. Group 4 (only Laboratory 56) which did not identify the correct cluster as they only reported strain2 and strain6 as a part of the cluster.

All other laboratories obtained similar results (0-1 SNP/ 0-3 AD) for the three technical triplets (strain2 strain6 and strain9 (sequence)), whereas laboratory 56 reported 128/129 ADs. Ion Torrent data was used by laboratory 56 and the results were not comparable with the results for other platforms.

Four of the test strains/provided sequences (strain4, strain5, strain10 and strain13) were also ST7, but not predefined by the EQA-provider as part of the cluster cgMLST (AD 17-52) / SNP (30-103). Based on cgMLST, the thirteen laboratories (Group1, Group2 and Group3) reported allele differences to the selected cluster strain at 9-74 for this group of strains. One sequence provided (strain11) was ST1504 and allele differences were reported to the selected cluster strain at 98−171. (Annex 8).

Figure 6. Reported SNP distances or allelic differences for each test strain to selected cluster representative strain9



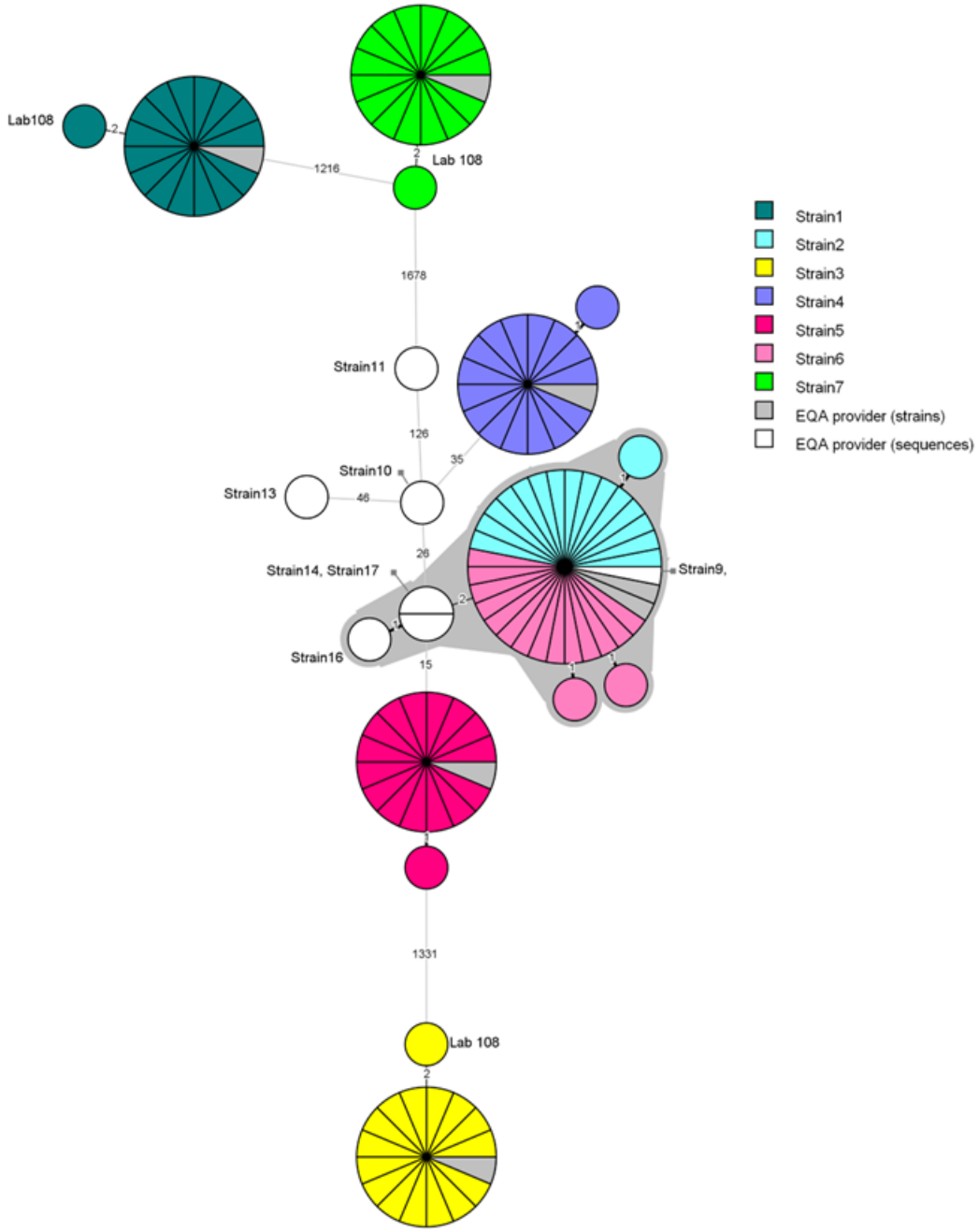
**: additional analysis.*
SNP: single nucleotide polymorphism.
Grey box around the modified strain17.
Dark green: reported cluster of closely related strains.
Light green: not reported as part of cluster.

3.3.2.4. 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 pipeline [10].

The overall cgMLST analysis, shown in the minimum spanning tree (MST), based on submitted raw reads (FASTQ files) from 16 laboratories, and the sequences provided by the EQA provider, excluding the three sequences with poor quality (strain8, strain12, strain15). Figure 7 reveals clear clustering of the results for each test strain, only two of the data notes from Laboratory 108 are separated with two AD from the other results.

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



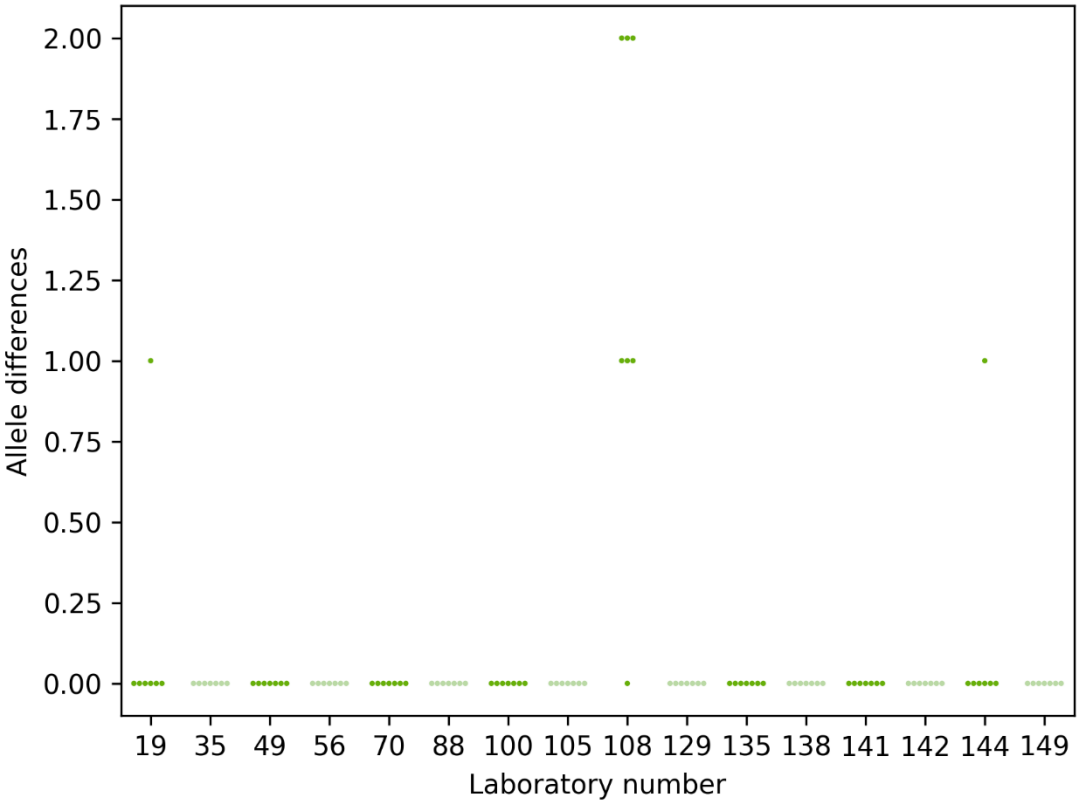
Minimum spanning tree (MST) in log scale of core genome multi-locus sequence typing (cgMLST) [6] based on submitted raw reads (FASTQ files). Each of the strain1–7 test strains have a different colour. EQA sequences provided (strain1-strain7) by the EQA provider are in grey, the sequences provided (strain9-17) are in white. The modified sequences of poor quality (strain8, strain12 and strain15) were not included in the analysis.

The allele differences in Figure 7 do not exactly match those illustrated in the individual reports and consequently those in Figure 8, as all are based on the same data. This discrepancy is caused by loci being dropped if they do not pass quality control for all strains 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 using 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 strains. As seen in Figures 7 and 8, all laboratories have small differences to the reference strains. Laboratory 108 had the 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 strain



Allele difference of participant strains from corresponding strain1-7 (EQA provider), based on submitted raw reads (FASTQ files).

For 104 of 112 results (93%), no difference was identified. As seen in Figure 8, for five results (4%), a difference of one allele from the corresponding EQA-provided strain was calculated and a difference of 2 alleles was seen for three results (3%). Results from Laboratory 108 showed allele difference for six of seven strains. The difference is for the results using Ion Torrent data analysed in BioNumerics, however Laboratory 56 also used Ion Torrent data and it had no differences.

The laboratories responded to QC parameters used to evaluate their data separately. Both confirmation of the genus and coverage were the most widely used QC parameters, with 100% and 94% of the laboratories using them (Table 10). Participants used different thresholds of coverage ranging from 29–50 x coverage. Only one laboratory did not use the coverage as a QC parameter. The laboratories reported the different programmes used for the contamination check of the genus, however Kraken was reported by five participants. The number of good cgMLST loci was used as a QC parameter by 81% of the participants, with thresholds ranging from 89–98%. Genomic size was used by 81% ranging from 2.5–3.5 Mb. Q score was only reported by 50%. Additional QC parameters were provided by some of the participants (listed in Annex 9). In addition, several participants listed N50 and GC% content as a parameter used, with a threshold of 30.000–300.000 and 37.6–38.2%.

Table 10. 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 analysis and <5% contamination with other genus	Minimum x 50	No	2.8- 3.1 Mbp	Minimum 95% core percent and maximum 30 loci with multiple consensus
35	SeqSphere in built feature	30-fold	No	Approx. 3.0 Mbp	95%
49	rMLST if problem with quality	x30	>Q30	2.6-3.2MB	97%
56	No	30	No	No	No
70	Contamination check with Mash Screen tool in SeqSphere or Species ID tool from PubMLST website	> = 30	No	Length of contigs assembled < ref genome + 10%	cgMLST alleles found and called > 95%
88	Kraken (as implemented in INNUca_v4.2.2)	INNUca_v4.2.2 employs several coverage thresholds throughout the analysis (e.g. 15x for the first estimated coverage 30x for the assembly coverage).	INNUca_v4.2.2, which performs reads quality analysis/improvement using FastQC and Trimmomatic (INNUca default settings were applied for these steps)	INNUca_v4.2.2 uses the genome size as a quality criteria during the analysis (we set 3.0 Mbp as the expected genome size and used INNUca default criteria for handling this QC parameter).	>=95% loci called using ReporTree (https://github.com/insapathogenomics/ReporTree) (exceptionally, a lower threshold is applied for outbreak investigation)
100	KmerFinder 3.1, Center for Genomic Epidemiology	40x	FastQC, threshold set to 30	SeqSphere assembler, genome size app. 2.9 Mb	SeqSphere cgMLST scheme, 95 % good targets threshold.
105	Assembled genomes were aligned against a *Listeria monocytogenes* genome (threshold:>90% nucleotide identity).	depth of coverage >45X	Trimming was performed with Trimmomatic, removing 3Â´ nucleotides with Phred <10 or an average Phred <15 in a sliding window of 4 nucleotides. Sequences with a length <70 bases were removed too.	<=3.3Mb	>=95%
108	BLASTing and mapping against local reference genome.	20x	No	2,8 - 3,3 Mbp	No
129	Presence of prfA gene (LIP).	>29	No	No	>89
135	There is a species identification tool built into our in-house assembly pipeline ("Juno")	>30	>30	2.7-3.23 Mb	>90% of alleles
138	Kraken2 read and contig based mapping	>25x	>Q30	2918000 - 3156000 bp	No
141	SeqSphere - Mash	30x	No	2.9-3.1 MB	Minimum 98 % good targets
142	Kraken2 (2.0.7)	> 30x (250bp reads)	> 30	2.5- 3.5 MB	> 95%
144	KmerFinder 3.2, SpeciesFinder 2.0	>50x	>30	2.7-3.3 Mb	>95%
149	KRAKEN	No	No	No	>90% good targets
Percentage of laboratories using QC parameter	**100%**	**94%**	**50%**	**81%**	**81%**

For each laboratory, the raw reads submitted (FASTQ files) were evaluated by the EQA provider's in-house quality control pipeline [10]. Table 11 shows the QC parameters and range of QC values per laboratory. For the full QC evaluation of all strains, see Annex 10.

According to the QC parameters, sequencing quality was uniformly good. Overall, coverage was high. Only three laboratories had issues. One laboratory (149) had *Pseudomonas tolaasii* contamination in four of the seven FASTQ files, one laboratory (105) had 'Average coverage' below 50 for some strains an additional one laboratory (108) had 'Length at >25X coverage' below 2.8, however Bifrost was developed for Illumina data and not Ion Torrent (Annex 10).

Table 11. Results of raw reads submitted by participants and evaluated by EQA provider QC pipeline, summarised by laboratory

Ranges*	{Lm}			{5%}	{2.8-3.1}	{<250}	{>0}	{<1 000}	{>50}					
Lab ID	Detected species	% Species 1	% Species 2	Unclassified reads (%)	Length at >25 x min. coverage (Mbp)	Length [1-25] x min. coverage (kbp)	No. of contigs at 25 x min. coverage	Contigs at [1,25]X coverage	Average coverage	No. of reads (x 1000	Average read length	Average insert size	N50 (kbp)	QC status (Bifrost)
19	Lm	94.3-94.8	0.0-0.1	5.0-5.6	2.9-3.0	0.6-17.8	81.0-155.0	1.0-16.0	90.0-112.0	1 836.0-2 347.0	140.0-143.0	213.0-228.0	34.0-62.0	OK
35	Lm	96.1-97.2	0.0-0.1	2.7-3.8	2.9-3.0	0.0	15.0-37.0	0.0	184.0-224.0	3 630.0-4 830.0	139.0-147.0	253.0-317.0	177.0-451.0	OK
49	Lm	96.6-98.7	0.0-0.3	1.2-2.9	2.9-3.0	0.0	13.0-22.0	0.0	75.0-183.0	830.0-2 197.0	245.0-277.0	269.0-422.0	344.0-1 500.0	OK
56	Lm	98.5-98.8	0.0-0.2	1.1-1.5	2.8-3.0	0.0	213.0-553.0	0.0	149.0-271.0	1 557.0-2 515.0	279.0-346.0	0.0-13.0	9.0-26.0	OK
70	Lm	97.8-98.0	0.0-0.1	2.0-2.2	2.9-3.0	0.0	12.0-24.0	0.0	77.0-92.0	1 518.0-1 867.0	147.0	262.0-281.0	344.0-1 500.0	OK
88	Lm	97.7-98.2	0.0-0.2	1.6-1.9	2.9-3.0	0.0-6.2	48.0-132.0	0.0-9.0	105.0-165.0	2 147.0-3 376.0	146.0-147.0	303.0-326.0	41.0-116.0	OK
100	Lm	97.9-98.3	0.0-0.1	1.7-2.0	2.9-3.0	0.0	12.0-24.0	0.0	391.0-1096.0	8 045.0-22 359.0	144.0-147.0	210.0-286.0	344.0-1 500.0	OK
105	Lm	96.5-97.3	0.0-0.1	2.7-3.5	2.9-3.0	0.0	13.0-21.0	0.0	36.0-57.0	738.0-1 164.0	143.0-146.0	283.0-369.0	344.0-1 500.0	Warning
108	Lm	97.4-98.0	0.0-0.1	1.9-2.6	2.6-3.0	0.0-3.1	612.0-2 825.0	0.0-13.0	61.0-119.0	710.0-1 321.0	263.0-281.0	0.0	2.0-8.0	Warning
129	Lm	96.4-97.9	0.0-0.1	2.0-3.5	2.9-3.0	0.0-29.6	35.0-282.0	0.0-30.0	69.0-249.0	1 386.0-5 348.0	137.0-149.0	195.0-382.0	20.0-183.0	OK
135	Lm	95.9-97.6	0.1-1.0	2.2-3.3	2.9-3.0	0.0	14.0-23.0	0.0	204.0-579.0	4 427.0-11 929.0	149.0	272.0-311.0	344.0-1 500.0	OK
138	Lm	98.5-98.9	0.0-0.2	1.1-1.3	2.9-3.0	0.0	12.0-24.0	0.0	261.0-331.0	5 364.0-6 510.0	148.0	362.0-401.0	225.0-1 500.0	OK
141	Lm	96.7-97.7	0.0-0.1	2.1-3.2	2.9-3.0	0.0-10.6	32.0-125.0	0.0-10.0	56.0-170.0	657.0-2 069.0	246.0-255.0	291.0-341.0	38.0-254.0	OK
142	Lm	97.0-98.0	0.0-0.2	1.8-2.7	2.9-3.0	0.0-3.6	15.0-23.0	0.0-5.0	56.0-66.0	693.0-788.0	242.0-245.0	428.0-452.0	302.0-549.0	OK
144	Lm	97.1-98.5	0.0-0.1	1.4-2.9	2.9-3.0	0.0	14.0-20.0	0.0	83.0-123.0	1 684.0-2 433.0	145.0-148.0	219.0-321.0	344.0-1 500.0	OK
149	Lm, Pt	82.0-91.3	1.9-6.9	6.2-9.4	2.9-3.0	0.0-0.7	17.0-38.0	0.0-1.0	68.0-104.0	1 552.0-2 283.0	151.0	289.0-320.0	237.0-1 500.0	Warning

**: indicative QC range.*
Lm: L. monocytogenes, Pt: Pseudomonas tolaasii
Warning: Some issues were noted in the submitted sequences (see Annex 10).

3.4 Feedback survey – evaluation of the EQA scheme

After the individual reports were sent to the participants, the EQA provider circulated a feedback survey to assess the EQA scheme of *L.monocytogenes*. The questionnaire contained questions on aspects of accreditation, information on the individual report, the actions taken if errors were detected, the usefulness of the QC evaluation of the participants' sequenced data, the usefulness of including low quality data and any suggestions for improvements. The survey response rate was 65% and the results are summarised in Table 12.

Table 12. Results of evaluation of the EQA scheme

Questions	Response (Yes)	Comments /actions
1) Used for accreditation/licensing purposes?	10/13 (77%)	One reported applying for accreditation
2) Satisfied with the format/comments?	13/13 (100%)	One reported the individual report was too comprehensive.
3) Did any of your analytical test results differ?	8/13 (62%)	One reported they re-cultured and sequenced strain3 and solved the serotype error.
		One reported implementing pipeline for the detection of mixed cultures.
		One reported that investigation of the 'high allele issue' in strain17 was related to the calling of partial alleles.
4) Usefulness of the manipulated sequences?	12/12 (100%)	One reported that the manipulated sequences resulted in awareness of certain pitfalls when analysing low quality data.
5) Usefulness of the QC status of your submitted sequences?	12/12 (100%)	Yes, but guidelines for interpreting data and QC should be published, after the EQA, in the final report.
6) Improvements/remarks		Deposit the EQA provider sequences (strain1-17) in public databases (BIGSdb) to serve as reference for validation.
		The individual report could be modified to improve readability.
		For small labs the many EQAs (both typing and AMR) within a short timeframe is very hard.
		Include identification genus/species level as it is requested by customers and third parties.
		We need to have more practice analysing our NGS sequences and performing cluster analysis.
		One reported they wished the individual report were easier to read.
		One would have liked the EQAs to be spread out over the year, as it was only a small lab with few staff.

N=13 for main questions (1-3+6), N=12 for WGS related questions (4-5).

4. Discussion

Based on the evaluations completed, we believe that most of the participants were satisfied with the format of the individual report and additional feedback from the EQA provider, as only one participant found the individual report too comprehensive and suggested unspecified modifications. As the evaluation is based on anonymised responses it is not possible to follow up, however the suggestion will be discussed during the planning of the next round. The inclusion of the modified sequences in the cluster analysis and the QC feedback of the uploaded sequences was well received by the participants. The suggestions are listed in Section 6.

4.1 Serotyping/serogrouping

Eighteen (90%) laboratories participated in the serotyping part of EQA-9 and of these 17 (94%) provided molecular serogrouping results.

4.1.1 Conventional serotyping

The number of participants in the conventional serotyping decreased from 10 laboratories in EQA-1 to five in EQA-9, highlighting the transition towards the use of molecular serogrouping. One laboratory participated in the conventional serotyping only. The decreasing number of laboratories have also affected the performance in general. From EQA-1 to 7 the performance was above 85% (87–100%), however in the last two EQAs the performance has been below 85% (82% and 80%). Three of the participants (Lab 100, 138 and 142) were the same as in EQA-8, however two additional laboratories again submitted conventional serotyping results, one performed 100% correctly and the other had 86% correct.

One laboratory (138) has had multiple errors since they started conventional serotyping in EQA-7 to EQA-9. In EQA-7 and -8 laboratory 138 correctly identified a single 1/2a strain, however in EQA-9 with four 1/2a strains only one was correctly identified and the others were reported as 1/2b, 3a or 3b. The same was observed for strains with 4b, the laboratory only once identified the serotype correctly. However, this year the laboratory has started to use WGS, and the reported molecular serogroups were correctly identified.

4.1.2 Molecular serogrouping

Since EQA-2, the number of participants in the molecular serogrouping has ranged from 13 to 17 participants. From EQA-6 to EQA-8, three laboratories reported the use of in silico PCR (WGS) serogrouping, and this increased to ten in EQA-9. Reducing the number of strains to sequence has probably had a positive effect on the number of participants. With regard to molecular serogrouping, the performance was also very good in EQA-9, with a score of 96% correct. Over the years from EQA-1 to EQA-9, the general performance among the participating laboratories has been strong: 98%; 94%; 94%; 94%; 99%; 97%; 100%; 99% and 96%. The serogroup of strain3 was incorrectly reported as IIa by four (24%) participants, which is more than in previous EQAs. However one laboratory has already re-cultured and sequenced the strain and found the correct serogroup. The EQA provider has also analysed the submitted sequences of strain3 from the participants and found the serogroup to be IIc for all. The analysis was conducted using BioNumerics plugin version 8.1 and an in-house (EQA provider) - in silico PCR.

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 has evolved alongside the development of surveillance methods used by NPHRLs in the EU/EEA. The adjustment of the EQA appears to have been well accepted by the countries, as 17 of the 20 laboratories (85%) participated in the cluster analysis, which is two more than in EQA-8. Only one laboratory participated in cluster identification using PFGE as a sole method, while another laboratory participated in the cluster identification using both PFGE- and WGS-derived data.

4.2.1 PFGE-derived data

Of the 17 laboratories participating in the cluster analysis, two (12%) 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 two cluster strains (defined by WGS) among the strains in the package, both included one or two other strains. Laboratory 114 only performed PFGE and Laboratory 142 also performed cluster analysis on WGS-derived data.

The number of participants only submitting cluster analysis based on PFGE-derived data has decreased with each of the EQAs and this time 94% (16/17) submitted analysis based on WGS-derived data.

4.2.3 WGS-derived data

Sixteen of the 17 laboratories (94%) 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 and since then the number of participants has varied, but increased overall. In addition, in EQA-9 one laboratory participated for the first time using WGS-derived data. Almost all (94%) laboratories reported that the sequencing was done at their own premises. The majority (14/16) also reported using an Illumina platform. All reported using commercial kits for library preparation.

The EQA provider QC evaluation of the raw reads submitted by the participants showed good-quality data. Only three participants received warnings from the Bifrost QC pipeline.

One laboratory (149) received warnings since *Pseudomonas tolaasii* was identified as an additional species when checking for contamination. The contamination assessment in Bifrost is based on Kraken [11]. The second laboratory received warnings as the 'average coverage' was below the threshold of 50 but within the range of 36−48, the cut-off of 50 in the Bifrost pipeline is very strict. The participant reported they used a 'depth of coverage of >45x', which would then only have resulted in a warning of one sequence using their threshold. The third laboratory received a warning due to the 'Length at >25X' coverage being below 2.8, however as the participant submitted IonTorrent data, the EQA provider knows that some of the QC values provided by Bifrost are unreliable due to assembly issues for Ion Torrent data.

As in previous years, the main QC parameters reported in EQA-9 were a threshold of coverage and the control of genus/species confirmation. The percentage of participants using assessment of the genome size has increased from 71% to 81% and only one of the participants using allele-based analysis was not using the number of cgMLST allele cited as a QC parameter.

Fourteen laboratories (88%) reported using an allele-based method as the main analysis and two laboratories (12%) reported using SNP analysis. Compared to EQA-6 to EQA-8 (75%, 85%, and 86%), this is a 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 switched between allele-based and SNV approaches (SNP-based).

As in previous EQAs, many participants used the Ruppitsch cgMLST scheme for the main analysis (57%).

For the first time in the EQA, the EQA-provided sequences had to be in included in the main cluster analysis by the participant. All the laboratories disregarded the strain with reduced coverage (strain12). Almost all (75%) laboratories identified the enlarged genome size of the sequences for strain15 and therefore disregarded the sequences in the analysis. Four laboratories accepted the sequence either as QC status A or B, although no laboratories (incorrectly) included the strain in the cluster of closely-related strains. The contamination with *L. innocua* in sequence strain8 was identified by 75% of the participants, however, one used the sequence with caution (for outbreak purposes) and one used the sequences despite the contamination. None of the participants included the strain in the cluster of closely-related strains.

Sequence strain17 was a cluster strain, made from strain14 with a slightly reduced coverage, and the participants found this challenging, assigning different QC statuses. Fourteen (88%) laboratories accepted the sequence either as QC status A or B, but two laboratories discarded the sequence as the cgMLST core percent was below the participant's threshold of 95%. The cut-off was reported in the QC section (Table 10).

The modification of strain17 increased the diversity of the results. This was partly because the QC status was determined differently and also because some laboratories achieved a higher number of AD/SNP than expected by the EQA provider (and other participants). However, when assessing the cluster of closely related strains with QC status A, (strain17 not included) 13/16 laboratories identified the cluster. If strain17 is included in the cluster assessment, only seven laboratories identified the cluster of closely-related strains. The ADs reported for the strains with QC status A were very comparable, both inside and outside the cluster, despite using different schemes with different numbers of loci.

Only one laboratory (56) submitted very different results for many of the test strains and sequences (mix of Ion Torrent (own sequencing) and Illumina data (EQA provided sequences)). However, when using BioNumerics on the participant's submitted data, the analysis showed the expected results, and the cluster was easily identified. The laboratories using the Pasteur scheme with 1 748 loci reported between 0−2 (inside the cluster, without strain17) and up to 0−5 (inside the cluster, with strain17) AD and the laboratories using the Ruppitsch scheme reported 0−3 and up to 0−9 AD. Laboratory 129, which has a scheme with a different number of loci (1503), reported 0−3 AD.

Neither of the two laboratories using SNP as the main analysis identified the cluster, either with or without strain17. One of the laboratories (138) was a new laboratory conducting cluster analysis using WGS for the first time and

therefore listed all ST7 as closely-related strains (0-130 SNPs), however, there is a small separation between the correct cluster (0−26 SNPs) and the closets ST7 strain5 (38 SNPs). The other laboratory (108) reported a cluster of four correct strains within 0−128 SNPs, however, they also reported 128 SNPs for a non-cluster strain (strain10). In addition, laboratory 108 reported strain16 and strain17 as being outside the cluster, with 143 and 163 SNPs. It was also one of the laboratories that assigned strain15 with OC status A and added the strain in the SNP analysis. The laboratory must have used a combination of Ion Torrent data (own sequencing of strain1-7) and the EQA provided Illumina data in their analysis, which might also have influenced the outcome of the analysis.

It is challenging to evaluate the results from the four laboratories achieving a high number of AD for strain17. From the reported details of the cluster analysis, laboratories 35 and 149 listed the same approach (SeqPhere, with only assembly-based using Velvet as mapping with the Ruppitsch (cgMLST) with 1701 loci). However, the number of AD reported is different, especially for strain16 and strain17 (nine and 31 ADs for laboratory 35, compared with four and four ADs for laboratory 149). See Annex 8 and Figure 6. Laboratory 35 did, however, include strain16 (9ADs) in the cluster despite the normal threshold of seven AD.

Laboratory 88 and 142 both reported using an assembly-based approach only, using SPAdes and the same Pasteur scheme of 1748 loci in INNUca and BIGSdb-Lm. The two laboratories obtained similar AD differences, although different from other laboratories using the same scheme, but used a different assembler (Skesa in BioNumerics). When trying to use an assembler-based approach only (Spades 3.15.3 or Velvet at different settings) in BioNumerics, the EQA provider was unable to achieve similarly high ADs to those obtained by the participants.

Laboratory 135 used a similar approach to Laboratory 141, an assembler-based approach only, using SPAdes and the same Ruppitsch scheme of 1701 loci in Seqphere. The lab obtained very similar results, except for strain17 (13 AD versus six AD). Laboratory 141 described the version of SPAdes used in detail (v3.15.2), however, the EQA provider did not ask the participants for version details or the different settings used for the assembly.

Palma et al., 2022 concluded that all the examined workflows (BIGSdb, INNUENDO, GENPAT, SeqSphere, MentaLIST) need a depth of coverage >40 and high loci detection >99.54% (BioNumerics only 97.78%) to have consistent cluster definitions when using the reference cut-off of seven AD. [12]. Similarly, the current EQA indicates that the modification created (~lower quality data for strain17) is more sensitive to possible small discrepancies in the settings and analysis tools used by the different participants. A proposed explanation from one of the participants (during the evaluation) is the calling of new alleles (probably partial alleles) and missing loci being counted as ADs in some of the approaches. A comparison of strain14 (original sequence before the modification) and strain17 in BioNumerics shows zero ADs, but 112 discrepancies (108 missing data in strain17, and four new alleles which were not present in strain14). Each approach probably has a QC threshold that affects the analysis and conclusions.

Once again this year, as seen in previous reports, when the data from Laboratory 108 was analysed by the EQA provider (using the standardised cgMLST/Pasteur analysis) it showed allele differences for most of the test strains (Figure 8). 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, but not for Laboratory 56 which also provided Ion Torrent data). Therefore, the AD observed may be method artifacts. However, the use of Ion Torrent data can complicate the communication and investigation of multi-country outbreaks when only using the allelic method.

Only two of the four laboratories performing SNP analysis obtained a clear separation of the cluster and the non-cluster strains using their approach/analysis. (Figure 6).

As seen in previous EQAs and in the publication by Henri et al. 2017 [13], the two approaches for analysing WGS-derived data (allele- and SNP-based analysis) show comparable results. This year, however, there was a great deal more variation in the SNP analysis.

5. Conclusions

Twenty laboratories participated in the EQA-9 scheme, with 18 (90%) performing serotyping and 17 (85%) performing cluster identification. It was very encouraging to see an increase in participation for both serotyping and cluster analyses.

Most laboratories (72%, 13/18) performed only molecular serogrouping, while 22% (four) performed molecular serogrouping in combination with conventional serotyping and only 6% (1) performed conventional serotyping alone. In general, there has been an increasing trend towards replacing conventional serotyping with molecular serogrouping throughout the nine EQAs. The average quality of conventional serotyping (80%) was below the range for the previous EQAs. The performance for molecular serogrouping was good, achieving 96% in EQA-9. 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 strains than expected in the cluster for WGS.

Most of the participants were able to identify the different characteristics (and modifications) of the EQA-provided sequences. For the sequence with very low coverage, all participants identified the QC issues and did not proceed with cluster analysis. However, the participants did not agree on the QC status of the modified cluster sequence with slightly reduced coverage: all three groups of QC status (A, B and C) were selected for the sequence. In addition, the modified sequence also resulted in very different AD, depending on the Approach/Assembler/Allelic calling method used. Only two (13%) laboratories did not detect the sequence with both ST1 and ST224 combined.

Sixteen laboratories performed cluster analysis using WGS-derived data. The performance was good when analysing the good quality data (QC status A) and 13/16 (81%) of the participants correctly identified the cluster of closely-related strains. However, the modified strain17 did interfere with the analysis, causing four participants to exclude strain17 as a cluster strain since they achieved higher AD than the rest of the participants. In addition, two participants excluded strain17 because of the reduced core percent.

In the WGS, an allele-based method was preferred, as 88% (14/16) used cgMLST, compared to 13% (2/16) using SNP as the main method reported for cluster analysis. In general, the reported cgMLST results for participants own sequence data and the non-modified sequences were at a comparable level of allelic difference (0–9) within the cluster strains, despite being analysed using different schemes. However, most did report only 0–4 ADs in the cluster. The reported SNP results showed that neither of the two participants (main analysis) could separate and report the correct cluster, however one of the participants had only recently started to use WGS. In addition, both laboratories reported SNP as an additional analysis and could clearly see the separation of the cluster. Therefore, both methods seem to work for cluster detection, even though it was less obvious than in previous EQAs. Furthermore, standardised cgMLST analyses leave little room for error when only including good quality data, which results in good inter-laboratory comparability.

The current EQA scheme for *L. monocytogenes* typing is the ninth EQA organised for NPHRLs 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 five 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. In addition to the WGS data submitted by the Member States for the listeriosis cluster detection, ECDC also collects and centrally analyses sequence data during multi-country outbreak investigations initiated by Member State public health institutes in EpiPulse.

6. Recommendations

6.1 Laboratories

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

When laboratories re-name/change the strains from the EQA-provided ID to an ID that fits into their pipelines, 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.

6.3 EQA provider

The EQA have evolved over the years as the EQA provider has included additional sequences, some modified to have QC issues. The positive feedback from evaluation suggests that this should be continued in future EQAs. Moreover, the QC assessment of the sequences submitted by the participants was appreciated, although one laboratory did suggest guidelines for interpreting data and sharing QC with the participants. In the next round, the EQA provider will share the available QC assessment documents with the participants. Another suggestion was that the EQA provider sequences (strain1-17) should be placed in public databases (BIGSdb) to serve as a reference for validation. This can be done immediately after the EQA has concluded. Two suggestions – the possibility of including identification genus/species level as an addition in future EQAs and a more even distribution of the EQA schemes for different pathogens across the year – will be discussed with ECDC.

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:https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex:32004R0851
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: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex%3A32013D1082
3. European Food Safety Authority (EFSA)/European Centre for Disease Prevention and Control (ECDC). The European Union One Health 2021 Zoonoses Report. EFSA J. 2022;20(12):7666 Available from: https://doi.org/10.2903/j.efsa.2022.7666
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. Wood, DE, Salzberg SL. Kraken: ultrafast metagenomic sequence classification using exact alignments. Genome Biology 2014, 15:R46.
12. Palma F, Mangone I, Janowicz A, et al. In vitro and in silico parameters for precise cgMLST typing of *Listeria monocytogenes*. BMC Genomics 23, 235 (2022). Available from: https://doi.org/10.1186/s12864-022-08437-4
13. 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	NRL for Listeria	Austrian Agency for Health and Food Safety
Belgium	NRC Listeria	Sciensano
Denmark	Laboratory of Gastrointestinal Bacteria	Statens Serum Institut
Germany	FG11/Consultant Laboratory for Listeria	Robert Koch Institute
Denmark	Laboratory of Gastrointestinal Bacteria	Statens Serum Institut
Spain	Neisseria, Listeria and Bordetella Unit	National Centre for Microbiology Instituto de Salud Carlos III
Finland	Expert Microbiology	Finnish Institute for Health and Welfare
France	National Reference Centre and WHO Collaborating Centre Listeria	Institut Pasteur
Hungary	FWD Laboratory	National Public Health Center
Ireland	Galway Microbiology Reference Laboratory	University Hospital Galway
Italy	Department of Food Safety, Nutrition and Veterinary Public Health	Istituto Superiore di Sanita
Lithuania	Clinical testing Department	National Public Health Surveillance Laboratory
Luxembourg	Epidemiology and Microbial Genomics	Laboratoire National de Santé
Latvia	Latvian Centre of Infectious Diseases National Microbiology Reference Laboratory	Riga East University Hospital
The Netherlands	IDS	RIVM
Norway	National Reference Laboratory for Enteropathogenic Bacteria	Norwegian Institute of Public Health
Portugal	URGI	National Institute of Health Dr. Ricardo Jorge (INSA)
Sweden	Department of Microbiology	Public Health Agency of Sweden (Folkhälsomyndigheten)
Slovenia	Department for Public Health Microbiology	National Laboratory of Health, Environment and Food
Slovakia	Department of Environmental Microbiology	Regional Authority of Public Health Košice
Türkiye	National Reference Laboratory for Enteric Pathogens	General Directorate of Public Health

Annex 2. Participation overview EQA-8 and 9

	2020 to 2021 (EQA-8)					2022 (EQA-9)				
	All[#]	Serotyping		Cluster		All[#]	Serotyping		Cluster	
Laboratory		Conventional	Molecular	PFGE	WGS		Conventional	Molecular	PFGE	WGS
19	X				X	X		X		X
35	X		X		X	X		X		X
49	X				X	X		X		X
56	X	X			X	X				X
70	X		X		X	X		X		X
88						X		X		X
96						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
114						X	X		X	
129	X		X		X	X		X		X
130	X		X							
135*	X				X	X				X
138	X	X		X		X	X	X		X
141	X				X	X		X		X
142	X	X	X	X	X	X	X	X	X	X
143	X		X			X		X		
144	X		X		X	X		X		X
145						X	X	X		
149	X		X		X	X		X		X
Number of participants	**17**	4	11	2	14	**20**	5	17	2	16

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

Annex 3. Serotyping result scores

Conventional serotyping

	Laboratory ID					
Strain number	**Provider**	**100**	**114**	**138**	**142**	**145**
Strain1	4b	4b	4b	4d	4b	4b
Strain2#	1/2a	1/2a	1/2a	3b	1/2a	1/2a
Strain3	1/2c	1/2c	1/2c	3c	1/2c	1/2a
Strain4	1/2a	1/2a	1/2a	1/2b	1/2a	1/2a
Strain5	1/2a	1/2a	1/2a	3a	1/2a	1/2a
Strain6#	1/2a	1/2a	1/2a	1/2a	1/2a	1/2a
Strain7	1/2b	1/2b	1/2b	1/2b	1/2b	4b

Molecular serogrouping

		Laboratory ID																
Strain number	Provider	19	35	49	70	88	96	100	105	108	129	138	141	142	143	144	145	149
Strain1	IVb	IVb	IVb	IVb	IVb	IVb	IVb	IVb	IVb	IVb	IVb	IVb	IVb	IVb	IVb	IVb	IVb	IVb
Strain2#	IIa	IIa	IIa	IIa	IIa	IIa	IIa	IIa	IIa	IIa	IIa	IIa	IIa	IIa	IIa	IIa	IIa	IIa
Strain3	IIc	IIc	IIc	IIc	IIc	IIa	IIc	IIc	IIc	IIc	IIa	IIc	IIa	IIc	IIc	IIc	IIa	IIc
Strain4	IIa	IIa	IIa	IIa	IIa	IIa	IIa	IIa	IIa	IIa	IIa	IIa	IIa	IIa	IIa	IIa	IIa	IIa
Strain5	IIa	IIa	IIa	IIa	IIa	IIa	IIa	IIa	IIa	IIa	IIa	IIa	IIa	IIa	IIa	IIa	IIa	IIa
Strain6#	IIa	IIa	IIa	IIa	IIa	IIa	IIa	IIa	IIa	IIa	IIa	IIa	IIa	IIa	IIa	IIa	IIa	IIa
Strain7	IIb	IIb	IIb	IIb	IIb	IIb	IIb	IIb	IIb	IIb	IIb	IIb	IIb	IIb	IIb	IIb	IVb	IIb
Method	WGS	WGS	PCR	WGS	WGS	WGS	PCR	PCR	WGS	WGS	WGS	WGS	WGS	PCR	PCR	PCR	PCR	WGS

#: technical duplicates
Purple shading: repeat strains in EQA-1 to 9 (strain1, strain2 and strain3). Strain2 was a different strain to previous years, but the same serotype/group.
Pink shading: incorrect results.

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

Lab ID	Reported cluster	Corresponding strains	Included the two strains from the WGS cluster*	Included additional strains in the cluster
Provider		Strain2 and Strain6		
114	1798 1874 1838 1219	Strain2, Strain4 Strain5, Strain6	Yes	Yes
142	1436 1596 1131	Strain2, Strain5, Strain6	Yes	Yes

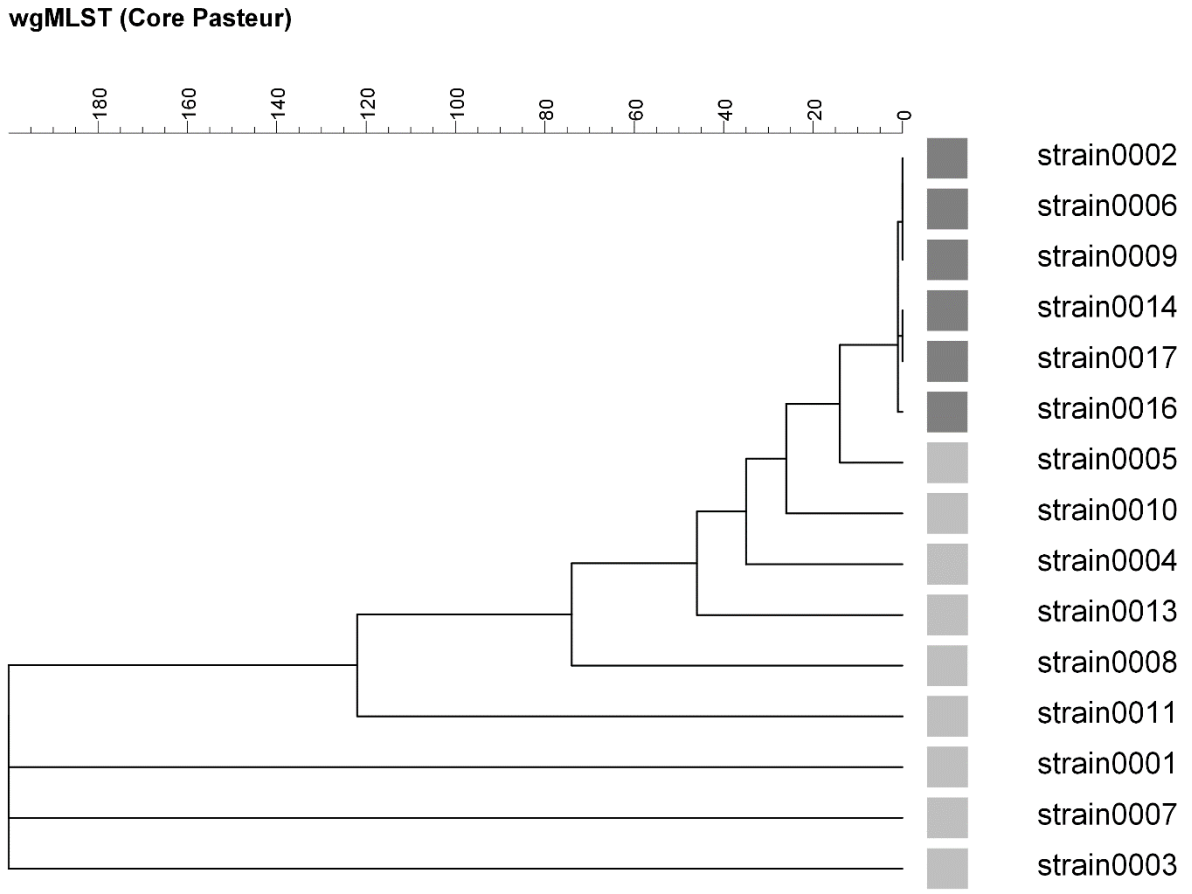
**pre-defined categorisation using WGS derived data.*

Annex 5. Reported sequencing details

Sequencing performed	Protocol (library prep)	Commercial kit	Sequencing platform
In own laboratory	Commercial kits	Nextera XT Kit (Illumina)	NextSeq
In own laboratory	Commercial kits	Nextera XT DNA Library Prep Kit	NextSeq
In own laboratory	Commercial kits	Illumina DNA Prep	MiSeq
Externally	Commercial kits	Biosearch genomics Masterpure Gram positive DNA purification kit	Ion Torrent PGM
In own laboratory	Commercial kits	DNA prep Illumina (Nextera Flex)*	MiniSeq
In own laboratory	Commercial kits	Nextera XT DNA Library*	NextSeq
In own laboratory	Commercial kits	Illumina DNA Prep	NextSeq
In own laboratory	Commercial kits	Nextera DNA flex library prep (Illumina)	MiSeq
In own laboratory	Commercial kits	Ion Xpress Plus Fragment Library Kit for AB Library Builder System (Thermo Fisher S.)	Ion GeneStudio S5 Prime system
In own laboratory	Commercial kits	Illumina Nextera XT	MiSeq
In own laboratory	Commercial kits	Illumina DNA prep	NextSeq
In own laboratory	Commercial kits	Illumina DNA prep	NextSeq
In own laboratory	Commercial kits	Nextera XT DNA Library Kit, Illumina	MiSeq
In own laboratory	Commercial kits	MiSeq Reagent Kit v3 - 500 cycles	MiSeq
In own laboratory	Commercial kits	DNA Prep	MiSeq
In own laboratory	Commercial kits	Kapa HyperPlus (Kapa Biosystems)	NextSeq

** The reaction was performed at 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-9 strains (cgMLST, Pasteur, Moura et al., 2016).*
Analysed in BioNumerics: maximum distance of 200 exceeded; results clipped.
Dark grey: cluster strains.
Light grey: outside cluster strains.

Annex 7. Reported cluster of closely-related strains, based on WGS-derived data

Lab ID	Reported cluster	Corresponding to EQA provider strains	Correct cluster without modified sequence (strain17)	Correct cluster
Provider		Strain2, Strain6, Strain9, Strain14, Strain16, Strain17 (2/6 and 9 technical triplicates)		
19	1422, 1111, 0009, 0014, 0016, 0017	Strain2, Strain6, Strain9, Strain14, Strain16, Strain17	Yes	Yes
35	1391, 1243, 0009, 0014, 0016	Strain2, Strain6, Strain9, Strain14, Strain16	Yes	No
49	1633, 1971, 0009, 0014, 0016, 0017	Strain2, Strain6, Strain9, Strain14, Strain16, Strain17	Yes	Yes
56	1582, 1999	Strain2, Strain6	No	No
70	1309, 1606, 0009, 0014, 0016, 0017	Strain2, Strain6, Strain9, Strain14, Strain16, Strain17	Yes	Yes
88	1417, 1478, 0009, 0014, 0016	Strain2, Strain6, Strain9, Strain14, Strain16	Yes	No
100	1680, 1645, 0009, 0014, 0016, 0017	Strain2, Strain6, Strain9, Strain14, Strain16, Strain17	Yes	Yes
105	1848, 1295, 0009, 0014, 0016	Strain2, Strain6, Strain9, Strain14, Strain16	Yes	No
108	1033, 1503, 0009, 0014	Strain2, Strain6, Strain9, Strain14	No	No
129	1203, 1564, 0009, 0014, 0016, 0017	Strain2, Strain6, Strain9, Strain14, Strain16, Strain17	Yes	Yes
135	1673, 1725, 0009, 0014, 0016	Strain2, Strain6, Strain9, Strain14, Strain16	Yes	No
138	1956, 1193, 1674, 1041, 0009, 0010, 0011, 0013, 0014, 0016, 0017	Strain2, Strain4, Strain5, Strain6, Strain9, Strain10, Strain11, Strain13, Strain14, Strain16, Strain17	No	No
141	1978, 1292, 0009, 0014, 0016, 0017	Strain2, Strain6, Strain9, Strain14, Strain16, Strain17	Yes	Yes
142	1436, 1131, 0009, 0014, 0016	Strain2, Strain6, Strain9, Strain14, Strain16	Yes	No
144	1139, 1954, 0009, 0014, 0016	Strain2, Strain6, Strain9, Strain14, Strain16	Yes	No
149	1036, 1284, 0009, 0014, 0016, 0017	Strain2, Strain6, Strain9, Strain14, Strain16, Strain17	Yes	Yes

Strains 8–17 were provided sequences. Strain 8, 12, 15 and 17 were modified by the EQA provider.

Annex 8. Reported SNP distance and allelic differences

SNP distances

			Laboratory ID			
Strain number	ST	Provider	19*	100*	108	138
Strain1	2	113224	9999	68771	138181	55004
Strain2‡#	7	0	0	0	1	0
Strain3	9	22869	9999	14042	26788	18715
Strain4	7	65	77	54	159	93
Strain5	7	30	31	15	39	38
Strain6‡#	7	0	0	0	1	0
Strain7	3	112427	9999	63317	132504	57031
Strain8^	-	9999	9999	198	9999	9999
Strain9‡#	7	0	0	0	0	0
Strain10	7	43	49	24	128	56
Strain11	1504	232	9999	143	292	280
Strain12^	-	9999	9999	9999	9999	9999
Strain13	7	103	115	71	268	130
Strain14‡	7	6	6	3	128	24
Strain15^	-	9999	9999	9999	126354	9999
Strain16‡	7	8	8	4	143	26
Strain17‡^	7	6	6	3	163	21

Allelic difference

			Laboratory ID													
Strain number	ST	Provider	19	35	49	56	70	88	100	105	129	135	141	142	144	149
Strain1	2	1726	1715	1655	1721	1727	1659	1731	2757	1651	1238	1568	1668	1737	1652	1659
Strain2‡#	7	0	0	3	0	128	0	0	0	0	0	0	0	0	0	0
Strain3	9	1353	1329	1282	1353	1385	1282	1358	1222	1278	901	1191	1291	1362	1278	1283
Strain4	7	42	42	41	42	167	38	43	56	38	27	34	38	44	38	38
Strain5	7	17	17	19	17	143	16	17	14	16	9	14	16	17	17	16
Strain6‡#	7	0	1	3	0	129	0	0	0	0	0	0	0	0	0	0
Strain7	3	1722	1708	1653	1722	1726	1655	1730	1581	1647	1235	1565	1665	1736	1648	1655
Strain8^	-	9999	9999	9999	9999	182	9999	62	9999	59	43	9999	9999	9999	58	59
Strain9‡#¤	7	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Strain10	7	28	28	30	28	155	26	29	25	26	16	22	26	29	26	26
Strain11	1504	130	130	141	129	154	139	143	133	138	98	133	140	171	138	139
Strain12^	-	9999	9999	9999	9999	9999	9999	9999	9999	9999	9999	9999	9999	9999	9999	9999
Strain13	7	52	52	60	51	249	56	54	74	53	38	50	56	56	55	55
Strain14‡	7	2	2	6	2	122	3	2	2	5	2	2	3	2	3	3
Strain15^	-	9999	9999	9999	9999	539	9999	9999	9999	1549	9999	9999	9999	9999	9999	9999
Strain16‡	7	3	3	9	3	131	4	4	3	4	3	3	4	5	4	4
Strain17‡^	7	1	1	31	1	1015	2	12	2	9999	3	13	6	14	9999	4

**: additional analysis*
‡: closely-related strains
^: modified sequences
#: technical triplicates
Strain9 used as the representative to report the AD/SNP distance.
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 ambiguous sites.	Available from QC analysis but no threshold.		
49	N50	Pass > 300 000						
70	Number of contigs	Threshold of 250						
88	Inter- and intra-species contamination	Kraken, ConFindr, etc.	Number of contigs	INNUca_v4.2.2 (default setting)				
100	N50	70 000	Contig count	Less than 100 contigs.	SAV-NextSeq run parameters	Clusters passing filter, no. of generated reads and Q30 score were all according to Illumina recommendations		
108	MLST	Coverage >10x and 100% ID for every detected gene.						
135*	N50	>30 000	Contig number	<=300	Contamination of assembly	CheckM, <4%	GC%	37.6–38.2%
138*	N50	>60 000 bp	Kraken2 read and contig-based mapping.	>90% mapped to single species	Number of full genes.	2800-3200	GC content	37.8–38 %
141	contig size	200 contigs shorter than 200 bases were ignored						
142	GC content of 38%:	Warning >2% deviation failure >4% deviation						
149	Contamination	KRAKEN						

135*: In addition, completeness (CheckM) threshold >96
138*: In addition, insert size distribution which adheres to Gaussian distribution.

Annex 10. Calculated qualitative/quantitative parameters

Qualitative/quantitative	Ranges*	Laboratory 19						
		1111	1145	1397	1422	1656	1692	1958
Detected species	{*Lm*}	*Lm*	*Lm*	*Lm*	*Lm*	*Lm*	*Lm*	*Lm*
% Species 1		94.7	94.5	94.7	94.3	94.6	94.6	94.8
% Species 2	{<5%}	0.1	0.0	0.1	0.1	0.1	0.1	0.1
Unclassified reads (%)		5.1	5.4	5.1	5.6	5.2	5.2	5.0
Length at >25 x min. coverage (Mbp)	{>2.8 ∧ <3.1}	2.9	2.9	3.0	2.9	3.0	2.9	3.0
Length [1-25] x min. coverage (kbp)	{<250}		8.5	17.8	1.5	9.2	2.9	4.6
No. of contigs at 25 x min. coverage	{>0}	82	155	121	81	94	123	97
No. of contigs [1-25] x min. coverage	{<1000}	1	13	16	2	7	4	3
Average coverage	{>50}	112	90	95	107	105	91	106
No. of reads (x 1000)		2 347	1 836	2 050	2 281	2 256	1 929	2 273
Average read length		141	143	141	140	142	142	141
Average insert size		213	228	217	214	216	218	218
N50 (kbp)		62	34	46	57	59	40	57
QC-status (Bifrost)		OK	OK	OK	OK	OK	OK	OK

Qualitative/quantitative	Ranges*	Laboratory 35						
		1020	1243	1391	1454	1505	1565	1566
Detected species	{*Lm*}	*Lm*	*Lm*	*Lm*	*Lm*	*Lm*	*Lm*	*Lm*
% Species 1		96.1	96.6	96.6	96.3	96.9	97.2	96.1
% Species 2	{<5%}	0.1	0.0	0.0	0.0	0.0	0.0	0.0
Unclassified reads (%)		3.8	3.3	3.4	3.6	3.0	2.7	3.8
Length at >25 x min. coverage (Mbp)	{>2.8 ∧ <3.1}	3.0	2.9	2.9	2.9	3.0	2.9	3.0
Length [1-25] x min. coverage (kbp)	{<250}	0.0	0.0	0.0	0.0	0.0	0.0	0.0
No. of contigs at 25 x min. coverage	{>0}	37	15	23	16	27	27	21
No. of contigs [1-25] x min. coverage	{<1000}	0	0	0	0	0	0	0
Average coverage	{>50}	212	222	216	224	198	184	223
No. of reads (x 1000)		4 577	4 555	4 455	4 714	4 131	3 630	4 830
Average read length		140	143	142	141	145	147	139
Average insert size		261	279	275	274	288	317	253
N50 (kbp)		177	451	201	447	389	266	302
QC-status (Bifrost)		OK	OK	OK	OK	OK	OK	OK

Qualitative/quantitative	Ranges*	Laboratory 49						
		1269	1343	1633	1819	1945	1971	1981
Detected species	{Lm}	*Lm*	*Lm*	*Lm*	*Lm*	*Lm*	*Lm*	*Lm*
% Species 1		97.6	97.4	98.7	98.0	98.2	98.7	96.6
% Species 2	{<5%}	0.0	0.0	0.0	0.1	0.0	0.0	0.3
Unclassified reads (%)		2.3	2.5	1.2	1.9	1.8	1.3	2.9
Length at >25 x min. coverage (Mbp)	{>2.8 ∧ <3.1}	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
No. of contigs at 25 x min. coverage	{>0}	13	14	19	22	14	19	20
No. of contigs [1-25] x min. coverage	{<1000}	0	0	0	0	0	0	0
Average coverage	{>50}	115	136	183	161	140	182	75
No. of reads (x 1000)		1 247	1 516	2 179	1 852	1 583	2 197	830
Average read length		272	271	248	264	266	245	277
Average insert size		352	340	276	327	328	269	422
N50 (kbp)		1500	478	447	476	557	447	344
QC-status (Bifrost)		OK	OK	OK	OK	OK	OK	OK

		Laboratory 56						
Qualitative/quantitative	**Ranges***	**1082**	**1161**	**1582**	**1640**	**1705**	**1873**	**1999**
Detected species	{*Lm*}	*Lm*	*Lm*	*Lm*	*Lm*	*Lm*	*Lm*	*Lm*
% Species 1		98.6	98.8	98.8	98.7	98.8	98.8	98.5
% Species 2	{<5%}	0.2	0.0	0.0	0.0	0.0	0.0	0.0
Unclassified reads (%)		1.2	1.1	1.2	1.2	1.2	1.1	1.5
Length at >25 x min. coverage (Mbp)	{>2.8 ∧ <3.1}	3.0	3.0	2.9	2.9	2.9	2.8	2.9
Length [1-25] x min. coverage (kbp)	{<250}	0.0	0.0	0.0	0.0	0.0	0.0	0.0
No. of contigs at 25 x min. coverage	{>0}	275	235	213	423	553	419	218
No. of contigs [1-25] x min. coverage	{<1000}	0	0	0	0	0	0	0
Average coverage	{>50}	210	220	201	271	267	211	149
No. of reads (x 1000)		1 940	2 041	1 869	2 515	2 319	1 838	1 557
Average read length		327	324	314	320	346	332	279
Average insert size		8	0	0	0	13	2	7
N50 (kbp)		17	23	26	13	9	11	22
QC-status (Bifrost)		OK	OK	OK	OK	OK	OK	OK

		Laboratory 70						
Qualitative/quantitative	**Ranges***	**1309**	**1398**	**1450**	**1598**	**1606**	**1632**	**1868**
Detected species	{*Lm*}	*Lm*	*Lm*	*Lm*	*Lm*	*Lm*	*Lm*	*Lm*
% Species 1		97.8	98.0	97.8	97.9	98.0	98.0	97.8
% Species 2	{<5%}	0.0	0.0	0.1	0.0	0.0	0.0	0.0
Unclassified reads (%)		2.2	2.0	2.1	2.0	2.0	2.0	2.1
Length at >25 x min. coverage (Mbp)	{>2.8 ∧ <3.1}	2.9	2.9	3.0	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	0.0	0.0
No. of contigs at 25 x min. coverage	{>0}	12	15	23	24	12	18	13
No. of contigs [1-25] x min. coverage	{<1000}	0	0	0	0	0	0	0
Average coverage	{>50}	82	92	83	91	88	88	77
No. of reads (x 1000)		1 634	1 860	1 710	1 867	1 747	1 799	1 518
Average read length		147	147	147	147	147	147	147
Average insert size		280	274	262	281	279	266	278
N50 (kbp)		1 491	480	344	477	1 491	478	1 500
QC-status (Bifrost)		OK	OK	OK	OK	OK	OK	OK

		Laboratory 88						
Qualitative/quantitative	**Ranges***	**1104**	**1235**	**1328**	**1417**	**1478**	**1638**	**1959**
Detected species	{*Lm*}	*Lm*	*Lm*	*Lm*	*Lm*	*Lm*	*Lm*	*Lm*
% Species 1		97.8	97.9	97.7	98.1	97.9	98.2	98.0
% Species 2	{<5%}	0.1	0.2	0.1	0.0	0.1	0.0	0.0
Unclassified reads (%)		1.9	1.8	1.9	1.8	1.9	1.6	1.8
Length at >25 x min. coverage (Mbp)	{>2.8 ∧ <3.1}	2.9	3.0	2.9	2.9	2.9	3.0	3.0
Length [1-25] x min. coverage (kbp)	{<250}	2.9	1.7	0.6	0.0	5.6	5.5	6.2
No. of contigs at 25 x min. coverage	{>0}	108	70	85	48	109	132	85
No. of contigs [1-25] x min. coverage	{<1000}	4	2	1	0	9	9	7
Average coverage	{>50}	162	165	154	161	120	117	105
No. of reads (x 1000)		3 201	3 376	3 127	3 206	2 381	2 372	2 147
Average read length		147	147	147	147	147	147	146
Average insert size		314	326	303	323	317	309	306
N50 (kbp)		49	84	70	116	45	41	71
QC-status (Bifrost)		OK	OK	OK	OK	OK	OK	OK

		Laboratory 100						
Qualitative/quantitative	**Ranges***	**1321**	**1360**	**1645**	**1680**	**1905**	**1946**	**1965**
Detected species	{*Lm*}	*Lm*	*Lm*	*Lm*	*Lm*	*Lm*	*Lm*	*Lm*
% Species 1		98.2	98.0	98.3	98.2	98.3	98.0	97.9
% Species 2	{<5%}	0.0	0.0	0.0	0.0	0.0	0.0	0.1
Unclassified reads (%)		1.8	1.9	1.7	1.8	1.7	2.0	1.9
Length at >25 x min. coverage (Mbp)	{>2.8 ∧ <3.1}	2.9	3.0	2.9	2.9	3.0	2.9	3.0
Length [1-25] x min. coverage (kbp)	{<250}	0.0	0.0	0.0	0.0	0.0	0.0	0.0
No. of contigs at 25 x min. coverage	{>0}	13	21	12	13	18	22	24
No. of contigs [1-25] x min. coverage	{<1000}	0	0	0	0	0	0	0
Average coverage	{>50}	858	948	902	913	1096	391	767
No. of reads (x 1000)		16 983	19 703	17 995	18 264	22 359	8 045	16 002
Average read length		146	145	147	146	147	144	145
Average insert size		280	270	284	286	275	210	260
N50 (kbp)		1 500	477	1 491	1 491	478	447	344
QC-status (Bifrost)		OK	OK	OK	OK	OK	OK	OK

		Laboratory 105						
Quali-/Quantitative	**Ranges***	**1059**	**1192**	**1295**	**1423**	**1747**	**1848**	**1985**
Detected species	{*Lm*}	*Lm*	*Lm*	*Lm*	*Lm*	*Lm*	*Lm*	*Lm*
% Species 1		96.7	97.0	97.3	97.3	96.5	97.0	97.2
% Species 2	{<5%}	0.0	0.1	0.0	0.0	0.0	0.0	0.0
Unclassified reads (%)		3.2	2.8	2.7	2.7	3.5	3.0	2.8
Length at >25 x min. coverage (Mbp)	{>2.8 ∧ <3.1}	2.9	3.0	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
No. of contigs at 25 x min. coverage	{>0}	13	21	14	19	16	14	17
No. of contigs [1-25] x min. coverage	{<1000}	0	0	0	0	0	0	0
Average coverage	{>50}	47	47	36	55	47	48	57
No. of reads (x 1000)		950	983	738	1 151	993	970	1 164
Average read length		144	146	145	145	143	144	144
Average insert size		335	334	341	341	369	358	283
N50 (kbp)		1 500	344	1 490	509	478	1 491	478
QC-status (Bifrost)		Warning	Warning	Warning	OK	Warning	Warning	OK

Warning: average coverage is below 50 for some strains.

		Laboratory 108						
Quali-/Quantitative	**Ranges***	**1033**	**1354**	**1503**	**1570**	**1575**	**1741**	**1863**
Detected species	{*Lm*}	*Lm*	*Lm*	*Lm*	*Lm*	*Lm*	*Lm*	*Lm*
% Species 1		97.6	97.4	97.9	97.7	97.9	98.0	97.9
% Species 2	{<5%}	0.0	0.0	0.0	0.0	0.0	0.0	0.1
Unclassified reads (%)		2.4	2.6	2.1	2.2	2.0	2.0	1.9
Length at >25 x min. coverage (Mbp)	{>2.8 ∧ <3.1}	2.7	2.6	2.7	2.9	2.8	2.7	3.0
Length [1-25] x min. coverage (kbp)	{<250}	0.2	0.7	1.3	0.0	3.1	0.3	0.0
No. of contigs at 25 x min. coverage	{>0}	2 684	2 688	2 630	632	2 563	2 825	612
No. of contigs [1-25] x min. coverage	{<1000}	1	3	6	0	13	1	0
Average coverage	{>50}	72	83	81	107	61	89	119
No. of reads (x 1000)		756	869	864	1170	710	983	1 321
Average read length		281	279	278	277	263	273	273
Average insert size		0	0	0	0	0	0	0
N50 (kbp)		2	2	2	8	2	2	8
QC-status (Bifrost)		Warning	Warning	Warning	OK	Warning	Warning	OK

Warning: length at >25X coverage: below 2.8, Some QC values are unreliable due to assembly issues for Ion Torrent data (contigs, average insert size, N50).

		Laboratory 129						
Qualitative/quantitative	**Ranges***	**1175**	**1199**	**1203**	**1549**	**1558**	**1564**	**1913**
Detected species	{*Lm*}	*Lm*	*Lm*	*Lm*	*Lm*	*Lm*	*Lm*	*Lm*
% Species 1		97.6	97.8	97.9	96.5	97.5	97.5	96.4
% Species 2	{<5%}	0.1	0.0	0.0	0.0	0.1	0.0	0.0
Unclassified reads (%)		2.2	2.1	2.0	3.4	2.3	2.4	3.5
Length at >25 x min. coverage (Mbp)	{>2.8 ∧ <3.1}	2.9	2.9	2.9	2.9	3.0	2.9	3.0
Length [1-25] x min. coverage (kbp)	{<250}	29.6	1.3	0.6	3.4	0.0	0.5	0.0
No. of contigs at 25 x min. coverage	{>0}	135	35	40	282	51	36	67
No. of contigs [1-25] x min. coverage	{<1000}	30	2	1	9	0	1	0
Average coverage	{>50}	69	140	171	249	197	142	161
No. of reads (x 1000)		1 386	2 790	3 408	5 348	4 116	2 849	3 392
Average read length		149	146	146	137	145	146	143
Average insert size		382	318	304	195	268	315	333
N50 (kbp)		36	183	175	20	95	168	92
QC-status (Bifrost)		OK	OK	OK	OK	OK	OK	OK

		Laboratory 135						
Qualitative/quantitative	**Ranges***	**1242**	**1263**	**1367**	**1432**	**1673**	**1725**	**1974**
Detected species	{*Lm*}	*Lm*	*Lm*	*Lm*	*Lm*	*Lm*	*Lm*	*Lm*
% Species 1		97.3	96.5	96.5	95.9	96.1	96.3	97.6
% Species 2	{<5%}	0.2	0.2	0.3	1.0	0.8	0.8	0.1
Unclassified reads (%)		2.3	3.3	3.1	3.0	3.0	2.8	2.2
Length at >25 x min. coverage (Mbp)	{>2.8 ∧ <3.1}	3.0	2.9	3.0	2.9	2.9	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
No. of contigs at 25 x min. coverage	{>0}	23	16	16	20	14	14	22
No. of contigs [1-25] x min. coverage	{<1000}	0	0	0	0	0	0	0
Average coverage	{>50}	471	361	204	253	279	403	579
No. of reads (x 1000)		9 746	7 226	4 427	5 410	5 847	8 402	11 929
Average read length		149	149	149	149	149	149	149
Average insert size		311	272	276	282	282	293	302
N50 (kbp)		344	1500	478	478	1491	1491	477
QC-status (Bifrost)		OK	OK	OK	OK	OK	OK	OK

		Laboratory 138						
Qualitative/quantitative	**Ranges***	**1038**	**1041**	**1193**	**1464**	**1473**	**1674**	**1956**
Detected species	{*Lm*}	*Lm*	*Lm*	*Lm*	*Lm*	*Lm*	*Lm*	*Lm*
% Species 1		98.9	98.9	98.8	98.8	98.5	98.7	98.9
% Species 2	{<5%}	0.0	0.0	0.0	0.0	0.2	0.0	0.0
Unclassified reads (%)		1.1	1.1	1.2	1.2	1.3	1.3	1.1
Length at >25 x min. coverage (Mbp)	{>2.8 ∧ <3.1}	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
No. of contigs at 25 x min. coverage	{>0}	13	14	12	16	24	19	13
No. of contigs [1-25] x min. coverage	{<1000}	0	0	0	0	0	0	0
Average coverage	{>50}	287	285	331	289	261	279	304
No. of reads (x 1000)		5 837	5 661	6 510	5 901	5 364	5 623	6 039
Average read length		148	148	148	148	148	148	148
Average insert size		401	381	362	372	365	372	376
N50 (kbp)		540	1491	1500	491	225	478	1491
QC-status (Bifrost)		OK	OK	OK	OK	OK	OK	OK

		Laboratory 141						
Qualitative/quantitative	**Ranges***	**1285**	**1292**	**1611**	**1637**	**1782**	**1978**	**1993**
Detected species	{*Lm*}	*Lm*	*Lm*	*Lm*	*Lm*	*Lm*	*Lm*	*Lm*
% Species 1		97.5	97.7	97.4	97.4	97.6	96.7	97.1
% Species 2	{<5%}	0.1	0.1	0.1	0.0	0.1	0.1	0.1
Unclassified reads (%)		2.4	2.1	2.4	2.5	2.2	3.2	2.7
Length at >25 x min. coverage (Mbp)	{>2.8 ∧ <3.1}	2.9	2.9	3.0	3.0	3.0	2.9	2.9
Length [1-25] x min. coverage (kbp)	{<250}	0.0	0.3	2.9	0.0	3.7	8.7	10.6
No. of contigs at 25 x min. coverage	{>0}	58	81	78	32	67	125	67
No. of contigs [1-25] x min. coverage	{<1000}	0	1	5	0	3	10	9
Average coverage	{>50}	68	73	85	170	75	56	70
No. of reads (x 1000)		800	858	1043	2069	904	657	799
Average read length		254	252	246	250	251	253	255
Average insert size		312	295	291	297	301	341	325
N50 (kbp)		83	54	81	254	85	38	90
QC-status (Bifrost)		OK	OK	OK	OK	OK	OK	OK

		Laboratory 142						
Qualitative/quantitative	**Ranges***	**1131**	**1386**	**1416**	**1436**	**1453**	**1596**	**1858**
Detected species	{*Lm*}	*Lm*	*Lm*	*Lm*	*Lm*	*Lm*	*Lm*	*Lm*
% Species 1		97.0	97.9	97.6	97.5	97.7	98.0	97.3
% Species 2	{<5%}	0.2	0.1	0.2	0.0	0.2	0.1	0.1
Unclassified reads (%)		2.7	1.9	2.0	2.2	2.1	1.8	2.4
Length at >25 x min. coverage (Mbp)	{>2.8 ∧ <3.1}	2.9	3.0	3.0	2.9	2.9	2.9	3.0
Length [1-25] x min. coverage (kbp)	{<250}	1.7	0.5	0.0	3.6	0.0	0.5	0.0
No. of contigs at 25 x min. coverage	{>0}	23	21	21	18	15	17	20
No. of contigs [1-25] x min. coverage	{<1000}	3	1	0	5	0	1	0
Average coverage	{>50}	63	57	56	62	66	63	62
No. of reads (x 1000)		771	701	693	751	788	770	776
Average read length		242	245	245	245	243	245	244
Average insert size		452	447	449	431	428	444	435
N50 (kbp)		405	449	302	360	549	445	478
QC-status (Bifrost)		OK	OK	OK	OK	OK	OK	OK

		Laboratory 144						
Qualitative/quantitative	**Ranges***	**1139**	**1164**	**1517**	**1573**	**1718**	**1954**	**1972**
Detected species	{*Lm*}	*Lm*	*Lm*	*Lm*	*Lm*	*Lm*	*Lm*	*Lm*
% Species 1		97.1	98.3	98.1	97.6	97.8	98.0	98.5
% Species 2	{<5%}	0.0	0.0	0.0	0.1	0.0	0.0	0.0
Unclassified reads (%)		2.9	1.7	1.9	2.1	2.1	2.0	1.4
Length at >25 x min. coverage (Mbp)	{>2.8 ∧ <3.1}	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
No. of contigs at 25 x min. coverage	{>0}	14	14	18	20	15	14	20
No. of contigs [1-25] x min. coverage	{<1000}	0	0	0	0	0	0	0
Average coverage	{>50}	111	123	113	104	84	94	83
No. of reads (x 1000)		2 208	2 433	2 303	2 131	1 696	1 881	1 684
Average read length		148	146	145	148	148	147	147
Average insert size		306	236	219	310	321	264	271
N50 (kbp)		1 491	1 500	447	344	553	1 491	478
QC-status (Bifrost)		OK	OK	OK	OK	OK	OK	OK

		Laboratory 149						
Quali-/Quantitative	**Ranges***	**1036**	**1054**	**1069**	**1284**	**1619**	**1667**	**1814**
Detected species	{*Lm*}	*Lm*	*Lm, Pt*	*Lm, Pt*	*Lm, Pt*	*Lm*	*Lm*	*Lm, Pt*
% Species 1		89.5	82.1	82.0	86.9	89.9	91.3	84.1
% Species 2	{<5%}	3.5	6.9	6.9	5.3	2.8	1.9	5.9
Unclassified reads (%)		6.5	9.1	9.4	6.9	6.7	6.2	7.4
Length at >25 x min. coverage (Mbp)	{>2.8 ∧ <3.1}	2.9	2.9	3.0	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	0.7
No. of contigs at 25 x min. coverage	{>0}	21	17	38	18	19	22	20
No. of contigs [1-25] x min. coverage	{<1000}	0	0	0	0	0	0	1
Average coverage	{>50}	90	88	82	68	104	101	71
No. of reads (x 1000)		1 951	2 125	2 057	1 552	2 283	2 180	1 705
Average read length		151	151	151	151	151	151	151
Average insert size		311	289	309	320	293	307	316
N50 (kbp)		447	1 500	237	1 491	369	477	478
QC-status (Bifrost)		OK	Warning	Warning	Warning	OK	OK	Warning

Warning: species % is below threshold 95%. Pseudomonas tolaasii identified as additional species. This contamination was also observed in the Salmonella EQA-11.

Quality assessment made using the EQA provider's in-house quality control pipeline.
**: indicative QC ranges*
NA: not available
Lm: L. monocytogenes
Pt: Pseudomonas tolaasii.

Annex 11. Accessing QC status of sequences provided

Lab ID	Sero	ST	Cluster	QC status	Description Strain8
EQA provider	**IIa**	**7**	**No**	**B/C**	**A non-cluster sequence contaminated with approx. 9% *L. innocua***
19				C	The quality of strain 0008 is not accepted. The assembled genome is too large, has a high number of contigs, many ambiguous sites/Ns, the N50 value is low and there are too many multiple consensus calls in the cgMLST analysis, which is an indication of a contamination. Kraken analysis further indicates a likely contamination with *Listeria innocua*. The strain has to be restreaked for pure culture and re-sequenced.
35				C	74.3% good cgMLST targets, estimated genome size 5.7 Mbp
49				C	Fail. Genome (5.68 MB) too large for a pure isolate of *L.monocytogenes*. Large number (927) of multiple alleles suggesting more than one *Listeria* strain. Core genome too low (89%), N50 too small (<15kb) and number of contigs too large (1081) to give satisfactory results. Analysis by rMLST showed both *L.monocytogenes* and *L.innocua* were present.
56	1/2a 3a	7	No	A	Good, phred score over 30
70				C	QC for this strain is bad. No evidence of contamination but a very high number of contigs (5211) and only 6.6% of targets found. It is not possible to give the serotype or any typing result.
88	Non-typeable		No	B	FAIL (potential contamination with *Listeria innocua* and % cgMLST loci called below 95%).
100				C	QC - low % of good cgMLST targets (5,9%) ST not detected, serotype/serogroup not detected, assembled genome size too short (2.7 Mb).
105	IIa		No	A	The 7-gene MLST genotyping failed because target QC procedure failed for dapE: multiple hits above thresholds were found in scan procedure (Seqsphere software). In the *Listeria* PasteurMLST sequence definition database, dapE is defined as allele 35 but ST was also not identified, could be a new ST belonging to CC7.
108				C	Strain 0008 has too large a genome size. Identifies both *L. monocytogenes* and *L. innocua.*
129	IIa	7	No	A	Percentage of good targets and average coverage were acceptable. QC passed.
135				C	Too many contigs, N50 too low, contamination too high, there seem to be multiple *Listeria* species (*welshimeri, monocytogenes, ivanovii* and *innocua*) present in this sample. If possible, this sample should be recultured to obtain pure cultures for each species and sequenced again as separate samples.
138				C	Low N50 (12922. bp), low coverage and low percentage of contigs mapped to *Listeria monocytogenes* (53%). Observed intraspecies contamination with *Listeria innocua*. No MLST or serotype/group could be obtained. Was not used in further cluster analysis.
141				C	QC failed, only 75.8 % good targets potential contamination with *L. innocua* (Mash result) expected genome size too big for *Listeria monocytogenes* (5.7 MB instead of ~ 3 MB).
142				C	Contamination with *L. innocua*
144	IIa		No	A	Acceptable quality, parameters:approximated genome size 3.0 Mb, average coverage: 98xcgMLST good target%: 98.2
149	IIa		No	A	% good targets 98.2 (>90%, ok). Average coverage (assembled) 98 (ok) approx. genome size 3.0 Mb (ok). KRAKEN: *L. monocytogenes* (dominating), *L. innocula* (present).

Lab ID	Sero	ST	Cluster	QC status	Description Strain12
EQA provider				**C**	**A non-cluster sequence with massively reduced coverage and removal of genes**
19				C	The quality of strain 0012 is not accepted. The read coverage is too low, resulting in a poor genome assembly with too small a genome, many contigs and low N50 value. Furthermore, the core % in the cgMLST analysis is too low and the genome is not accepted for further analysis. The strain has to be re-sequenced.
35				C	55.1% good cgMLST targets, estimated genome size 1.8 Mbp, average coverage <30-fold.
49				C	Fail. Average read coverage (14) too low, N50 *10kb) too small, genome length (1.7MB) too small and core percent (38%) too low to give satisfactory results.
56				C	Insufficient coverage (15X)
70				C	39.9% of targets found (cut-off is 95%) number of contigs = 875 (cut-off is 250) coverage = 25 (cut off is 30) approximated genome size = 1.6 Mb instead of 3 Mb.
88				C	OVERALL QC: FAILED (low depth of coverage).
100				C	QC: low % of good targets (39,2 %), low coverage (24x), genome size too low (1.6 Mb)
105				C	Strain 0012 genome FAILS several QCs:- cgMLST percentage of good targets found was only 40.7%- Genome size was 1.5 Mb.

Lab ID	Sero	ST	Cluster	QC status	Description Strain12
108				C	Strain 0012 has too small genome size. Serotype could not be identified.
129				C	Quality of sequences was poor. Percentage of good targets was 43.3 and average coverage was 22.
135				C	Total length too low, too many contigs, N50 too low, GC% too high, average coverage too low, completeness too low. The sequence quality of this sample is too low, but there does not seem to be a contamination. If isolated DNA had enough yield and purity, sequencing can be repeated from DNA.
138				C	Low coverage (23x), GC content within expected range, low N50 (7580 bp) and high percentage of contigs mapped to *Listeria monocytogenes* (99,9%). Number of obtained reads too low. Serotype/group could not be determined. QC fail.
141				C	54.6 % good targets (too low), coverage is not good (below 30). Expected genome size only 1.7 MB instead of ~ 3MB for *Listeria monocytogenes.*
142				C	Low coverage, low % cgMLST
144				C	Unacceptable quality. Parameters: approximated genome size 1.7 Mb (<2.7 Mb). Average coverage: 22x (<50x)cgMLST good target%: 42.4 (<95).
149				C	% good targets cgMLST: 42.4 (<90%, not acceptable) avg. coverage (assembled): 22 (too low) approx. genome size: 1.7 (too small). KRAKEN: *L. monocytogenes.*

Lab ID	Sero	ST	Cluster	QC status	Description Strain15
EQA provider				**C**	**Two non-cluster sequences of ST1 and ST224 combined**
19				C	The quality of strain 0015 is not accepted. The assembled genome is slightly too large, has a high number of contigs, many ambiguous sites/Ns, the N50 value is slightly low and there are too many mulitple consensus calls in the cgMLST analysis, which is an indication of a contamination. Kraken analysis does not indicate contamination with other species and therefore it is probably a contamination with two *Listeria monocytogenes* strains. The strain has to be re-streaked for pure cultures and re-sequenced.
35				C	54.7% good cgMLSt targets, estimated genome size 4.1 Mbp, average coverage >30-fold.
49				C	Fail. Length slightly too long, N50 too small, a lot of Ns and very low (32) core percent. The number of perfect alleles was ok but number of consensus alleles was low. rMLST = *L.monocytogenes*.
56	4b 4d 4e		No	A	Phred score over 30.
70	IVb		No	B	52.7% of targets found, number of contigs = 335 (cut-off is 250) coverage = 137
88				C	OVERALL QC: FAILED Mixture of two strains (intra-species contamination).
100				C	QC: low % of good targets (49.9%), genome size higher than expected (3.2 Mb)
105	IIb		No	B	The QC 'cgMLST percentage of good targets' was 93% (WARN), and 7-gene MLST genotyping FAILS due to incomplete number of targets (no sequence found for 'abcZ' and absence of 'bglA').
108	IVb	1	No	A	Strain 0015. QC status ok.
129				C	Sequence quality was poor. Percentage of good targets was 51.3 % and average coverage was 136.
135				C	Too many contigs, N50 too low, completeness too low, contamination too high. This sample seems to be contaminated with *L. innocua*. It should be repeated from a pure culture.
138				C	Good coverage (66x), GC content within expected range, low N50 (1517 bp) and high percentage of contigs mapped to *Listeria monocytogenes* (98,8%). Sufficient number of reads obtained. MLST and serotype/group could not be determined. Non-Gaussian insert size distribution, inability to assemble contigs. Suspected library preparation error. One explanation would be that products of two different library preparation methods were mixed which contained inserts of differing sizes and amounts. QC fail.
141				C	39.1 % good targets (too low). Coverage is good, expected genome size 3.7 MB is too big for *Listeria monocytogenes* (~ 3 MB) - possibly contaminated.
142				C	Poor sequence quality, per base sequence content low % cgMLST.
144				C	Unacceptable quality. Parameters: approximated genome size 3.3 Mb. Average coverage: 98xcgMLST good target %: 49.9 (<95).
149				C	% good targets cgMLST: 49.9 (unacceptable). Average coverage (assembled): 98 (ok). Genome size: 3.3 KRAKEN: *L. monocytogenes* and some *Listeria* phage LP-030-2.

Lab ID	Sero	ST	Cluster	QC status	Description Strain17
EQA provider	**IIa**	**7**	**Yes**	**B**	**A cluster sequence (Strain14) with reduced core percent from 100% to 94%**
19	IIa	7	Yes	B	Strain 0017 has borderline read coverage and the cgMLST core % is slightly below accepted value, but the strain clusters nicely with the outbreak cluster in the cgMLST and SNP analysis and the genome is accepted for outbreak.
35	IIa	7	No	B	99.1% good cgMLSt targets, estimated genome size 2.9 Mbp, average coverage in the range of 30-fold.
49	IIa	7	Yes	B	Core percent (93) a bit low for good analysis.
56	1/2 a 3a	7	No	A	Phred score over 30.
70	IIa	7	Yes	A	95.4% of targets foundnumber of contigs = 165coverage = 43 (cut-off is 30)
88	1/2a,3a (IIa)	7	No	B	OVERALL QC: FAILED (low depth of coverage and % cgMLST loci called was 91% (below 95%).
100	IIa	7	Yes	B	QC coverage is slightly low (38x).
105	-	-	-	C	Strain 0017 genome FAILS QC 'cgMLST percentage of good targets' which was below 90% (82.9%).
108	IIa	7	No	A	Strain 0017. QC status ok.
129	IIa	7	Yes	A	Percentage of good targets and average coverage were acceptable. QC passed.
135	-	7	No	B	Average coverage too low, number of reads low. Sequencing of this sample did not yield enough reads for an adequate coverage. This should be taken into account when interpreting the result. The existing library prep can probably be used to generate additional reads.
138	1/2a, 3a (IIa)	7	Yes	A	Average coverage (33x), GC content within expected range, high N50 (195811 bp) and high percentage of contigs mapped *to Listeria monocytogenes* (99,8%).QC pass.
141	IIa	7	Yes	A	98.5 % good targets. Coverage is low but acceptable, no contamination found, expected genome size 2.9 MB OK for *Listeria monocytogenes*
142	IIa	7	No	B	Poor quality, per base sequence quality. Reverse low coverage.
144	-	-	-	C	Unacceptable quality. Parameters: approximated genome size 2.9 Mb. Average coverage: 36x (<50x)cgMLST good target%: 92.7 (<95).
149	IIa	7	Yes	A	% good targets cgMLST: 92.7 (OK). Average coverage (assembled): 36 (low). Approx. genome size: 2.9 Mb (OK). KRAKEN: *L. monocytogenes.*

Annex 12. EQA-9 laboratory questionnaire

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

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

1. *Listeria* EQA-9 2022-2023

Dear participant,

Welcome to the ninth External Quality Assessment (EQA-9) scheme for typing of *Listeria* in 2022–2023.
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)

☐ Australia
☐ Austria
☐ Belgium
☐ Denmark
☐ Finland
☐ France
☐ Germany
☐ Hungary
☐ Iceland
☐ Ireland
☐ Italy
☐ Israel
☐ Latvia
☐ Lithuania
☐ Luxembourg
☐ New Zealand
☐ Norway
☐ Portugal
☐ Slovakia
☐ Slovenija
☐ Spain
☐ Sweden
☐ The Netherlands
☐ Turkey
☐ United Kingdom
☐ United States.

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. Listeria EQA-9 Strain IDs

Please enter the strain ID (4 digits)

Listeria

Strain 1 ___
Strain 2 ___
Strain 3 ___
Strain 4 ___
Strain 5 ___
Strain 6 ___
Strain 7 ___

8. Serotyping/grouping of *Listeria*

9. Submitting results

(State one answer only)

☐ Online here
☐ Did not participate in the serotyping/grouping part - Go to 14

10. Submitting results - Serotyping/grouping of *Listeria*

(State one answer only)

☐ Both molecular and conventional serogrouping/serotyping - Go to 11
☐ Molecular serogrouping - Go to 11
☐ Conventional serotyping - Go to 13

11. Method used for molecular serogrouping of *Listeria*

(State one answer only)

☐ PCR-based
☐ WGS-based

12. Results for serotyping/grouping *Listeria* - molecular serogrouping

Please select the serogroup

(State only one answer per question)

Strain	Molecular serotype					
	IIa	IIb	IIc	IVb	L	Un-typeable
Strain 1						
Strain 2						
Strain 3						
Strain 4						
Strain 5						
Strain 6						
Strain 7						

13. Results for serotyping *Listeria* - Conventional serotyping

Please select the serotype

(State only one answer per question)

Strain		Conventional serotype													
Strain 1	1/2a	1/2b	1/2c	3a	3b	3c	4a	4ab	4b	4c	4d	4e	7	Autoagglutinable	Un-typeable
Strain 2															
Strain 3															
Strain 4															
Strain 5															
Strain 6															
Strain 7															

14. Submitting cluster results

(State one answer only)

☐ Cluster analyses based on PFGE and/or WGS
☐ Did not participate in the Cluster part - Go to 222.

15. Submitting cluster results

(State one answer only)

☐ Cluster analysis based on PFGE - Go to 16
☐ Do not wish to submit any cluster results based on PFGE analysis - Go to 20

16. Cluster analysis based on PFGE data

17. 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.

18. *ApaI* - Total number of bands (>33kb) in a cluster strain

(Use 9999 if not analysed)

19. *AscI* – Total number of bands (>33kb) in a cluster strain

(Use 9999 if not analysed)

20. Submitting cluster results

(State one answer only)

☐ Cluster analysis based on WGS data - Go to 21
☐ Do not wish to submit any cluster results based on WGS data - Go to 222.

21. Cluster analysis based on WGS data

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

The results of the cluster detection can only be reported once (main analysis). If more than one analysis is performed, please report later in this submission.

(State one answer only)

☐ SNP-based – Go to 24
☐ Allele-based – Go to 31
☐ Other – Go to 23.

23. If another analysis is used, please describe in detail your approach (including: assembler, number of loci, variant caller, read mapper or reference ID etc.)– Go to 38.

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

25. Please select the approach used for the SNP analysis

(State one answer only)

☐ Reference based – Go to 26
☐ Assembly based – Go to 39.

26. Reference genome used

Preferable use EQA strain 0009 (downloaded sequences) as reference. Otherwise indicate Multi-locus Sequence Type (e.g. ST8) and identification of the reference used.

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

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

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

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

31. Please select tools used for the allele analysis

(State one answer only)

☐ BioNumerics – Go to 33
☐ SeqPhere – Go to 33
☐ BIGSdb-*Lm* – Go to 33
☐ Other – Go to 32.

32. If another tool is used please enter here:

33. Please indicate allele calling method

(State one answer only)

☐ Assembly based and mapping based – Go to 34
☐ Only assembly based – Go to 34
☐ Only mapping based – Go to 35.

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

35. Please select scheme used for the allele analysis

(State one answer only)

☐ Applied Math (wgMLST) – Go to 37
☐ Applied Math (cgMLST/Pasteur) – Go to 37
☐ Pasteur (cgMLST) – Go to 37
☐ Ruppitsch (cgMLST) – Go to 37
☐ Other – Go to 36.

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

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

38. Cluster detected by analysis on data derived from WGS

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

Please fill in all the data for the strains one by one.

39. Strain 1

Report the MLST, serotype/group, part of the cluster and SNP distance/allele difference

40. Report the MLST, serotype/group, part of the cluster and SNP distance /allele difference

41. Report the 7-gene MLST

(State value between 0 and 1000000)

42. Report if this strain is a part of identified cluster

(State one answer only)

☐ Yes
☐ No

43. Report the allele difference/SNP distance to the strain 9 (as 0009 downloaded sequence)

(State value)

44. Strain 2

Report the MLST, serotype/group, part of the cluster and SNP distance /allele difference

45. Report the MLST, serotype/group, part of the cluster and SNP distance /allele difference

46. Report the 7-gene MLST

(State value between 0 and 1000000)

47. Report if this strain is a part of identified cluster

(State one answer only)

☐ Yes
☐ No

48. Report the allele difference/SNP distance to the strain 9 (as 0009 downloaded sequence)

(State value)

49. Strain 3

Report the MLST, serotype/group, part of the cluster and SNP distance /allele difference

50. Report the MLST, serotype/group, part of the cluster and SNP distance /allele difference

51. Report the 7-gene MLST

(State value between 0 and 1000000)

52. Report if this strain is a part of identified cluster

(State one answer only)

☐ Yes
☐ No.

53. Report the allele difference/SNP distance to the strain 9 (as 0009 downloaded sequence)

(State value)

54. Strain 4

Report the MLST, serotype/group, part of the cluster and SNP distance /allele difference

55. Report the MLST, serotype/group, part of the cluster and SNP distance /allele difference

56. Report the 7-gene MLST

(State value between 0 and 1000000)

57. Report if this strain is a part of identified cluster

(State one answer only)

☐ Yes
☐ No

58. Report the allele difference/SNP distance to the strain 9 (as 0009 downloaded sequence)

(State value)

59. Strain 5

Report the MLST, serotype/group, part of the cluster and SNP distance /allele difference

60. Report the MLST, serotype/group, part of the cluster and SNP distance /allele difference

61. Report the 7-gene MLST

(State value between 0 and 1000000)

62. Report if this strain is a part of identified cluster

(State one answer only)

☐ Yes
☐ No.

63. Report the allele difference/SNP distance to the strain 9 (as 0009 downloaded sequence)

(State value)

64. Strain 6

Report the MLST, serotype/group, part of the cluster and SNP distance/allele difference.

65. Report the MLST, serotype/group, part of the cluster and SNP distance /allele difference

66. Report the 7-gene MLST

(State value between 0 and 1000000)

67. Report if this strain is a part of identified cluster

(State one answer only)

☐ Yes
☐ No.

68. Report the allele difference/SNP distance to the strain 9 (as 0009 downloaded sequence)

(State value)

69. Strain 7

Report the MLST, serotype/group, part of the cluster and SNP distance/allele difference.

70. Report the MLST, serotype/group, part of the cluster and SNP distance /allele difference

71. Report the 7-gene MLST

(State value between 0 and 1000000)

72. Report if this strain is a part of identified cluster

(State one answer only)

☐ Yes
☐ No.

73. Report the allele difference/SNP distance to the strain 9 (as 0009 downloaded sequence)

(State value)

74. Strain 0008 (as downloaded sequence)

Report the QC status, MLST, serotype/group, part of the cluster and SNP distance/allele difference.

75. QC observations

Please evaluate the QC results of the strain and explain what you observed.

76. Please select the QC status that fits with your assessment of the strain

(State one answer only)

☐ Acceptable quality
☐ Quality only acceptable for outbreak situations (less good quality)
☐ Unacceptable quality - strain not analysed - Go to 82.

77. Strain 0008 continue

Report the QC status, MLST, serotype/group, part of the cluster and SNP distance allele difference.

78. (Optional) Report the serotype/group

79. Report the 7-gene MLST

(State value between 0 and 1000000)

80. Report if this strain is a part of identified cluster

(State one answer only)

☐ Yes
☐ No.

81. Report the allele difference/SNP distance to the strain 9 (as 0009 downloaded sequence)

Please use 9999 for not analysed.

(State value)

82. Strain 0009 (as downloaded sequence)

Report the QC status, MLST, serotype/group, part of the cluster and SNP distance/allele difference.

83. QC observations

Please evaluate the QC results of the strain and explain what you observed.

84. Please select the QC status that fits with your assessment of the strain

(State one answer only)

☐ Acceptable quality
☐ Quality only acceptable for outbreak situations (less good quality)
☐ Unacceptable quality - strain not analysed - Go to 90.

85. Strain 0009 continue

Report the QC status, MLST, serotype/group, part of the cluster and SNP distance/allele difference.

86. (Optional) Report the serotype/group

87. Report the 7-gene MLST

(State value between 0 and 1000000)

88. Report if this strain is a part of identified cluster

(State one answer only)

☐ Yes
☐ No.

89. Report the allele difference/SNP distance to the strain 9 (as 0009 downloaded sequence)

Please use 9999 for not analysed.

(State value)

90. Strain 0010 (as downloaded sequence)

Report the QC status, MLST, serotype/group, part of the cluster and SNP distance/allele difference.

91. QC observations

Please evaluate the QC results of the strain and explain what you observed.

92. Please select the QC status that fits with your assessment of the strain

(State one answer only)

☐ Acceptable quality
☐ Quality only acceptable for outbreak situations (less good quality)
☐ Unacceptable quality – strain not analysed – Go to 98.

93. Strain 0010 continue

Report the QC status, MLST, serotype/group, part of the cluster and SNP distance/allele difference.

94. (Optional) Report the serotype/group

95. Report the 7-gene MLST

(State value between 0 and 1000000)

96. Report if this strain is a part of identified cluster

(State one answer only)

☐ Yes
☐ No.

97. Report the allele difference/SNP distance to the strain 9 (as 0009 downloaded sequence)

Please use 9999 for not analysed

(State value)

98. Strain 0011 (as downloaded sequence)

Report the QC status, MLST, serotype/group, part of the cluster and SNP distance/allele difference.

99. QC observations

Please evaluate the QC results of the strain and explain what you observed.

100. Please select the QC status that fits with your assessment of the strain

(State one answer only)

☐ Acceptable quality
☐ Quality only acceptable for outbreak situations (less good quality)
☐ Unacceptable quality - strain not analysed - Go to 106.

101. Strain 0011 continue

Report the QC status, MLST, serotype/group, part of the cluster and SNP distance/allele difference.

102. (Optional) Report the serotype/group

103. Report the 7-gene MLST

(State value between 0 and 1000000)

104. Report if this strain is a part of identified cluster

(State one answer only)

☐ Yes
☐ No.

105. Report the allele difference/SNP distance to the strain 9 (as 0009 downloaded sequence)

Please use 9999 for not analysed

(State value)

106. Strain 0012 (as downloaded sequence)

Report the QC status, MLST, serotype/group, part of the cluster and SNP distance/allele difference.

107. QC observations

Please evaluate the QC results of the strain and explain what you observed.

108. Please select the QC status that fits with your assessment of the strain

(State one answer only)

☐ Acceptable quality
☐ Quality only acceptable for outbreak situations (less good quality)
☐ Unacceptable quality - strain not analysed - Go to 114.

109. Strain 0012 continue

Report the QC status, MLST, serotype/group, part of the cluster and SNP distance/allele difference.

110. (Optional) Report the serotype/group

111. Report the 7-gene MLST

(State value between 0 and 1000000)

112. Report if this strain is a part of identified cluster

(State one answer only)

☐ Yes
☐ No.

113. Report the allele difference/SNP distance to the strain 9 (as 0009 downloaded sequence)

Please use 9999 for not analysed

(State value)

114. Strain 0013 (as downloaded sequence)

Report the QC status, MLST, serotype/group, part of the cluster and SNP distance/allele difference.

115. QC observations

Please evaluate the QC results of the strain and explain what you observed.

116. Please select the QC status that fits with your assessment of the strain

(State one answer only)

☐ Acceptable quality
☐ Quality only acceptable for outbreak situations (less good quality)
☐ Unacceptable quality - strain not analysed - Go to 122.

117. Strain 0013 continue

Report the QC status, MLST, serotype/group, part of the cluster and SNP distance/allele difference.

118. (Optional) Report the serotype/group

119. Report the 7-gene MLST

(State value between 0 and 1000000)

120. Report if this strain is a part of identified cluster

(State one answer only)

☐ Yes
☐ No

121. Report the allele difference/SNP distance to the strain 9 (as 0009 downloaded sequence)

Please use 9999 for not analysed

(State value)

122. Strain 0014 (as downloaded sequence)

Report the QC status, MLST, serotype/group, part of the cluster and SNP distance /allele difference

123. QC observations

Please evaluate the QC results of the strain and explain what you observe

124. Please select the QC status that fits with your assessment of the strain

(State one answer only)

☐ Acceptable quality
☐ Quality only acceptable for outbreak situations (less good quality)
☐ Not acceptable quality - strain not analysed - Go to 130

125. Strain 0014 continue

Report the QC status, MLST, serotype/group, part of the cluster and SNP distance /allele difference

126. (Optional) Report the serotype/group

127. Report the 7-gene MLST

(State value between 0 and 1000000)

128. Report if this strain is a part of identified cluster

(State one answer only)

☐ Yes
☐ No

129. Report the allele difference/SNP distance to the strain 9 (as 0009 downloaded sequence)

Please use 9999 for not analysed

(State value)

130. Strain 0015 (as downloaded sequence)

Report the QC status, MLST, serotype/group, part of the cluster and SNP distance /allele difference

131. QC observations

Please evaluate the QC results of the strain and explain what you observe

132. Please select the QC status that fits with your assessment of the strain

(State one answer only)

☐ Acceptable quality
☐ Quality only acceptable for outbreak situations (less good quality)
☐ Not acceptable quality - strain not analysed - Go to 138

133. Strain 0015 continue

Report the QC status, MLST, serotype/group, part of the cluster and SNP distance /allele difference

134. (Optional) Report the serotype/group

135. Report the 7-gene MLST

(State value between 0 and 1000000)

136. Report if this strain is a part of identified cluster

(State one answer only)

☐ Yes
☐ No

137. Report the allele difference/SNP distance to the strain 9 (as 0009 downloaded sequence)

Please use 9999 for not analysed

(State value)

138. Strain 0016 (as downloaded sequence)

Report the QC status, MLST, serotype/group, part of the cluster and SNP distance /allele difference

139. QC observations

Please evaluate the QC results of the strain and explain what you observe

140. Please select the QC status that fits with your assessment of the strain

(State one answer only)

☐ Acceptable quality
☐ Quality only acceptable for outbreak situations (less good quality)
☐ Not acceptable quality - strain not analysed - Go to 146

141. Strain 0016 continue

Report the QC status, MLST, serotype/group, part of the cluster and SNP distance /allele difference

142. (Optional) Report the serotype/group

143. Report the 7-gene MLST

(State value between 0 and 1000000)

144. Report if this strain is a part of identified cluster

(State one answer only)

☐ Yes
☐ No

145. Report the allele difference/SNP distance to the strain 9 (as 0009 downloaded sequence)

Please use 9999 for not analysed

(State value)

146. Strain 0017 (as downloaded sequence)

Report the QC status, MLST, serotype/group, part of the cluster and SNP distance /allele difference

147. QC observations

Please evaluate the QC results of the strain and explain what you observe

148. Please select the QC status that fits with your assessment of the strain

(State one answer only)

☐ Acceptable quality
☐ Quality only acceptable for outbreak situations (less good quality)
☐ Not acceptable quality - strain not analysed - Go to 154.

149. Strain 0017 continue

Report the QC status, MLST, serotype/group, part of the cluster and SNP distance /allele difference

150. (Optional) Report the serotype/group

151. Report the 7-gene MLST

(State value between 0 and 1000000)

152. Report if this strain is a part of identified cluster

(State one answer only)

☐ Yes
☐ No.

153. Report the allele difference/SNP distance to the strain 9 (as 0009 downloaded sequence)

Please use 9999 for not analysed

(State value)

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

For example, 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 155
☐ No - Go to 192.

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

(State one answer only)

☐ SNP-based – Go to 157
☐ Allele-based – Go to 164
☐ Other – Go to 156.

156. If another analysis is used please describe in detail your approach (including: assembler, number of loci, variant caller, read mapper or reference ID etc.) - Go to 171

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

158. Please select the approach used for the SNP analysis

(State one answer only)

☐ Reference based – Go to 159
☐ Assembly based – Go to 162.

159. Reference genome used: (preferable use EQA strain 0009, downloaded sequences as reference)

Otherwise indicate Multi-locus Sequence Type (e.g. ST8) and isolate ID

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

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

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

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

164. Please select tool used for the allele analysis

(State one answer only)

☐ BioNumerics – Go to 166
☐ SeqPhere – Go to 166
☐ BIGSdb-*Lm* – Go to 166
☐ Other – Go to 165.

165. If another tool is used please list here:

166. Please indicate allele calling method:

(State one answer only)

☐ Assembly based and mapping based – Go to 167
☐ Only assembly based – Go to 167
☐ Only mapping based – Go to 168.

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

168. Please select scheme used for the allele analysis

(State one answer only)

☐ Applied Math (wgMLST) – Go to 170
☐ Applied Math (cgMLST/Pasteur) – Go to 170
☐ Pasteur (cgMLST) – Go to 170
☐ Ruppitsch (cgMLST) – Go to 170
☐ Other – Go to 169.

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

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

171. Additional analysis on data derived from WGS.

172. Results for an additional cluster analysis.

Reporting allele differences/SNP distances to strain 0009 (as downloaded sequence) (e.g. SNP or allele-based)

Please use 9999 for not analysed.

Isolate	Distance/difference (e.g. SNP/allele) to the strain 0009 (downloaded sequence)
Strain 1	
Strain 2	
Strain 3	
Strain 4	
Strain 5	
Strain 6	
Strain 7	
Strain 0008 (as downloaded sequence)	
Strain 0009 (as downloaded sequence)	
Strain 0010 (as downloaded sequence)	
Strain 0011 (as downloaded sequence)	
Strain 0012 (as downloaded sequence)	
Strain 0013 (as downloaded sequence)	
Strain 0014 (as downloaded sequence)	
Strain 0015 (as downloaded sequence)	
Strain 0016 (as downloaded sequence)	
Strain 0017 (as downloaded sequence)	

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

For example, 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 174
☐ No – Go to 192.

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

(State one answer only)

☐ SNP-based – Go to 176
☐ Allele- based – Go to 183
☐ Other – Go to 175.

175. If another analysis is used, please describe in detail your approach (including: assembler, number of loci, variant caller, read mapper or reference ID etc.) - Go to 190

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

177. Please select the approach used for the SNP analysis

(State one answer only)

☐ Reference-based – Go to 178
☐ Assembly-based – Go to 181.

178. Reference genome used: (preferable use EQA strain 0009, downloaded sequences as reference)

Otherwise indicate Multi-locus Sequence Type (e.g. ST8) and isolate ID.

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

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

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

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

183. Please select tool used for the allele analysis

(State one answer only)

☐ BioNumerics – Go to 185
☐ SeqPhere – Go to 185
☐ BIGSdb-*Lm* – Go to 185
☐ Other – Go to 184.

184. If another tool is used please enter here:

185. Please indicate allele calling method:

(State one answer only)

☐ Assembly-based and mapping-based – Go to 186
☐ Only assembly-based – Go to 186
☐ Only mapping-based – Go to 187.

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

187. Please select scheme used for the allele analysis

(State one answer only)

☐ Applied Math (wgMLST) – Go to 189
☐ Applied Math (cgMLST/Pasteur) – Go to 189
☐ Pasteur (cgMLST) – Go to 189
☐ Ruppitsch (cgMLST) – Go to 189
☐ Other – Go to 188.

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

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

190. Third analysis on data derived from WGS

191. Results for the third cluster analysis.

Reporting allele differences /SNP distances to strain 0009 (as downloaded sequence) (e.g. SNP or allele-based)

Please use 9999 for not analysed.

Isolate	Distance/difference (e.g. SNP/allele) to the strain 0009 (downloaded sequence)
Strain 1	
Strain 2	
Strain 3	
Strain 4	
Strain 5	
Strain 6	
Strain 7	
Strain 0008 (as downloaded sequence)	
Strain 0009 (as downloaded sequence)	
Strain 0010 (as downloaded sequence)	
Strain 0011 (as downloaded sequence)	
Strain 0012 (as downloaded sequence)	
Strain 0013 (as downloaded sequence)	
Strain 0014 (as downloaded sequence)	
Strain 0015 (as downloaded sequence)	
Strain 0016 (as downloaded sequence)	
Strain 0017 (as downloaded sequence)	

192. Additional questions for the WGS part

193. Where was the sequencing performed?

(State one answer only)

☐ In own laboratory
☐ Externally.

194. Protocol used to prepare the library for sequencing

(State one answer only)

☐ Commercial kits – Go to 195
☐ Non-commercial kits – Go to 197.

195. Please indicate name of commercial kit:

196. If relevant, please list deviation from commercial kit briefly in a few bullet points: - Go to 198

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

198. The sequencing platform used

(State one answer only)

☐ Ion Torrent PGM - Go to 200
☐ Ion Torrent Proton - Go to 200
☐ Genome Sequencer Junior System (454) - Go to 200
☐ Genome Sequencer FLX System (454) - Go to 200
☐ Genome Sequencer FLX+ System (454) - Go to 200
☐ PacBio RS - Go to 200
☐ PacBio RS II - Go to 200
☐ HiScanSQ - Go to 200
☐ HiSeq 1000 - Go to 200
☐ HiSeq 1500 - Go to 200
☐ HiSeq 2000 - Go to 200
☐ HiSeq 2500 - Go to 200
☐ HiSeq 4000 - Go to 200
☐ Genome Analyzer Iix - Go to 200
☐ MiSeq - Go to 200
☐ MiSeq Dx - Go to 200
☐ MiSeq FGx - Go to 200
☐ ABI SOLiD - Go to 200
☐ NextSeq - Go to 200
☐ MinION (ONT) - Go to 200
☐ Other - Go to 199.

199. If another platform is used please list here:

200. 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 five selected criteria which were the most frequently reported by in previous EQAs.

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

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

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

☐ Yes
☐ No – Go to 203.

202. Procedure used to evaluate confirmation of genus:

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

(State one answer only)

☐ Yes
☐ No – Go to 205.

204. Procedure or threshold used for coverage:

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

(State one answer only)

☐ Yes
☐ No – Go to 207.

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

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

(State one answer only)

☐ Yes
☐ No – Go to 209.

208. Procedure or threshold used for genome size:

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

(State one answer only)

☐ Yes
☐ No – Go to 211.

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

211. ONLY list additional information related to other criteria used to evaluate the quality of sequence data.

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

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

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

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

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

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

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

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

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

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

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

222. Comment(s):

For example, remarks on the submission, the data analyses or the laboratory methods.

223. Thank you for your participation

Thank you for filling out the submission form for the *Listeria* EQA-9.

For questions, please email: list.eqa@ssi.dk or telephone: +45 3268 8341.

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.

European Centre for Disease Prevention and Control (ECDC)

Gustav III:s Boulevard 40
16973 Solna, Sweden

Tel. +46 858601000
ECDC.info@ecdc.europa.eu

www.ecdc.europa.eu

Follow ECDC on social media
Twitter: @ECDC_EU
Facebook: www.facebook.com/ECDC.EU
Linkedin: www.linkedin.com/company/ecdc/

Publications Office
of the European Union