EXTERNAL QUALITY ASSESSMENT REPORT

Twelfth external quality assessment scheme for typing of Shiga toxin-producing *Escherichia coli*

ECDC TECHNICAL REPORT

Twelfth external quality assessment scheme for typing of Shiga toxin-producing *Escherichia coli*

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 Nadia Boisen, Susanne Schjørring, Gitte Sørensen, Anne Sophie Majgaard Uldall, Flemming Scheutz, and Eva Møller Nielsen of the Foodborne Infections Unit at Statens Serum Institut, Copenhagen, Denmark.

Suggested citation: European Centre for Disease Prevention and Control. Twelfth external quality assessment scheme for typing of Shiga toxin-producing *Escherichia coli*. Stockholm: ECDC; 2024.

Stockholm, June 2024

ISBN 978-92-9498-730-3
doi 10.2900/68503
Catalogue number TQ-05-24-530-EN-N

Contents

Abbreviations

AEEC	Attaching and effacing E. coli
aggR	Gene encoding the master regulator in enteroaggregative *E. coli*
BN	BioNumerics
bp	Base pair
cgMLST	Core genome multilocus sequence typing
DEC	Diarrhoeagenic *E. coli*
EAEC	Enteroaggregative *E. coli*
EFSA	European Food Safety Authority
EQA	External quality assessment
esta	heat stable (ST) enterotoxin gene
ETEC	Enterotoxigenic *E. coli*
FWD	Food- and waterborne diseases
FWD-Net	Food- and Waterborne Diseases and Zoonoses Network
HUS	Haemolytic uraemic syndrome
ND	Not done
NPHRL	National public health reference laboratory
NSF	Non-sorbitol fermenter
PCR	polymerase chain reaction
PFGE	Pulsed-field gel electrophoresis
QC	Quality control
R1/R2	Read1 (forward)/Read2 (reverse) from a paired-end run (FASTQ file)
SF	Sorbitol fermenting
SKESA	Strategic k-mere extension for scrupulous assemblies
SNP	Single nucleotide polymorphism
SPAdes	St. Petersburg genome assembler
SSI	Statens Serum Institut
ST	Sequence type
STEC	Shiga toxin-producing *E. coli* (synonymous with verocytotoxin-producing *E. coli*; VTEC)
Stx1	Shiga toxin 1
stx1	Gene encoding Shiga toxin 1
Stx2	Shiga toxin 2
stx2	Gene encoding Shiga toxin 2
TESSy	The European Surveillance System
wgMLST	Whole genome multilocus sequence typing
WGS	Whole genome sequencing

Executive summary

This report presents the results of the 12th round of the external quality assessment (EQA-12) scheme for typing of Shiga toxin-producing *Escherichia coli* (STEC). This EQA was organised for national public health reference laboratories (NPHRLs) providing data to the Food- and Waterborne Diseases and Zoonoses Network (FWD-Net) managed by the European Centre for Disease Prevention and Control (ECDC). Since 2012, the unit of Foodborne Infections at Statens Serum Institut (SSI) in Denmark has arranged the EQA under a framework contract with ECDC. EQA-12 contained serotyping, detection of virulence genes, and molecular typing-based cluster analysis.

Twenty-six laboratories participated in the EQA-12 scheme, with 25 (96%) performing the serotyping part, 25 (96%) determining the virulence profile, and 23 (88%) engaging in cluster identification using WGS data analysed by different approaches. In O:H serotyping, an average score of 97% was achieved by participants. The performance in detecting the virulence genes was also high: 97% for *stx1* and 98% for s*tx2*, 98% for aggR and *esta* genes, and 96% for *eae* gene. The average score of laboratories that correctly performed the *stx* subtyping were 95% for *stx1*, 94% for *stx2*, and 93% combined stx1 and stx2. In general, the performance of the cluster detection was high, with 15/23 (65%) laboratories correctly identifying the cluster of closely related strains.

Human STEC infection is a zoonotic disease. For 2022, 8 565 confirmed cases of STEC infection were reported by 29 EU/EEA countries. Twenty-three countries reported at least one confirmed STEC case and three countries reported zero cases (Bulgaria, Cyprus and Lithuania). In 2022, the EU notification rate was 2.1 per 100 000 population. For 2022, there was an increase of 8.8% in the annual notification rate reported compared with year 2021 (1.9 cases per 100 000 population). In 2022, the six most frequently reported serogroups were O157 (21.3%), O26 (19.4%), O103 (6.6%), O146 (5.5%), O145 (4.4%), and O91 (2.9%) [1].

Since 2007, ECDC has been responsible for the EU-wide surveillance of STEC, including facilitating the detection and investigation of food-borne outbreaks. Surveillance data, including basic typing parameters and molecular typing data for the isolated pathogen, are reported by Member States to The European Surveillance System (TESSy). The surveillance system relies on the capacity of NPHRLs in FWD-Net providing data to produce comparable typing results. To ensure that the EQA is linked to the development of surveillance methods used by NPHRLs, a molecular typing-based cluster analysis using whole genome sequencing (WGS)-derived data has been included since EQA-8.

The objectives of the EQAs are 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 currently relevant to public health in Europe and represent a broad range of clinically relevant types of STEC. Twelve test strains were selected for serotyping/virulence profile determination and molecular typing-based cluster analyses. Additional eight strains (sequences) were included for the molecular typing-based cluster analysis. Twenty-six laboratories registered and 26 completed the exercise, comparable to EQA-11.

The full O:H serotyping was performed by 85% (22/26) of participating laboratories, with an average score of 97%. In general, the more common European serotypes generated the highest scores compared to the less common serotypes, such as O8:H4, O80:H2, and O154:H31, which proved more difficult to identify, particularly if participants used phenotypic methods. Notably, not all laboratories demonstrated the capacity to determine all O groups and H types, and the participation in H typing was lower (22/26) compared to the O grouping (25/26) but higher than H typing in EQA-11 (19/26) most likely reflecting a shift towards WGS-based methods. The shift towards WGS was also seen in the reported O-grouping results: 68% (17/25) used WGS-based methods, which is higher than EQA-11 (60%), EQA-10 (52%), EQA-9 (50%), and EQA-8 (26%).

The quality of the virulence profile determination results was generally good, with high average scores of 96%, 97%, and 98% for *eae, stx1*, and *stx2*, respectively, similar to previous EQAs.

In EQA-12, two other diarrhoeagenic *E. coli* (DEC) pathotypes were included, EAEC strain3 (*aggR* gene) and ETEC Strain11 (*esta* gene) testing the participating laboratories in their abilities to detect STEC hybrid strains. The detection performance of the *aggR* gene was higher (23/24, 98%) than in both EQA-11 (95%) and EQA-10 (94%). Similar to *aggR*, the performance for *esta* was also higher (98%) than EQA-11 (89%). This variance in performance was attributed to four laboratories that couldn't identify the gene in strain11. All laboratories, except one, utilised a WGS-based method.

Of the 26 laboratories participating in the EQA-12, 23 (88%) performed molecular typing-based cluster analysis using WGS data analysed by different approaches. Notably, all laboratories used WGS in both EQA-12 and EQA-11 and none chose PFGE, a decrease from EQA-10 (2 laboratories) and EQA-9 (8 laboratories). The purpose of the cluster analysis part of the EQA was to assess the NPHRL's ability to identify a cluster of genetically closely related strains, i.e. to correctly categorise the cluster test strains regardless of the method used. The focus is on the result, not a specific procedure.

Fifteen participants (65%) correctly identified the cluster of closely related ST301 strains defined by pre-categorisation from the EQA provider among the 12 test strains and eight test strains (genomic sequences).

In this EQA, participants were free to choose their preferred analytical method for the WGS-based cluster identification. An allele-based method was most frequently used; 87% (20/23) used core genome MultiLlocus Sequence Type (cgMLST) compared to 13% (3/23) using single nucleotide polymorphism (SNP) for the reported cluster analysis as the main analysis.

In general, for cgMLST the reported results from the participants were at a comparable level despite using various analysis and different allelic calling methods.

For inter-laboratory comparability and communication about cluster definitions, cgMLST using a standard scheme (e.g. Enterobase) gives a very high degree of homogeneity in the results, while the use of non-standardised SNP analysis may be more challenging. There are two main challenges: difficulty in comparing SNP with cgMLST results, and variations between SNP analyses in general, as demonstrated in this EQA, which makes the comparison and communication of the results between laboratories difficult. The latter was reflected in the reported results, as all three of the laboratories that used SNP-based analysis did not identify the pre-determined cluster.

The participants assessed additional genomes, some of which were modified by the EQA provider to provide a realistic view of various quality issues. Notably, only 48% (11/23) of the participants reported quality issues with the modified sequence containing 8% contamination with *Shigella sonnei*. In contrast, all participants (100%) correctly identified the poor quality of strain18, a non-cluster sequence with reduced coverage and removal of genes. Assessing both contamination with a different species and poor quality is crucial before conducting WGS analysis.

A feedback survey was sent to assess the STEC EQA scheme. The questionnaire contained both questions related to accreditation and information on the individual report; 58% (15/26) responded. Overall, the survey revealed an appreciation for QC assessment but highlighted the need for the EQA provider to optimise analyses also to IonTorrent data in addition to the standardised Illumina data. Streamlining the reporting form, especially for virulence gene determination, was suggested, along with exploring sending isolates for multiple EQAs simultaneously during sequencing runs. All of the responders appreciated the format, but some listed recommendation for improvements.

1 Introduction

1.1 Background

ECDC is an EU agency with a mission 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 the diagnosis, detection, identification, and characterisation of infectious agents that may threaten public health. ECDC maintains and extend such cooperation and support the implementation of quality assurance schemes [2].

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

ECDC has outsourced the organisation of EQA schemes for EU/EEA countries in the disease networks. EQAs aim to identify areas for improvement in laboratory diagnostic capacities relevant for epidemiological surveillance of communicable diseases as set forth in Decision No 1082/2013/EU [3] and ensure the reliability and comparability of results generated by laboratories across all EU/EEA countries.

The main purposes of EQA schemes are to:

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

Since 2012, the unit of Foodborne Infections at SSI, Denmark, has been the EQA provider for the three EQA schemes covering typing of *Salmonella enterica* ssp. *enterica*, Shiga toxin/verocytotoxin-producing *Escherichia coli* (STEC/VTEC) and *L. monocytogenes*. In 2021, SSI was granted the new round of tenders (2022–2025) for *Listeria* and STEC. The STEC EQA covers serotyping, virulence profile determination, and molecular typing-based cluster analysis. This report presents the results of the STEC EQA-12.

1.2 Surveillance of STEC infections

STEC is a group of *E. coli* characterised by the ability to produce Shiga toxins (Stxs). Human pathogenic STEC often harbour additional virulence factors important to the pathogenesis of the disease. A large number of serotypes of *E. coli* have been recognised as Stx producers. Notably, the majority of reported human STEC infections are sporadic cases. Symptoms associated with STEC infection in humans vary from mild diarrhoea to life-threatening haemolytic uraemic syndrome (HUS), which is clinically defined as a combination of haemolytic anaemia, thrombocytopenia and acute renal failure.

In 2022, 8 565 confirmed cases of STEC infection were reported by 29 EU/EEA countries. The overall EU/EEA notification rate was 2.5 cases per 100 000 population, which exceeded the pre-pandemic level and represented a 25% increase compared to the notification rate in 2021 [1]. In 2022, the six most frequently reported serogroups were O157 (21.3%), O26 (19.4%), O103 (6.6%), O146 (5.5%), O145 (4.4%), and O91 (2.9%). These serogroups together accounted for over 60% of the total number of confirmed STEC cases with known serogroups in 2022. For 2 414 cases (28.2%) the full serotype was reported, i.e. both the O type and the H type. The most common serotype was O157:H7 (18.4%), followed by O26:H11 (18.1%) and O103:H2 (7.6%). For cases with STEC-associated HUS, serogroup was reported for 393 cases; O26 was most frequently reported (51.4%), followed by O157 (14.5%), O80 (6.0%), and O145 (5.8%).

One of ECDC's key objectives is to improve and harmonise the surveillance system in the EU/EEA to increase scientific knowledge of aetiology, risk factors, and burden of FWDs and zoonoses. Surveillance data, including some basic typing parameters for the isolated pathogen, are reported by TESSy. In addition to the basic characterisation of the pathogens isolated from human infections, there is public health value in using more discriminatory typing techniques for pathogen characterisation in the surveillance of food-borne infections. Since 2012, ECDC has enhanced EU surveillance by incorporating molecular typing data through isolate-based reporting. Three selected FWD pathogens were included: *Salmonella enterica* ssp. *enterica*, *L. monocytogenes*, and STEC. The overall aims of integrating molecular typing into EU level surveillance are to:

- foster the rapid detection of dispersed international clusters/outbreaks;
- facilitate the detection and investigation of transmission chains and relatedness of isolates across Member States and contribution to global investigations;

- detect the emergence of new evolving pathogenic isolates;
- support investigations to trace the source of an outbreak and identify new risk factors; and
- aid the study of a particular pathogen's characteristics and behaviour in a community of hosts.

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

1.3 STEC characterisation

State-of-the-art characterisation of STEC includes O:H serotyping in combination with a few selected virulence genes, i.e. the two genes for production of Shiga toxin Stx1 (*stx1*) and Stx2 (*stx2*) and the intimin *(eae)* gene associated with attaching and effacing lesion of enterocytes, also seen in attaching and effacing non-STEC *E. coli* (AEEC), including enteropathogenic *E. coli* (EPEC). The combination of virulence genes and subtypes of toxin genes is clinically relevant. The *stx2a* in *eae*-positive STEC and the activatable[i] [4] *stx2d* subtype in *eae*-negative STEC appear to be highly associated with the serious sequela HUS [4–7]. In the recent Scientific Opinion by the European Food Safety Authority (EFSA), analysis of the confirmed reported human STEC infections in the EU/EEA (2012–2017) reveals that all Stx toxin subtypes may be associated with some cases of severe illness defined as bloody diarrhoea, HUS and/or hospitalisation [8]. Understanding the epidemiology of the *stx* subtypes is therefore important to prevent the risk of STEC infection and for the surveillance of STEC.

The recommended method for *stx* subtyping is a specific polymerase chain reaction (PCR) [4]. STEC serotype O157:H7 may be divided into two groups: non-sorbitol fermenters (NSF) and a highly virulent sorbitol fermenting (SF) variant of O157. STEC EQA-12 included O:H serotyping, detection of virulence genes (*eae, stx1* and *stx2*, including subtyping of *stx* genes), the *aggR* gene specific for enteroaggregative *E. coli* (EAEC), the *esta* gene specific for enterotoxigenic *E. coli* (ETEC), and molecular typing-based cluster analysis.

Notably, hybrid *E. coli* pathotypes represents an emerging public health threat with enhanced virulence from different pathotypes, where O104:H4 EAEC-STEC is well known. Hybrids of other STECs include enterotoxigenic *E. coli* (STEC/ETEC) and extraintestinal pathogenic *E. coli* (STEC/ExPEC) which have both been reported to be associated with diarrheal disease and HUS in humans.

1.4 Objectives of the EQA-12 on STEC

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

As a result, and part of the recommendations in EQA-10, the EQA provider does not include *aaiC* gene in EQA-12. This is based on the newest published recommendation defining enteroaggregative *E. coli* (EAEC) strains as harbouring *aggR* and a complete cluster of AAF-encoding genes (usher, chaperone, and both major and minor pilin subunit genes) or the enterotoxigenic *E. coli* (ETEC) colonisation factor (CF) CS22 gene [9].

1.4.1 Serotyping

The objectives of STEC serotyping in EQA-12 were to assess the ability to assign correct O groups and H types by using either serological (detection of somatic 'O' and flagellar 'H' antigens) or molecular typing methods (PCR or WGS).

1.4.2 Virulence profile determination

The objectives of the virulence gene determination of STEC EQA-12 were to assess the ability to assign the correct virulence profile; the presence/absence of *stx1*, *stx2*, *eae, esta,* and *aggR* genes and subtyping of *stx* genes (*stx1a, stx1c, stx1d, stx2a, stx2b, stx2c, stx2d, and stx2e*).

1.4.3 Molecular typing-based cluster analysis

The objective of the molecular typing-based cluster analysis of STEC EQA-12 was to assess the ability of the participants to correctly identify the cluster of closely related strains. Laboratories could perform analysis using PFGE and/or derived data from WGS. The cluster analysis should be conducted on the 12 test strains and eight additional test strains (provided genomic sequences). Some of the provided sequences were modified to have quality control (QC) issues.

[i] Activated by mucus containing elastase which increase the cytotoxicity [4].

2 Study design

2.1 Organisation

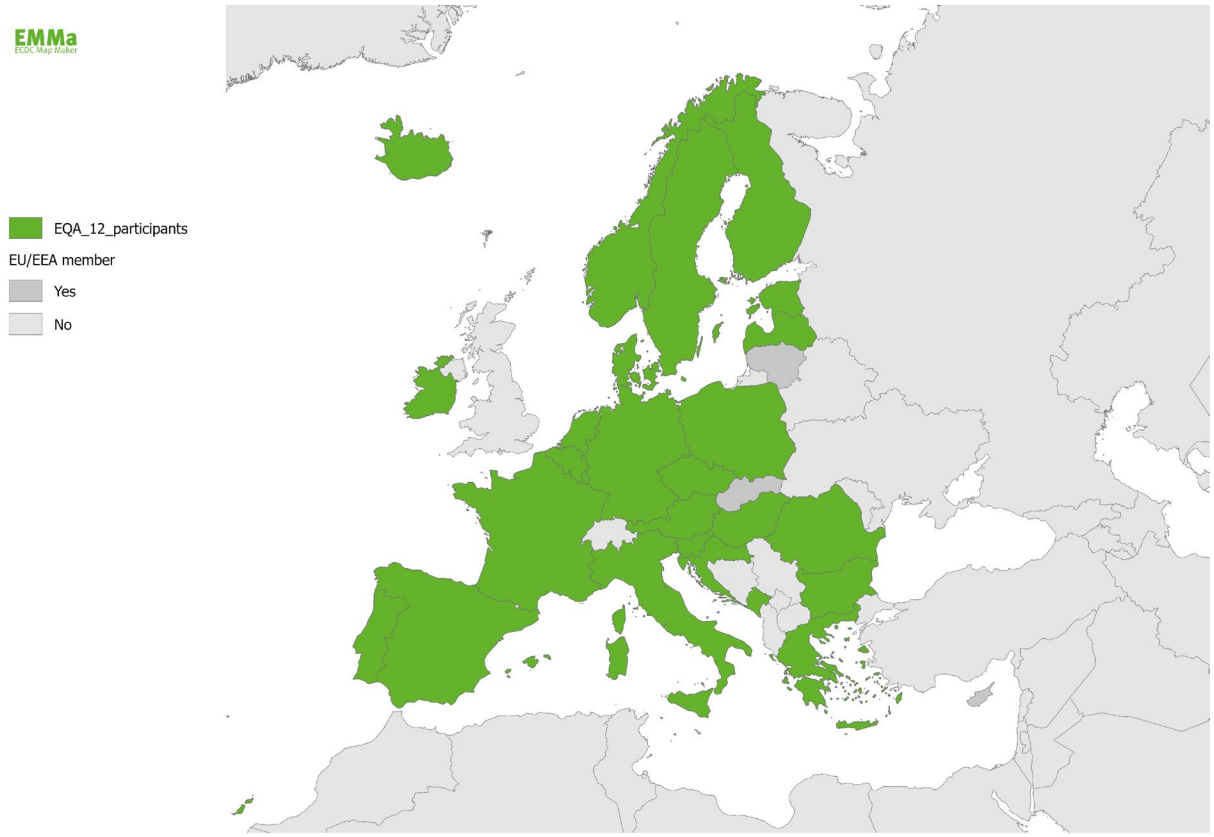
STEC EQA-12 was funded by ECDC and arranged by SSI following ISO/IEC 17043:2010 [10]. EQA-12 included serotyping, virulence gene determination, and a molecular typing-based cluster analysis, and was carried out between May and December 2023.

Invitations were emailed by the EQA provider to the ECDC's contact points in the FWD-Net (30 countries) by 1 May 2023, with a deadline to respond by 22nd of May 2023. In addition, invitations were sent to the EU candidate countries.

Twenty-six NPHRLs in EU/EEA and EU candidate countries accepted the invitation to participate, and all 26 submitted their results (Figure 1, Annex 1). EQA test strains were sent to participants from 21 June to 28 June 2022. In Annex 2, participation details in EQA-11 and EQA-12 are listed to give an overview of the trend in the number of participants. Participants were asked to submit their raw reads (FASTQ files) to a secure file transfer protocol (SFTP) -site and complete the online form for results by 15 October 2023 (Annex 12).

The EQA submission protocol, invitation letter, and a blank submission form were available online.

Figure 1. Countries participating in the 12th round of the external quality assessment (EQA-12) scheme for typing of Shiga toxin-producing *Escherichia coli* (STEC)



EQA-12 participating countries are shown in green. Administrative boundaries: © EuroGeographics © UN-FAO © Turkstat. The boundaries and names shown on this map do not imply official endorsement or acceptance by the European Union.

2.2 Selection of test strains/genomes

Seventeen test strains were selected to fulfil the following criteria:

- represent commonly reported strains in Europe;
- remain stable during the preliminary test period at the organising laboratory;
- include same serotypes as in the previous years;
- include a set of technical duplicates in the serotyping/grouping/cluster; and
- include genetically closely related strains.

The 14 selected strains were analysed with the methods used in the EQA (serotyping and virulence profile determination or WGS) before and after having been re-cultured 10 times. All candidate strains remained stable using these methods and the final test strains and additional sequences were selected. The selected 12 test strains (Table 1) for serotyping/detection of virulence gene were selected to cover different serotypes and *stx* subtypes relevant for the current epidemiological situation in Europe (Annexes 3-4).

Similarly to EQA-11, we included three hybrid *E. coli* pathotype test strains; Shiga toxin-producing and enterotoxigenic *E. coli* (STEC/ETEC), Shiga toxin-producing and enteroaggregative *E. coli* (STEC/EAEC), and extraintestinal pathogenic *E. coli* (STEC/ExPEC). As was seen with the emergence of Shiga Toxin producing enteroaggregative *E. coli* (Stx-EAEC), hybrid strains can possess a major challenge for the public health, due to the needs to now implement diagnostic procedures that will identify the most virulent clones. The selected hybrid strains comprised of O8:H4 (STEC/ETEC), O111:H21 (STEC/EAEC), and O80:H2 (STEC/ExPEC). It is noteworthy, that the latter strain was specifically chosen for the ST301 cluster analysis in this EQA.

Based on the WGS-derived data, the selected cluster of closely-related strains consisted of five STEC ST301 strains (including the technical duplicate set strain4/strain10). Characteristics of all the STEC test strains are listed in Table 1 and Annexes 3-9. The EQA provider found at most four allele differences or 0 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 [11]) and SNP analysis (NASP [12]). The participants using PFGE as a cluster method could only evaluate the 12 test strains from the package and only two belonged to the cluster of closely related strains based on WGS. The cluster categorisation is based on WGS data and the correct cluster delineation might be difficult to obtain by the use of less discriminatory methods, e.g. PFGE. An additional eight strains (sequences) for cluster analysis were selected to include strains with different varying relatedness of sequence types (ST301) and other STs. A set of duplicates were included in the test strains (strain4 and strain10) (Annexes 5–7, 9–10). Three of the sequences were modified by the EQA provider. Further, one sequences with reduced coverage, one sequence with low quality of R2, and one contaminated with 8% *S. sonnei.* The characteristics of all the strains and sequences are listed as 'EQA provider' in Annexes 4–10.

Table 1. Characteristics of test strains and sequences

Method	**Serotyping**		**Virulence profile**		**Cluster analysis**			
No. strains/sequences	12 strains		12 strains		12 strains / 8 sequences			
Annex	3		4		5–6, 7-9			
Strain ID						ST	QC-status	Cluster
Strain1	Strains for Serotyping	O146:H28	Strains for virulence profile	*stx2b*	Strains/sequences for cluster analysis	738	-	
Strain2		O103:H2		*stx1a, stx2d, eae*		17	-	
Strain3		O111:H21		*stx2a, aggR*		40	-	
Strain4#‡		O80:H2		*stx2a, eae*		301	-	Yes
Strain5		O145:H-/H28		*stx2a, eae*		32	-	
Strain6		O157:H-/H7		*stx2c, eae*		11	-	
Strain7		O128:H2		*stx1c, stx2b*		4 748	-	
Strain8		O154:H31		*stx1d*		1 892	-	
Strain9		O26:H11		*stx1a, stx2a, eae*		21	-	
Strain10#‡		O80:H2		*stx2a, eae*		301	-	Yes
Strain11		O8:H4		*stx2e, esta*		88	-	
Strain12		O80:H2		*stx2a, eae*		301	-	
Strain13 -sequence	-	O80:H2		*stx2a, eae*		301	A	
Strain14‡ -sequence	-	O80:H2		*stx2a, eae*		301	A	Yes
Strain15‡ -sequence	-	O80:H2		*stx2a, eae*		301	A	Yes
Strain16 - sequence	-	O80:H2		*stx2a, eae*		301	A	
Strain17^ -sequence	-	O80:H2		*stx2a, eae*		301	B	
Strain18^ - sequence	-	-		-		-	C	
Strain19‡ - sequence	-	O80:H2		*stx2a, eae*		301	A	Yes
Strain20‡^ - sequence	-	O80:H2		*stx2a, eae*		301	B	Yes

‡: closely related strains; #: technical duplicates strains; ST: sequence type; ^modified sequences: strain17, a nonCluster sequence with low quality of R2, strain18, A nonCluster sequence with low coverage and, strain20, a cluster sequence (strain14) contaminated with approx. 8% S. sonnei; A: Acceptable quality, B: Quality only acceptable for outbreak situations (less good quality) and C: Not acceptable quality – strain not analysed.

2.3 Distribution of strains and sequences

The 12 test strains were blinded and shipped from 21–28 June 2023 as UN2814. Letters stating the unique strain IDs were included in the packages and distributed individually to the participants by email on the day of shipment as an extra precaution. Fifteen participants received the strains within two days, eight within five and six days and four within seven to nine days after shipment, respectively. No participants reported damage to the shipment or errors in the unique strain IDs.

In July 2023, instructions for the submission of results procedure were emailed to the participants. This included the links to the online site for downloading the additional sequences, viewing the empty submission form and uploading the produced FASTQ files.

2.4 Testing

The serotyping part comprised 12 STEC test strains and the purpose was to assess the participants' ability to obtain the correct serotype. The participants could perform conventional serological methods according to suggested protocol [113] or molecular-based serotyping (PCR or WGS). The results of serotyping were submitted in the online form.

The same set of the above 12 STEC test strains were also used to generate the virulence profile. The analyses were designed to assess the participants' ability to obtain the correct virulence profile. The participants could choose to perform detection of *the aggR* (EAEC associated gene), *esta* (ETEC associated gene) *eae* and *stx1* and *stx2,* as well as subtyping of subtyping of *stx* genes (*stx1a, stx1c, stx1d, stx2a, stx2b, stx2c, stx2d, and stx2e*) according to suggested protocol [14, 15]. The results were submitted in the online form.

For the molecular typing-based cluster analysis the participants could choose to use either WGS-derived data or PFGE-derived data. In this EQA-12, all the participants for the first time only chose WGS-derived data. 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 Multi Locus Sequence Typing (wgMLST/cgMLST) (allele-based) and were asked to submit the strains identified as a cluster of closely related strains based on the analysis used. Laboratories could report results from up to three analyses (one main and up to two additional), but the detected cluster were required to be based on results from the main analysis. The laboratories reported SNP distance or allelic differences between each test strain and a strain (strain15) selected by the EQA provider.

In addition, each participant needed to assess the QC of the provided sequences (three 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: Not acceptable 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 three sequences (strain17, strain18 and strain20). Table 5, Annex 11.

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

2.5 Data analysis

The submitted serotype, virulence profile, and cluster analysis results, as well as the raw reads, were imported to a dedicated STEC EQA-12 BioNumerics (BN) database. The EQA provider contacted eight participants in order to ensure sequences were uploaded to the SFTP site. Two additional laboratories were contacted as one or more of the sequences was uploaded with incomplete data due to upload errors or wrong filenames. Another laboratory zipped the sequences twice.

Serotyping results were evaluated according to the percentage of correct results, generating a score from 0–100% for O group, H type and O:H serotype.

The virulence profile determination results were evaluated according to the percentage of correct results, generating a score from 0–100% for *eae*, *aggR*, *esta*, *stx1*, *stx2*, subtyping of *stx1* and *stx2* and combined subtype (Table 1).

Molecular typing-based cluster analysis was evaluated according to correct or incorrect identification of the expected cluster of closely related strains based on a pre-defined categorisation by the organiser. The EQA provider's WGS-derived cluster analysis was based on allele-based cgMLST [11] and SNP analysis (NASP) [12]. The cluster categorisation is based on WGS data and the correct cluster delineation might be difficult to obtain by the use of less discriminatory methods, e.g. PFGE. The ST301 cluster comprised six strains or sequences: strain4, strain10, strain14, strain15, strain19, and strain20, with duplicates represented by strain4 and strain10. Notably, the sequence of strain20 originated from strain14 but was modified by contaminating the sequence with 8% *S. sonnei* by the EQA-provider. Consequently, laboratories were given the flexibility to either include or exclude

strain20 in the ST301 cluster if they identified the contamination. The EQA provider determined that there were, at most, four allele differences or 0 SNPs between any two strains within the cluster.

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

Individual evaluation reports and certificates of attendance were distributed to participants in December 2023. If WGS data were used, the evaluation report included a quality assessment made by the EQA provider's in-house quality control pipeline (e.g. coverage, N50, sequence length, and number of contigs). The QC-status of the submitted sequences were commented in the evaluation report.

3 Results

3.1 Participation

Laboratories could either participate in the full EQA scheme or one part only (serotyping, virulence profile determination or molecular typing-based cluster analysis). Of the 26 participants, who signed up, 26 completed and submitted their results. Eighty-five percent of the participants (22/26) completed all three parts of the EQA-12 (serotyping, virulence determination, and cluster analysis). In total, 25 (96%) of the participants performed serotyping, 25 (96%) participated in the detection of one or more of the virulence genes and 23 (88%) in cluster analysis. (Table 2).

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

	Serotyping[1]	Virulence profile determination[2]	Cluster analysis[3]
Number of participants	25	25	23
% of participants	96*	96*	88*

[1]: O grouping and/or H typing
[2]: detection of at least one gene (aggR, eae, esta, stx1 and stx2) and/or subtyping of stx1 and stx2
[3]: molecular typing-based cluster analyses based on WGS-derived data
**: percentage of the total number (26) of participating laboratories.*

O grouping results were provided by 25 participants (96%) and H typing results were provided by 22 (85%). Almost two-thirds, 16/25 (64%), used molecular-based serotyping, two reported PCR-based method (8%), and seven performed phenotypic serotyping (28%). (Annex 3). The majority of the participants (96%, 25/26) performed the detection of virulence genes *stx1*, *stx2, eae,* and the detection of the enteroaggregative gene, *aggR.* Slightly fewer participants reported the heat stable (ST) enterotoxin gene, *esta* (92%, 24/26). Additionally, the *stx*1 and *stx2* subtyping detection were reported by 92% (24/26) (Annex 4). The majority of the participants performed the cluster analyses (88%, 23/26), all used WGS-derived data (Table 3).

Table 3. Detailed participation information for the parts of serotyping, virulence profile determination and molecular typing-based cluster analysis

	Serotyping n=25		Virulence profile determination n=25					Cluster analysis n=23
	O group	H type	*aggR*	*eae*	*esta*	*stx1* and *stx2*	*stx* subtyping	WGS
Number of participants	25[#]	22Δ	25	25	24	25	24	23
Percentage of participants^	100%	88%	100%	100%	96%	100%	96%	100%
Percentage of participants *	96%	85%	96%	96%	92%	96%	92%	88%

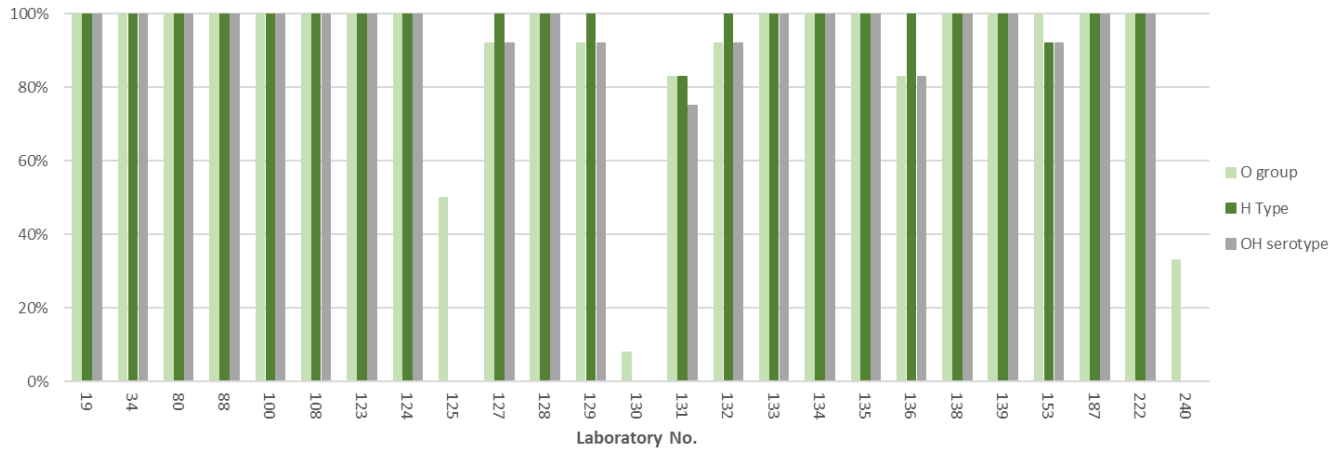
^: percentage of participants in respective part of EQA
**: percentage of total number of participating laboratories (26)*
[#]: phenotypic (n=7)/PCR-based (n=2)/WGS-based (n=16)
Δ: phenotypic (n=3)/PCR-based (n=0)/WGS-based (n=20)

3.2 Serotyping

Twenty-five (96%) laboratories performed O grouping and 17 (68%) of the 25 participants were able to correctly O-type all 12 test strains, and three laboratories had a score of $\leq$ 50%, giving an average score of 89% (Figure 2). Twenty-three laboratories (92%) reported the correct O group for the Hybrid strains O111 (strain3) and 20 (80%) correctly reported O8 (Strain11) (Figure 3). The highest performances were obtained for the O157 (100%), O145 (96%), and O26 (96%) positive strains (Figure 3).

Twenty-two (85%) laboratories performed H typing. Of the 25 laboratories participating in O grouping, 88% (22/25) also reported H type. The general performance for H typing was higher than O grouping, with the majority (91%; 20/22) of participants correctly H typing all 12 test strains, resulting in an average score of 99% (Figure 2). In two out of the 12 strains reporting H- was accepted as a correct result when using phenotypical H-typing as these strains were non-motile (strain5 and strain6). One laboratory (153) reported H9 instead of H28, and laboratory 131 reported two incorrect H types (H- instead of H31 and H14 instead of H2) (Annex 3).

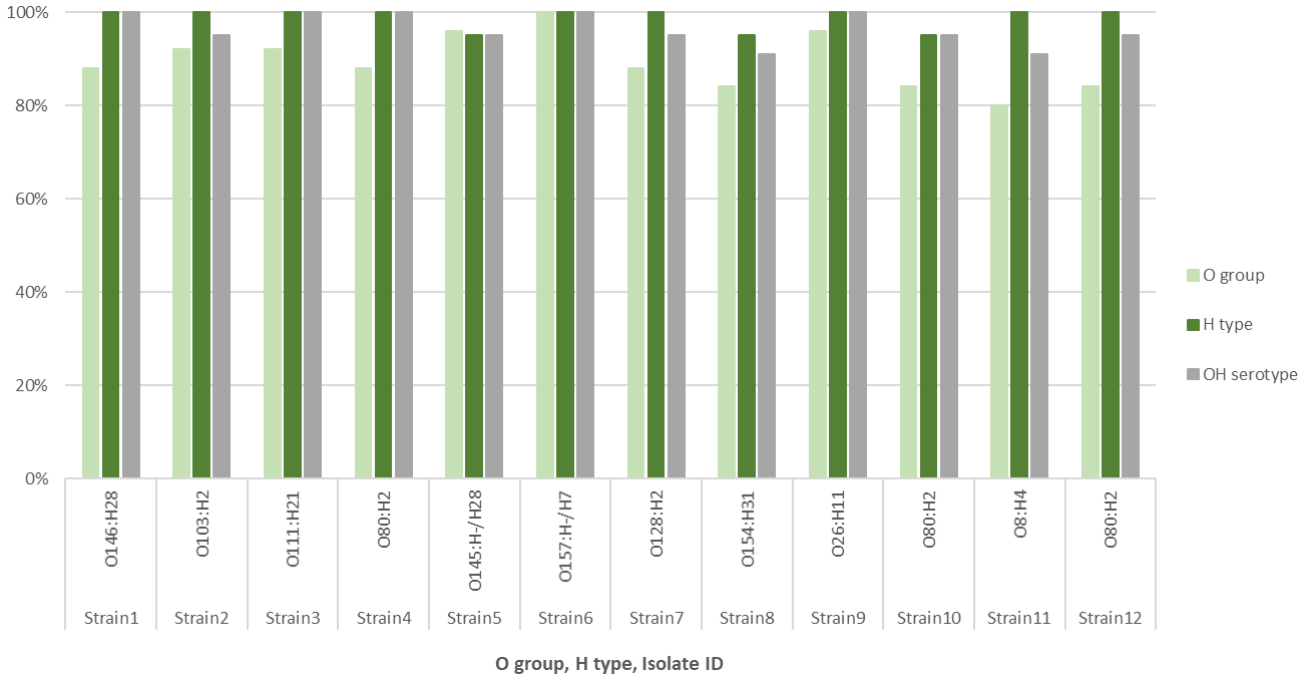
Figure 2. Participant percentage scores for O grouping and H typing



Arbitrary numbers represent participating laboratories.
Bars represent the percentage of correctly assigning O groups (light green), n= 25 participants, H types (dark green), n=22 participants, Combined O:H serotypes (grey), n=22 participants.

Complete O:H serotyping was performed by 22 out of the 25 (88%) participants with an average score of 97%, and for each strain the score ranged from 91% (20/22) for Strain8 (O154:H31) and strain 11 (O8:H4) to 100% (22/22) for Strain1 (O146:H28), Strain3 (O111:H21), Strain4 (O80:H2), Strain6 (O157:H-/H7), and Strain9 (O26:H11). The correct serotype of all 12 strains were reported by 73% (16/22) of the participants who performed the O:H serotyping (Figure 3, Annex 3).

Figure 3. Average percentage test strain score for serotyping of O and H



Bars represent the percentage of laboratories correctly assigning O groups (light green): n=25 participants.
H types (dark green): n=22 participants. Combined O:H serotypes (grey): n=22 participants.
Average scores: O group, 89%; H type, 99% and combined O:H serotype, 97%.

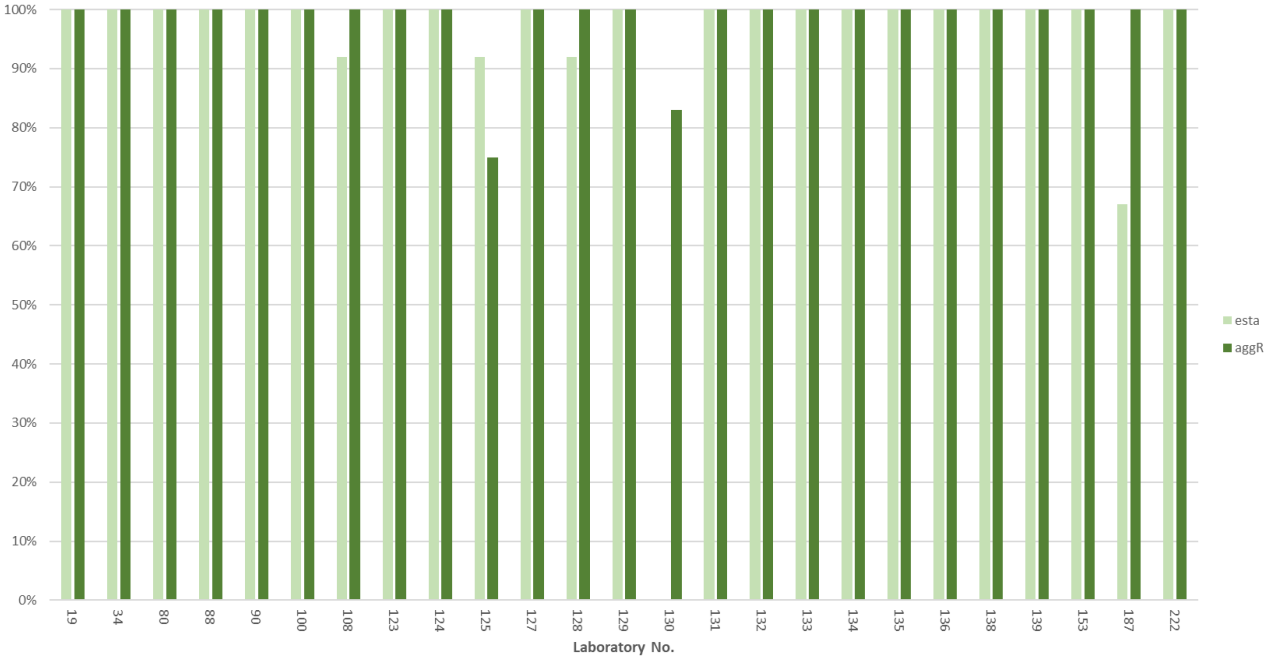
3.3 Virulence profile determination

Between 23 and 26 laboratories submitted results for some, or all, of the following virulence genes; *aggR* (25 participants), *eae* (25 participants), *esta* (24 participants), *stx1* (25 participants), *stx2* (25 participants), and subtyping of *stx1* (24 participants), and *stx2* (24 participants).

3.3.1 Detection of the EAEC and ETEC genes (*aggR* and *esta*)

Among the strains in EQA-12 two test strains harboured other pathotype-defining virulence genes; strain11 harbouring the ETEC associated *esta* gene and strain3 harbouring the EAEC defining gene *aggR*. All laboratories, except for two (125 and 130), correctly identified *aggR* in strain3, as such the performance for *aggR* was 98% (23/25). The performance for *esta* was 98% corresponding to four laboratories (108, 125, 128, and 187) that couldn't identify the gene in strain11 (Figure 4, Annex 4).

Figure 4. Participant percentage scores for genotyping of *aggR* and *esta*

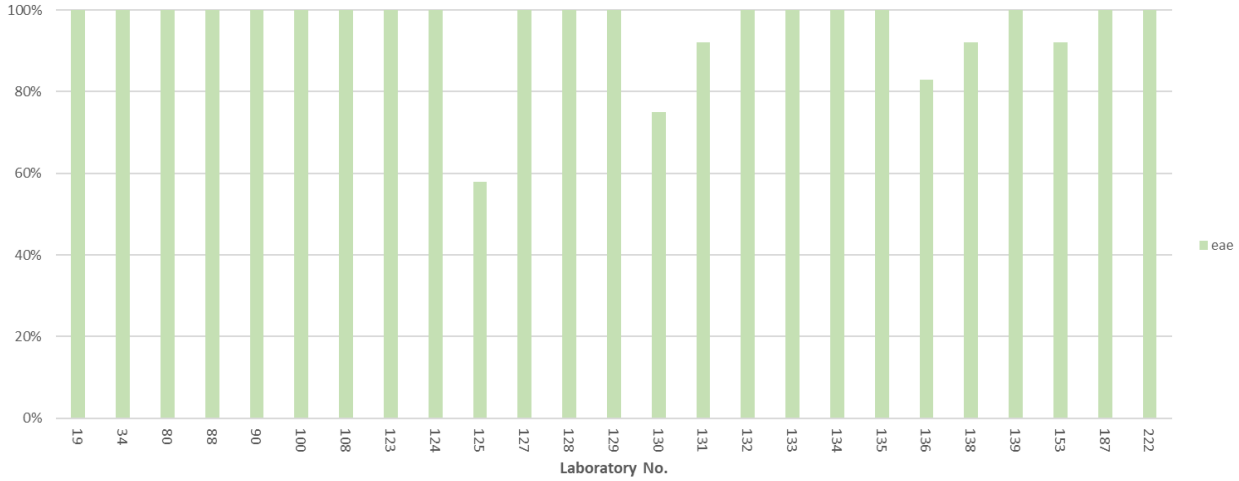


Arbitrary numbers represent participating laboratories.
Bars represent percentage of correct genotyping of esta (light green) n=24 participants and aggR (dark green): n=25 participants. Average scores: esta, 98%; aggR, 98%.

3.3.2 Detection of virulence genes *eae, stx1* and *stx2*

Detection of virulence genes *eae*, *stx1* and *stx2* was performed by 25 (96%) laboratories with a generally high performance (Figures 5–6). For *eae* detection, 19 (76%) laboratories obtained a 100% score and six laboratories (125, 130, 131, 136, 138, and 153) reported incorrect results for the *eae* gene (Figure 5). Eleven of the 13 incorrect results were false negative and eight of these were found by laboratories 125 and 130 in multiple strains. Laboratories 136 and 138 reported a false positive in strain7 and strain1, respectively (Annex 4).

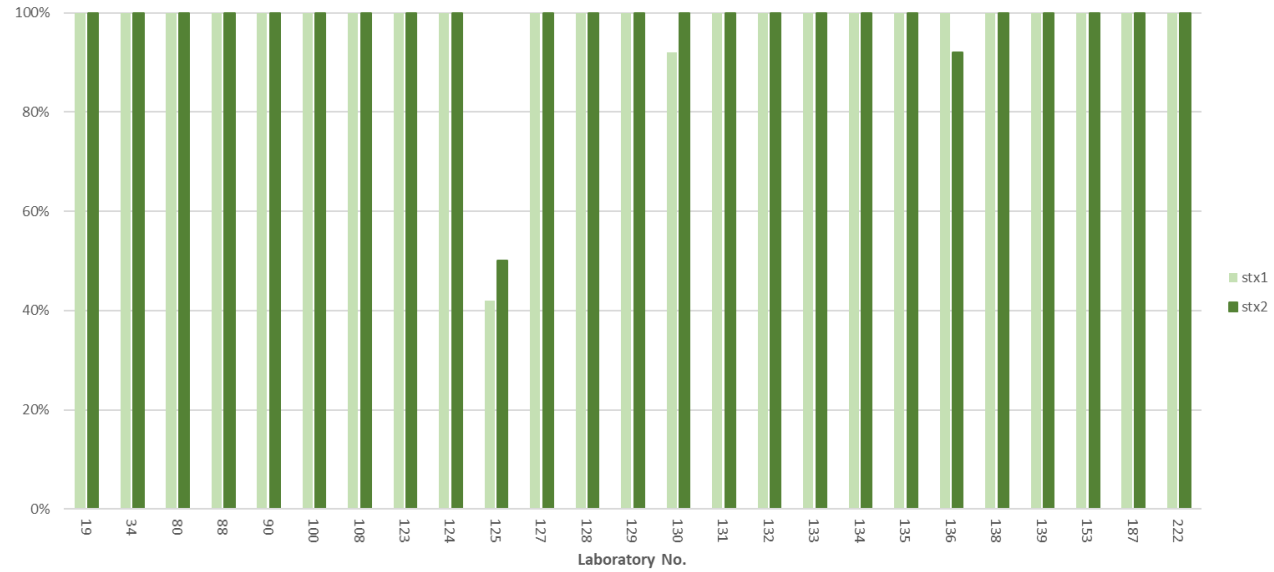
Figure 5. Participant percentage scores for genotyping of *eae*



Arbitrary numbers represent participating laboratories.
Bars represent percentage of correct genotyping of eae (light green): n=25 participants. Average score: eae, 96%.

The performance for the detection of both *stx1* and *stx2* was high, and 23 laboratories reported 100% accuracy for both *stx1* and *stx2* (Figure 6). There were eight incorrect results for stx1, with seven reported by laboratory 125 and one by laboratory 130. Laboratory 125 reported six false positives and one false negative, while laboratory 130 reported one false negative. For *stx2*, there were seven incorrect results reported, and six of those were false negatives. Primarily, laboratory 125 reported six of the incorrect *stx2* results. One laboratory (136), reported a false negative result for strain 6. The one false positive result was reported by laboratory 125 in strain 8 (Annex 4).

Figure 6. Participant percentage scores for detection of *stx1* and *stx2*



Arbitrary numbers represent participating laboratories.
Bars represent percentage of correct genotyping of stx1 *(light green) and* stx2 *(dark green): n=25 participants.*
Average scores: stx1, 97%; stx2, 98%.

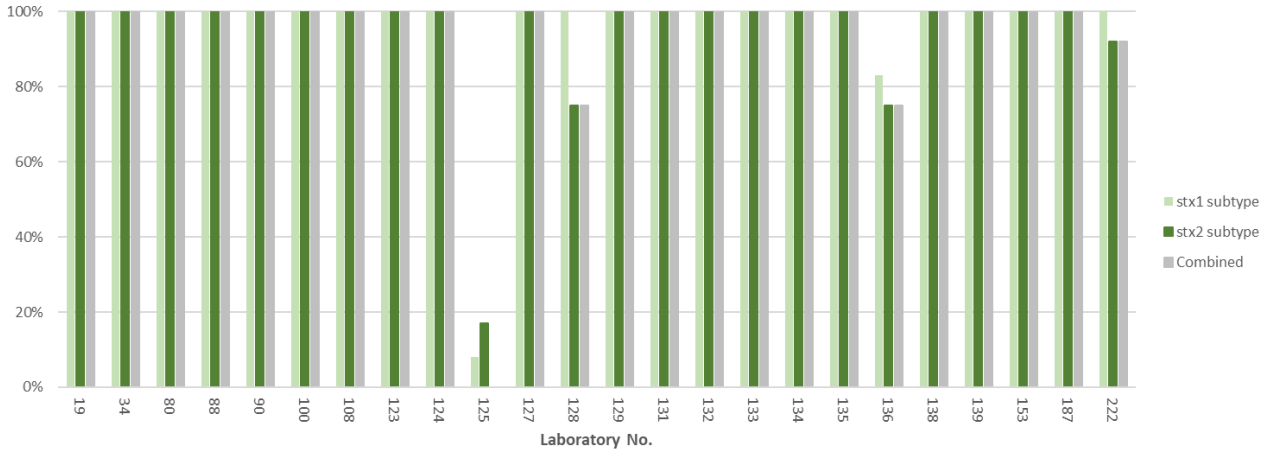
3.3.3 Subtyping of *stx1* and *stx2*

Subtyping of *stx1* and *stx2* was performed by 24 laboratories. For all 12 test strains, 22 laboratories subtyped *stx1* correctly (92%; 22/24) and 20 laboratories correctly subtyped *stx2* (83%; 20/24) (Figure 7; Annex 4). The combined *stx* subtyping were reported correctly by 83% of the laboratories (20/24).

Laboratories were not allowed to only report results for selected test strains for a particular test, so reporting ND was considered as an incorrect result if the laboratory reported results of other strains for that test.

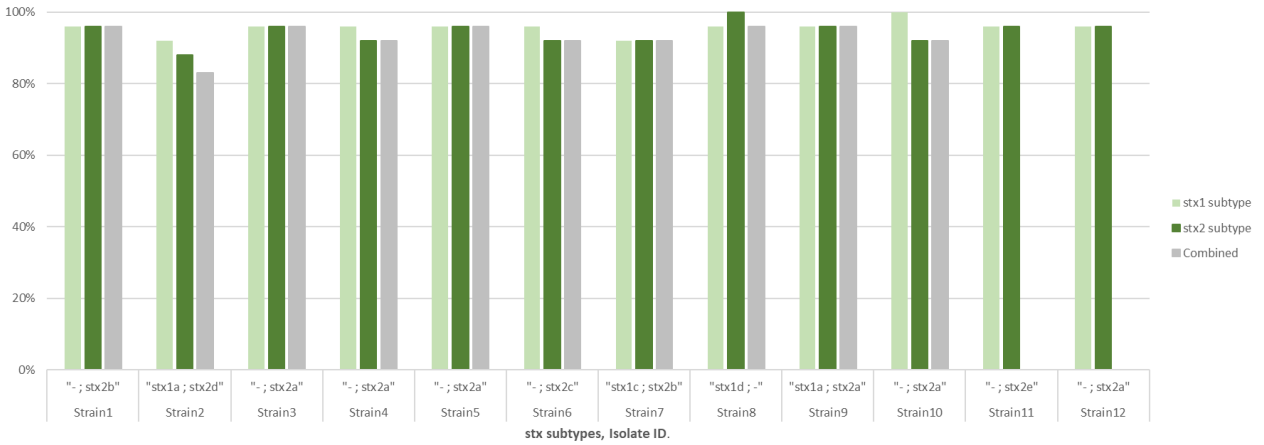
Only four laboratories (18%) reported an incorrect subtyping of either/or both *stx1* and *stx2*. For *stx1* subtyping, one laboratory (136) incorrectly reported *stx1c* instead of *stx1a* for strain2 and vice versa for strain7. Notably, Laboratory only reported *stx1* subtyping for Strain10. As such, the average score of the 12 test strains were 95% for the *stx1* subtyping. The average score for stx2 subtyping was 94%, with the majority of the mis-subtyped stx2 results attributed to three laboratories: 125, 128, and 136, corresponding to 17%, and the latter two achieving 75% correct *stx2* subtyping. The average score of the combined subtyping was 93%.

Figure 7. Participant percentage scores for subtyping of *stx1* and *stx2*



Arbitrary numbers represent participating laboratories.
Bars represent percentage of correct subtyping of stx1 (light green), stx2 (dark green), combined stx1 and stx2 (grey), n=24 participants. Reporting ND (not done) evaluated as incorrect.

Figure 8. Average percentage test strain score for subtyping of *stx1* and *stx2*



Bars represent percentage of laboratories correctly subtyping stx1 (light green), stx2 (dark green) and combined stx1 and stx2 (grey), n=24. Average scores: stx1, 95%; stx2, 94% and combined stx1 and stx2, 93%.

One laboratory (125) reported a 'ND' (not done) result, as such most incorrect results are no longer due to reporting ND instead of negative result, as in EQA-8.

The incorrect results of the *stx2* subtyping are shown in Table 4, which is divided into two categories: false negatives (3/17), incorrect reported *stx2* subtype 11/17.

Table 4. Incorrect *stx2* subtype results

		Incorrect subtype results		
Strain ID	**EQA provider**	**False negative**	**Incorrect**	**Total true errors**
Strain1	*stx2b*		*stx2c; stx2d* (1)	1
Strain2	*stx2d*		*stx2b* (1), *stx2c* (1), *stx2c; stx2d* (1)	3
Strain3	*stx2a*		*stx2g* (1)	1
Strain4	*stx2a*		*stx2d* (1), *stx2a; stx2c* (1)	2
Strain5	*stx2a*		*stx2g* (1)	1
Strain6	*stx2c*	1	*stx2d* (1)	2
Strain7	*stx2b*		*stx2d* (2)	2
Strain8	-			
Strain9	*stx2a*		*stx2d* (1)	1
Strain10	*stx2a*	1	*stx2a; stx2c* (1)	2
Strain11	*stx2e*		*stx2d* (1)	1
Strain12	*stx2a*	1	-	1
Total		**3**		**17**

3.4 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 12 cluster test strains and eight provided sequences. The pre-categorised cluster of closely related strains contained five Shiga toxin-producing *E. coli* ST301, based on WGS-derived data (Tabel1). The EQA provider's cluster analysis of WGS-derived data was based on an allele-based (cgMLST [11]) and SNP analysis (NASP [12]).

The correct cluster based on WGS-derived data contained six ST301 strains: strain4, strain10, strain14, strain15, strain19, and strain20 (strain4/strain10 were duplicates). As previously mentioned, the strain20 sequence originated from strain14 but was contaminated with 8% *S.a sonnei* by the EQA provider. The EQA provider found at most four allele differences or 0 SNPs between any two strains in the cluster. All downloaded sequences should be QC evaluated and included in an analysis with the own produced WGS data. (Annexes 5-11).

3.4.1 WGS-derived data

3.4.1.1 Reported details on equipment and method

Twenty-three participants (88%) performed cluster analysis using WGS-derived data. One laboratory reported using external assistance for sequencing. The participants utilised various sequencing platforms, including 1 Miniseq, 8 MiSeq, 11 NextSeq, 1 Novaseq, and 2 Ion Torrent (Ion GeneStudio S5 System and Ion Torrent S5XL). All laboratories reported using commercial kits for library preparation. Of the 23 participants, 18 (78%) used Illumina's Nextera kit. One participant reported making changes to the shearing time compared to the manufacturer's protocol (Annex 6).

3.4.1.2 Assessment of the QC- status of the provided sequences

The participants were instructed to describe their QC observations and considerations leading to the QC status decision and the following cluster analysis for the additional test strains (provided genome sequences) strain13-20. The three level of QC-status were A: Acceptable quality, B: Quality only acceptable for outbreak situations (less good quality) and C: Not acceptable quality - strain not analysed. The EQA-provider had modified three sequences (strain17, strain18, and strain20): one with low coverage, one with low quality of R2, and one with contamination. (Table 5). A participant reported QC-status 'C' for all sequences but did not conduct any QC analysis; nonetheless, they included the sequences in the cluster analysis.

The manipulations of the three strains were as followed:

Strain17, a non- cluster sequence with low quality of R2. The provider's QC pipeline only accepts paired-end reads. Consequently, the quality of the R2 read is too low for proper analysis. Therefore, a single-end assembly (using R1) was employed in the cluster analysis by the EQA-provider. The majority of the participants (18/23, 78%) reported the sequence as either QC-status B or C and 5 participants reported the sequence a having an acceptable quality (QC-status A) for further analysis.

Strain18, a non-cluster sequence with reduced coverage and removal of genes, exhibited a quality issue. All participants (100%) correctly identified the problem with the sequence. However, three participants accepted the

quality for outbreak investigation (QC-status B), and only for one participant did the result negatively impact the cluster identification.

Strain20, a cluster sequence (strain14) contaminated with approx. 8% *S. sonnei*, 48% (11/23) reported the contamination of the sequence as either QC-status B or C. Conversely, 12 participants reported the sequence to have acceptable quality (QC-status A). The EQA provider anticipated that all participants would recognise the contamination. Among the 11 who acknowledged the contamination, three reported a C level. Yet, one participant included the contaminated sequences in the analysis and cluster identification despite noting the issue.

Table 5. Results of the participants' QC assessment of the EQA modified provided sequences

Genome	Characteristics	Provider	A	B	C
Strain17	A nonCluster sequence with low quality of R2	B	5	11	7
Strain18	A nonCluster sequence with low coverage	C	0	3	20
Strain20	A Cluster sequence contaminated with approx. 8% *S. sonnei*	B	12	8	3

A: Acceptable quality
B: Less good quality
C: Not acceptable quality
Raw data available in Annex 11

Four of the five sequences provided without modification were reported as acceptable quality (QC-status A) from all participants, except for one who did not assess the quality. Additionally, Strain19 was deemed acceptable quality for outbreaks (QC-status B) by ten participants, as the number of contigs were either near or above their specified threshold.

3.4.1.3 Cluster analysis

Each participant is required to employ both their self-generated sequences and the provided sequences (post-assessment of QC status) during the cluster analysis. Thereafter, participants should report the strains/sequences that form a closely related cluster, simulating an outbreak scenario. In this context, it is essential to assess the sequences even in cases of poor quality, illustrating a situation where rerunning the sequence is not feasible.

Over the years the cluster analysis has developed to also encompass QC -borderline sequences. Strain20 from this year is an example - a cluster sequence with approximately 8% *S. sonnei,* influencing the cluster identification. Performance in the cluster analysis with WGS-derived data was high (65%) when accepting cluster identification without considering strain20, especially if its exclusion was based on QC rather than the cut-off. Thirteen participants (57%) accurately identified the cluster of closely related strains, as defined by pre-categorisation from the EQA provider, among the 12 test strains and eight sequences (Table 6).

Laboratories were instructed to report the data analysis used for cluster identification and use strain15 (sequence) as a representative in the cluster for reporting SNP distance or allelic differences. Laboratories could report results up to three analyses (one main and up to two additional), but the detected cluster had to be based on results from the main analysis.

Table 6. Results of cluster identification based on WGS-derived data

	Strain ID																				
Lab No.	1	2	3	4‡#	5	6	7	8	9	10‡#	11	12	13	14‡	15‡	16	17	18	19‡	20‡	Main Analysis	Cluster identified
19	-	-	-	+	-	-	-	-	-	+	-	-	-	+	+	-	-	ND	+	+	A[a]	+
34	-	-	-	+	-	-	-	-	-	+	-	-	-	+	+	-	-	ND	+	+	A	+
80	-	-	-	+	-	-	-	-	-	+	-	-	-	+	+	-	-	ND	+	+	A	+
88	-	-	-	+	-	-	-	-	-	+	-	-	-	+	+	-	-	ND	+	+	A[a]	+
90	-	-	-	+	-	-	-	-	-	+	-	-	-	+	+	-	-	ND	+	+	A[c]	+
100	-	-	-	+	-	-	-	-	-	+	-	-	-	+	+	-	-	-	+	+	A	+
108	-	-	-	+	-	-	-	-	-	+	-	-	-	**-**	+	-	-	ND	+	**-**	S	No
123	-	-	-	+	-	-	-	-	-	+	-	-	-	+	+	-	-	**+**	+	+	A	No
124	-	-	-	+	-	-	-	-	-	+	-	-	-	+	+	-	-	ND	+	ND	A[c]	(+)
127	-	-	-	+	-	-	-	-	-	+	-	-	-	+	+	-	ND	ND	+	ND	A	(+)
128	-	-	-	+	-	-	-	-	-	+	-	-	-	+	+	-	-	-	+	+	A[a]	+
129	-	-	-	+	-	-	-	-	-	+	-	-	-	+	+	-	-	ND	+	+	A	+
131	-	-	-	+	-	-	-	-	-	+	-	-	-	+	+	-	-	ND	+	+	A	+
132	-	-	-	+	-	-	-	-	-	+	-	-	-	+	+	-	-	**+**	+	+	S	No
133	-	-	-	+	-	-	-	-	-	+	-	-	-	+	+	-	ND	ND	+	+	A	+
134	-	-	-	+	-	-	-	-	-	+	-	-	-	+	+	-	-	ND	+	+	A	+
135	-	-	-	+	-	-	-	-	-	+	-	-	-	+	+	-	ND	ND	**-**	+	A	No
136	-	-	-	+	-	-	-	-	-	+	-	-	-	+	+	-	-	ND	+	+	A	+
138	-	-	-	+	-	-	-	-	-	+	-	-	-	+	**-**	-	ND	ND	**-**	**-**	A	No
139	-	-	-	+	-	-	-	-	-	+	-	-	-	+	+	-	ND	ND	+	+	A	+
153	-	-	-	+	-	-	-	-	-	+	-	-	-	+	+	-	-	ND	+	**-**	S	No
187	-	**+**	-	+	-	-	-	-	-	+	-	-	-	+	+	-	ND	ND	+	+	A	No
222	-	-	-	+	-	-	-	-	-	**-**	-	-	-	+	+	-	-	ND	+	+	A	No

‡: closely related strains (in grey)
#: technical duplicates strains
A: Allele-based
S: single-nucleotide polymorphism (SNP-based)
Additional analysis: [a] = SNP-based, [b] = single-nucleotide variant (SNV-based), [c] = Allele-based
ND: not done
+: Reported to be a closely related strain
-: Reported not to be closely related strain
Errors in bold
(See Annex 7).

Three participants (108, 132, and 153) utilised SNP as their main analysis method, and three laboratories reported SNP as an additional analysis. Both Laboratories 108 and 153 employed a reference-based approach with EQA strain15 as the reference. However, none of them could correctly identify the cluster, each for different reasons. Laboratory 108 did not include SNP distances above five (excluding strain14 and strain20). Laboratory 132 omitted the quality control check of the provided data, incorporating the modified strain18 with very low coverage into the analysis and incorrectly concluding that it belonged to the cluster. Laboratory 153 did not recognise strain20 as contaminated and consequently excluded the strain from the cluster because of eight SNPs. Additionally, of the six laboratories reporting SNP results, three used the Burrows-Wheeler Aligner (BWA) as their read mapper, one employed BWA-maximal-exact-matches (MEM), another utilised CLC assembly cell, and the final laboratory employed CSI Phylogeny. Four different variant callers were utilised (Table 7/Figure 9).

Table 7. Results of SNP-based cluster analysis

Lab No.	SNP-based							
	SNP Pipeline	Approach	Reference	Read mapper	Variant caller	Identified Pre-defined Cluster	Distance within cluster	Distance outside cluster
Provider	NASP [11]	Rb	Strain15	BWA	GATK	Yes	0–0	121-514
19*	NASP	Rb	Strain15	BWA	GATK	Yes	0-0	93-128
88*	Snippy	Rb	Strain15	BWA	Freebayes	Yes	0-4	-
108	In-house pipeline	Rb	Strain15	CLC assembly cell	CLC assembly cell	No	0-5	8-101211
128*	CSI Phylogeny	Ab	SPAdes	-	SAMtools	Yes	0-46	47-38339
132	CSI Phylogeny	Rb	GCF_000008865.2_ASM886v2	BWA-MEM	SAMtools	No	0-4	12-18479
153	CSI Phylogeny	Rb	Strain15	CSI Phylogeny	CSI Phylogeny	No	0-0	8-127

**: additional SNP-based analysis*
Ab: Assembly-based
Rb: Reference-based
(See Annex 8).

Twenty participants employed allele-based analysis as the main method for cluster detection, with two reporting additional allele-based analyses (Laboratories 90 and 124, as shown in Table 8) and three reporting additional SNP analyses. The majority of laboratories (13 out of 20, 65%) used only assembly-based (OAB) allelic calling methods, while six laboratories (30%) utilised assembly- and mapping-based (A&M) methods. One laboratory (128) solely employed mapping-based (OMB) methods. Laboratory 124 initially used A&M for allelic calling and adopted OAB for additional analysis (Table 8).

Table 8. Results of allele-based cluster analysis

Lab No.	Allele-based analysis							
	Approach	Allelic calling method	Assembler	Scheme	No. of loci	Identified Pre-defined Cluster	Difference within cluster	Difference outside cluster
Provider	BioNumerics	A&M	SPAdes	Applied Mathss (cgMLST/Enterobase)	2513	Yes	0–4	24-2346
19	BioNumerics	A&M	SPAdes	Applied Mathss (cgMLST/Enterobase)	2513	Yes	0-3	21-2350
34	SeqSphere	OAB	Skesa	Enterobase (cgMLST)	2513	Yes	0	25-2336
80	SeqSphere	OAB	Skesa	Enterobase (cgMLST)	2513	Yes	0	21-2290
88	INNUca, chewBBACA and ReporTree	OAB	SPAdes[a]	INNUENDO wgMLST	7601	Yes	0-13	34-2766
90	SeqSphere	A&M	SKESA	Enterobase (cgMLST)	2513	Yes	0	25-2335
90*	SeqSphere	A&M	Enterobase	Enterobase (cgMLST)	2513	Yes	0-2	26-2359
100	SeqSphere	OAB	SPAdes	Enterobase (cgMLST)	2513	Yes	0-7	25-7393
123	SeqSphere	OAB	SPAdes	Enterobase (cgMLST)	2513	No	0-6	22-2291
124	BioNumerics	OAB	SPAdes	Applied Mathss (cgMLST/Enterobase)	2506	(Yes)	0-1	24-2350
124*	Enterobase	A&M	SPAdes	Enterobase (cgMLST)	2513	Yes	0-4	28-2391
127	Enterobase	OAB	SPAdes	Enterobase (cgMLST)	2513	(Yes)	0-4	27-2393
128	Enterobase	OMB	-	Enterobase (cgMLST)	2513	Yes	0-1	19-2355
129	SeqSphere	OAB	Velvet	SeqSphere/ Target Definer	1514	Yes	0	13-1437
131	SeqSphere	OAB	SKESA	Enterobase (cgMLST)	2513	Yes	0	25-2336
133	BioNumerics	OAB	SPAdes	Applied Mathss (cgMLST/Enterobase)	2513	Yes	0-1	12-200
134	SeqSphere	A&M	SPAdes v3.15.4	Enterobase (cgMLST)	2513	Yes	0-1	25-2347
135	SeqSphere	OAB	SPAdes	Enterobase (cgMLST)	2513	No	0-1	25-2341
136	SeqSphere	A&M	Unicycler	Enterobase (cgMLST)	2513	Yes	0-3	25-7353
138	Enterobase	OAB	shovill,1.1.0	Enterobase (cgMLST)	2513	No	8-13	16-2318
139	Enterobase	A&M	SPAdes	Enterobase (cgMLST)	2513	Yes	0-5	27-2389
187	SeqSphere	A&M	Skesa	Enterobase (cgMLST)	2513	No	0-2138	23-2328
222	PHANtAsTiC pipeline, IRIDA-ARIES webserver as calculation engine.	OAB	SPADES[b]	Innuendo-curated Enterobase scheme	2360	No	0-16	17-2235

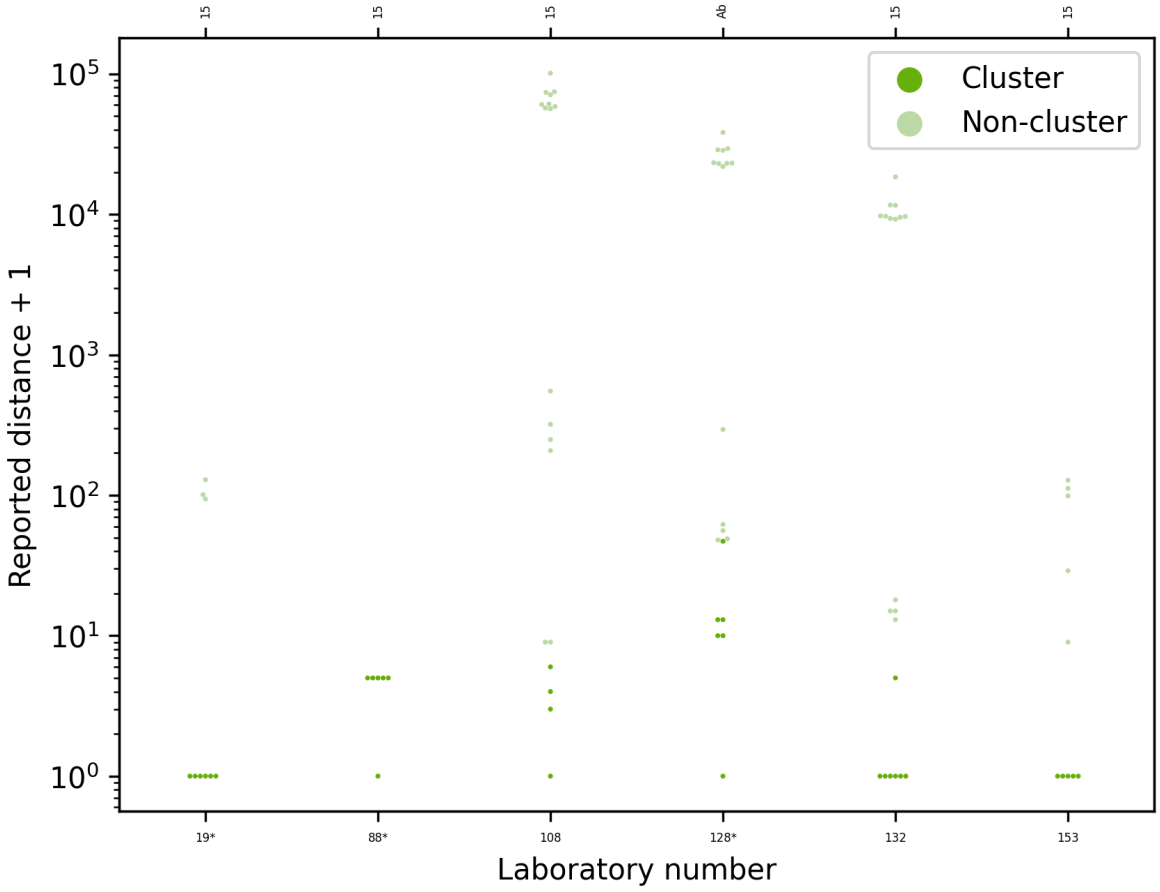
**: additional analysis*
A&M: Assembly- and mapping-based
OAB: Only assembly-based
(See Annex 8).
OMB: Only mapping-based
[a]: SPAdes 3.14.0, implemented in INNUca v4.2.2
[b]: SPADES 3.15, default parameters, filtering with the tool.

Of the 20 laboratories using allele-based methods as their main analysis, 75% (15 out of 20 laboratories) correctly identified the cluster of five (six if strain20 is included) closely related strains (Table 8). Sixteen laboratories performed cgMLST using the same scheme as the EQA provider (cgMLST/Enterobase [10]) with 2 513 loci as their main analysis. Three laboratories (124, 129, and 222) used a scheme with a lower number of loci (2 506, 1 514, and 2 360). Additionally, one laboratory (88) used wgMLST as the main analysis and SNP as an additional analysis, obtaining allelic differences within the correct cluster ranging from 0 to 13 (7 601 loci).

Eleven of the 15 laboratories that identified the correct cluster reported allele differences of 0–3 within the cluster of closely related strains in their main analysis (Figure 10, Table 8). Laboratories 100, 127, and 139 reported a slightly higher number of allelic differences of 0–7 within the cluster strains (strain4, strain10, strain14, strain15, strain19, and strain20). Laboratory 88 used wgMLST as described above.
Thirteen of the 20 laboratories that accurately identified the ST301 cluster incorporated strain20 into their analysis. Twenty-two laboratories (22 out of 23, 96%) correctly included the duplicate strains, strain4 and strain10, in the cluster. However, Laboratory 222 excluded strain10 but not strain4 from the cluster, despite strain4 and strain10 having 16 and 17 allelic differences from strain15, respectively. Two of the laboratories (123 and 100), using allele-based analysis, included the manipulated strain18 (with low coverage), believing it could be used in an outbreak situation, into their analysis, and one mistakenly reported it in the final cluster (with 6 allelic differences). Laboratory 187 mistakenly reported strain2 (ST17) as part of the cluster, accounting for their reported allelic differences spanning 0-2138. Additionally, two laboratories (135 and 138) excluded either strain19 or both strain15 and strain20 from the reported cluster. Both laboratories reported a high number of allelic differences for strain19 and were among the 10 participants that reported QC issues for strain19.

Of note, one additional test strain (strain12) was also identified as ST301 but was not pre-defined by the EQA provider as part of the cluster. Based on the main analysis of cgMLST, 20 laboratories reported allele differences between the selected cluster strain and this strain at distances 12-45 (Table 8, Annex 8).

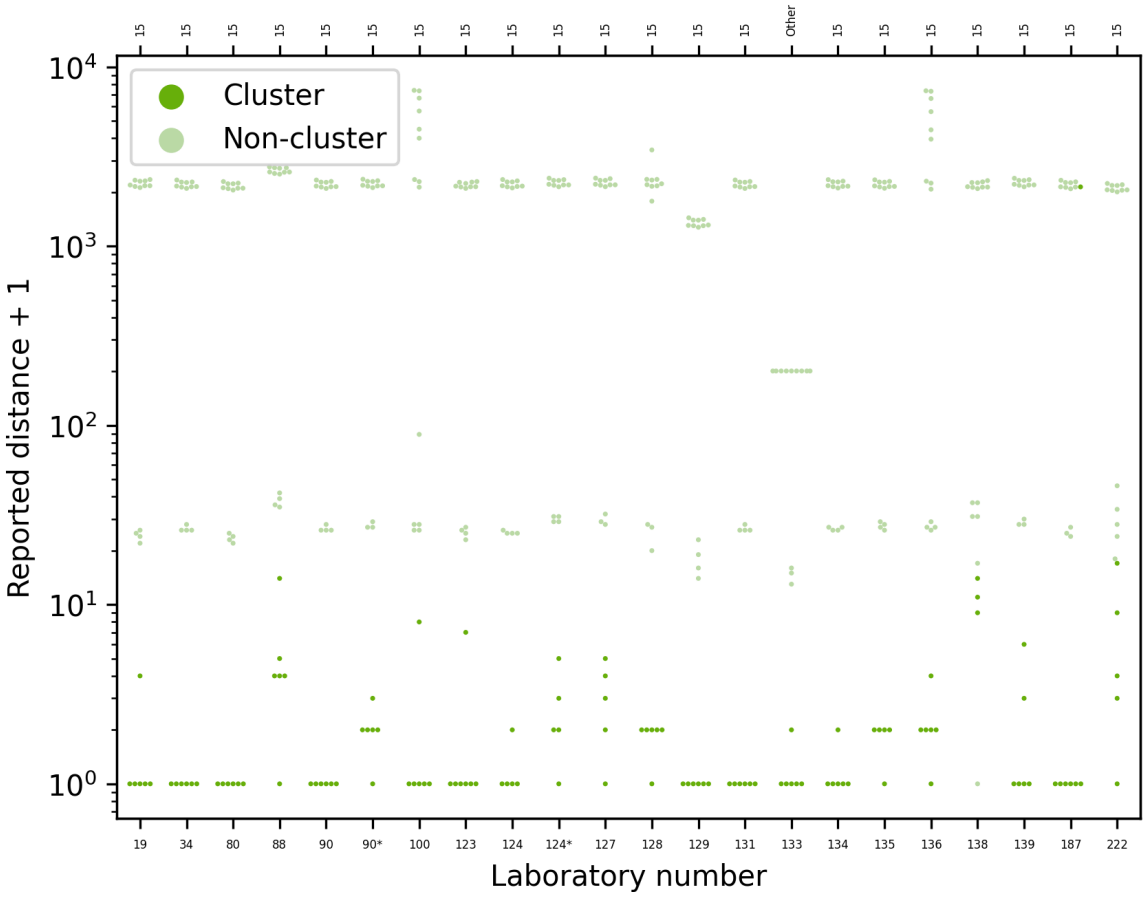
Figure 9. Reported SNP distances for each test strain to selected cluster representative strain



SNP: single nucleotide polymorphism; Ab: assembly-based. Participants were instructed to select strain15 as reference (listed as '15' on the top scale). Dark green: reported cluster of closely related strains, Light green: not reported as part of cluster.

Laboratory 153 did only identify the cluster without strain20, as they excluded all above 5 SNP from the cluster. (Table 7/Figure 9). Laboratories 108 and 132 did not identify the correct cluster of closely related strains, as 108 excluded strain14 from the analysis and laboratory 132 included the manipulated strain18 with low coverage.

Figure 10. Reported allelic differences for each test strain to selected cluster representative strain



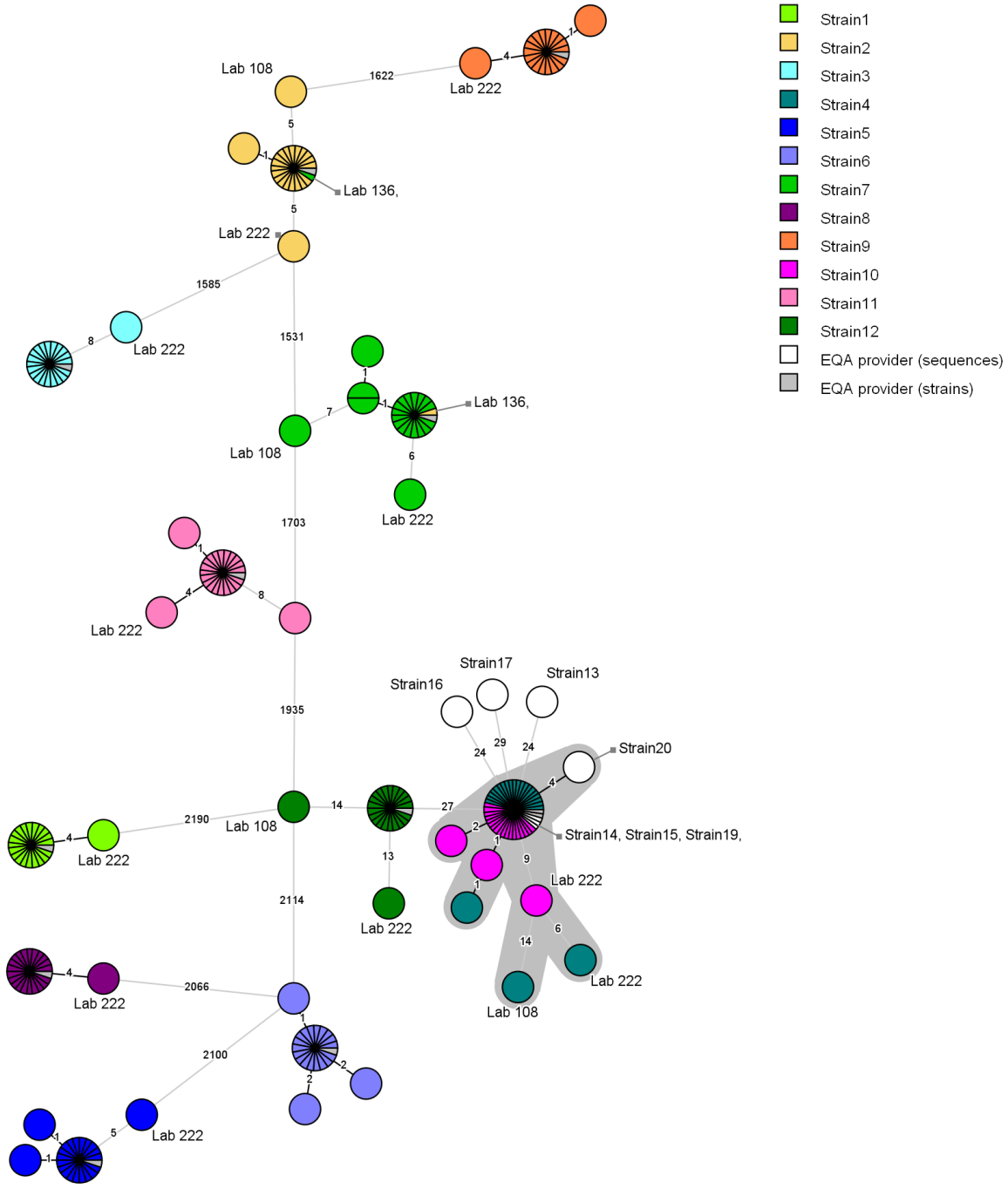
Other: One participant used another reference: GCF_000008865.2_ASM886v2. Participants were instructed to select strain15 as reference (listed as '15' on the top scale). Dark green: reported cluster of closely related strains, Light green: not reported as part of cluster.

3.4.1.4 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´s calculation engine for allele calling (Enterobase) [11] and evaluated by the EQA provider's in-house quality control (QC) pipeline [116].

The overall cgMLST analysis, shown in the minimum spanning tree (MST) based on submitted raw reads (FASTQ files) from the 23 laboratories revealed a clear clustering of the results for each test strain (Figure 11). Laboratory 108 and 122 did fall slightly outside of the clusters from each of the test strains (1-14 alleles) this is likely due to artefacts from comparing Ion Torrent generated data with Illumina data. Laboratory 136 mistakenly switched around strain2 and strain7 (Figure 11).

Figure 11. Minimum spanning tree of core genome multilocus sequence typing participant FASTQ files

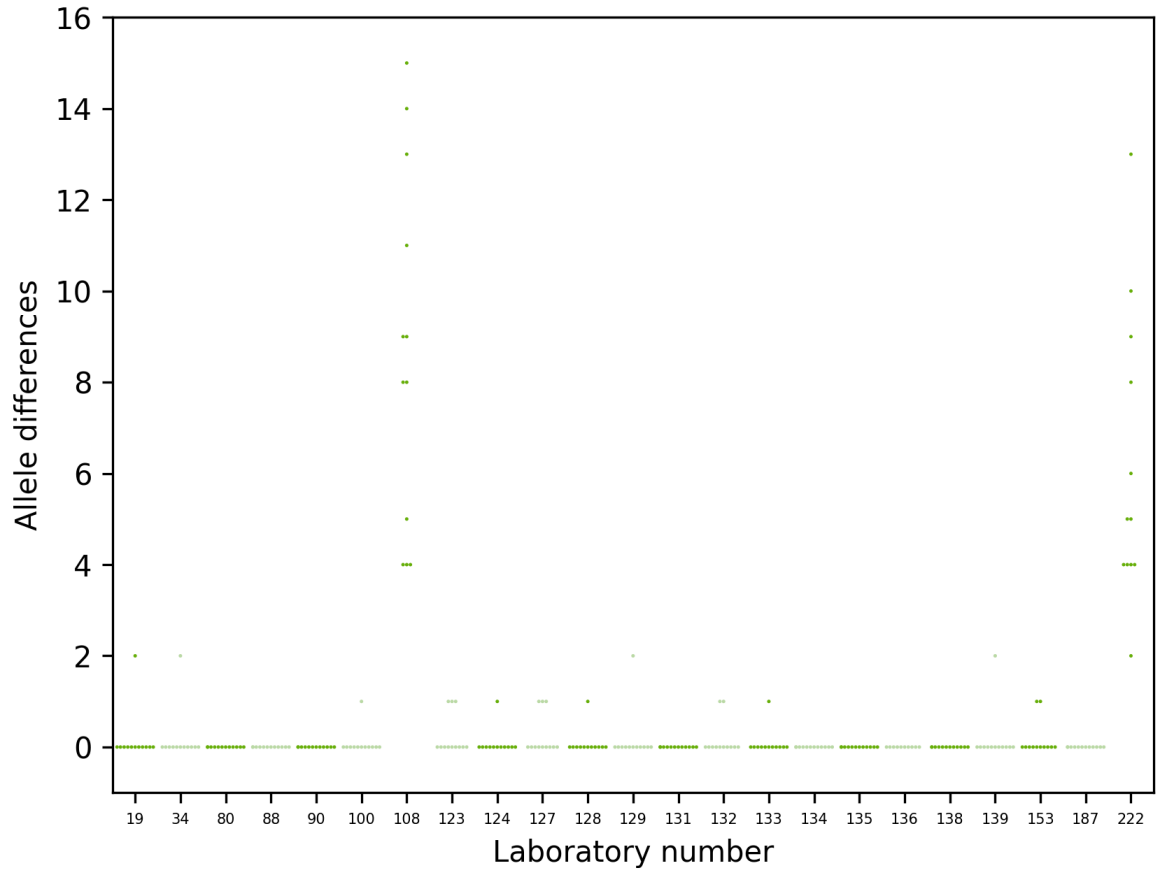


Minimum spanning tree (MST) in log scale of core genome multilocus sequence typing (cgMLST) [11] based on submitted raw reads (FASTQ files). Each of the strain1–12 test strains have a different colour. The EQA-provided sequences for strain1-strain12 from the EQA provider are in grey, and the provided sequences (strain13-20) are in white. The provided modified sequence with poor quality (strain18) was not included in the analysis; however, strain20 (modified with 8% S. sonnei) and the single-end assembly of strain17 (using R1) were used in the cluster analysis. Strain4 and strain10 were technical duplicates. Fourteen sequences were excluded as the core percent was below 94 (seven from laboratory 108, caused by the Ion Torrent data). Results from laboratories 108 and 222 were run in CE, using the Ion Torrent setup for allele calling.

The allele differences in Figure 11 do not exactly match those illustrated in the individual reports, and consequently there are discrepancies in Figure 11, where the same data are used. This discrepancy is caused by loci being dropped if they did not pass QC for all strains in the analysis. As a result, the joint analysis contains fewer loci.

For each laboratory, cgMLST was performed on the submitted raw reads (FASTQ files), applying Applied Mathss allele calling with the Enterobase scheme [11]. A hierarchical single linkage clustering was performed on the submitted data for each laboratory along with the EQA provider's reference strains (strain15). Figure 12 shows the allele differences between each submitted sequence and the corresponding reference.

Figure 12. Participant allele difference from reference result (EQA-provider) for each test strain



Allele difference from corresponding stain1-12 (EQA provider) based on submitted raw reads (FASTQ files) and analysed by EQA provider, two result from laboratory 136 was excluded from the figure.

For 220 of 276 results (80%), no allele difference was identified. For 19 results (7%), a difference of one to two alleles from the reference strain was calculated, and for 23 results (8%), a difference of two to 15 alleles was observed. These differences were primarily reported by Laboratories 108 and 222. Note that for Laboratory 136, only 10 strains were included in the plot due to the assumed inadvertent swapping of two strains (strain2 and strain7), resulting in allele differences of 1675 and 1679 from the reference strain. Consequently, the EQA provider excluded the results from Figure 12.

Separately, the laboratories listed quantitative and qualitative QC parameters used to evaluate their data. As seen in Table 9, almost all laboratories have implemented QC threshold for accepting the data. The most reported parameters were coverage with acceptance thresholds at 20-100X followed by confirmation of genus. Genome size and difference Q score parameters were also included. The number of good cgMLST *loci* was also listed as an important parameter for QC. The additional QC parameters reported by the participants are listed in Annex 9.

Table 9. Summary of selected QC parameters reported by participants

Laboratory	Confirmation of genus	Coverage	Q score (Phred)	Genome size	No. of good cgMLST loci
19	Kraken and Bracken analysis and <5% contamination with other genus	Min. x 50 coverage	No	4.64 – 5.56 Mb	Minimum 95% core percent and maximum 50 loci with multiple consensus
34	Mash Screen used by Ridom SeqSphere	more than 70fold is optimal	No	approx. 5 Mb	more than 95%
80	Kraken and Mash in SeqSphere	>50	No	4.4 - 5.3	>90%
88	Kraken (as implemented in INNUca v4.2.2) and ConFindr.	INNUca v4.2.2 15x for the first estimated coverage; 30x for the assembly coverage.	INNUca v4.2.2 (FastQC + trimming/filtering of the reads with Trimmomatic, default settings)	INNUca v4.2.2 5.0 Mb	Allele calling chewBBACA v3.1.2 and filtering of loci called in <90% of the samples, and of samples with <95% loci called with ReporTree clustering analysis
90	PubMLST rMLST, CGE KmerFinder, Ridom Mash Distance	40x	No	4.9-5.9 Mb	>95% good targets
100	KmerFinder3.1 Center for Genomic Epidemiology	50x	FastQC, threshold to 30	SPAdes assembler, 5.0-5.5 Mb	SeqSphere cgMLST sheme, 95% good targets threshold
108	No	≥20x coverage	No	4,8-6,0 Mb + ≥20x coverage	>10x
123	Contamination Check (Mash Screen) in SeqSphere	>50	No	5.0-6.0 Mb	>98
124	length, GC% and in silico PCR E coli det	>100 (acceptable >30 in BioNumerics)	Q30 >60	3.9 - 6.5 Mb	% alleles called available in BioNumerics (>80%)
127	EntroBase (Kraken)	No	No	3.7 - 6.4 Mbp	No
128	KmerFinder 3.2	Enterobase, at least 50x	FASTQC, >Q20	SPAdes assembler	Enterobase
129	No	30	No	No	90%
131	Mash screen (incl. in SeqSphere)	50x<	30<	4.5-5.5 Mb	>95%
132	Bifrost	over 50, however 25-50 are sometimes used.	No	No	No
133	Predicted pathotype in BioNumerics & PubMLST Species ID	>30	>30	5-5.8 Mb	corePercent >=96
134	Mash Screen in SeqSphere and ID species in PubMLST	>= 50	No	length of contigs assembled < Ref genome + 10%	> = 98%
135	There is a species identification tool based on Kraken2/Bracken built in our in-house assemebly pipeline ('Juno')	>30	>30	4.6 - 5.8 Mb	>90% of alleles
136	K-mer	50x	No	5.0	99 % good targets

Laboratory	Confirmation of genus	Coverage	Q score (Phred)	Genome size	No. of good cgMLST loci
138	Kraken2 with database built from all refseq genomes; rMLST from Pubmlst	in house calculation. >50x	>=q30; fastp 0.22.0.	4.909 < x ≤ 5.493 Mb	no more than 2% of loci missing
139	In house blastn based script	> 45 X	Discarding reads with Q scores < 15 on a minimum length of 50 bp	4.7 – 5.9 Mb	No
153	KmerFinder, SpeciesFinder	>30	20	4 - 6,5 Mbp	No
187	Kmer finder, rmlst	30	No	5.1 Mb	>95%
222	No mismatches in the alignment with the 7 housekeeping genes of MLST panel (Warwick); Result of rMLST on https://pubmlst.org/species-id	Minimum 50x average depth of coverage across the genome was considered as threshold value	No	No	Quality threshold for reliability of cluster analysis was set at 80% of loci found out of those part of the scheme (1880/2360)
% of laboratories using the QC parameter	**91%**	**96%**	**43%**	**87%**	**83%**

See Annex 9 for additional information.

For each laboratory, the submitted raw reads (FASTQ files) were evaluated by the EQA provider's in-house quality control pipeline [16]. For the full QC evaluation of all strains, see Annex 10. According to the QC parameters, the sequencing quality was uniformly good. The EQA-provider has disregarded any contamination warnings for strain3 or strain11, as similar results were observed for the EQA provider (Warning *, Annex 10). *Shigella flexneri* and *E. coli* are highly genetically related and species identification using WGS is not always sufficient. However, nine laboratories (Table 10) received warnings of their QC status, primarily warnings with average coverage below 50.

Table 10. Results of raw reads submitted by participants evaluated by EQA provider QC pipeline summarised by laboratory

*Ranges**	*{Ec}*			*{5%}*	*{4.5-5.8}*	*{<250}*	*{>0}*	*{<1000}*	*{>50}*					
Lab No.	**Detected species**	*% Species 1*	*% Species 2*	*Unclassified reads (%)*	*Length at >25 x min.*	*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 (Biffret)*
19	*Ec*	*68.8-95.7*	*0.1-21.7*	*4.0-14.0*	*5.0-5.5*	*1.9-33.1*	*189.0-419.0*	*2.0-41.0*	*61.0-134.0*	*2485.0-4953.0*	*140.0-144.0*	*250.0-297.0*	*39.0-81.0*	
34	*Ec*	*45.9-94.1*	*0.0-43.9*	*5.5-16.1*	*5.0-5.5*	*0.0-0.9*	*84.0-377.0*	*0.0-2.0*	*47.0-120.0*	*1936.0-5239.0*	*126.0-143.0*	*203.0-278.0*	*64.0-248.0*	*W*
80	*Ec*	*62.8-94.0*	*0.6-25.8*	*5.0-15.7*	*5.0-5.5*	*0.0-0.0*	*92.0-342.0*	*0.0-0.0*	*119.0-235.0*	*4510.0-8488.0*	*151.0-151.0*	*233.0-289.0*	*60.0-221.0*	
88	*Ec*	*48.6-97.8*	*0.0-45.5*	*2.0-19.9*	*5.0-5.5*	*0.0-9.2*	*90.0-374.0*	*0.0-11.0*	*103.0-158.0*	*3887.0-5743.0*	*146.0-148.0*	*323.0-405.0*	*54.0-119.0*	
90	*Ec*	*77.2-98.0*	*0.1-15.7*	*1.6-11.3*	*5.0-5.5*	*0.3-48.8*	*111.0-335.0*	*1.0-42.0*	*50.0-152.0*	*1103.0-3677.0*	*211.0-240.0*	*262.0-408.0*	*55.0-141.0*	*W*
100	*Ec*	*51.8-97.9*	*0.1-43.5*	*1.4-10.8*	*5.0-5.5*	*0.0-286.6*	*74.0-441.0*	*0.0-70.0*	*66.0-89.0*	*1361.0-2085.0*	*236.0-270.0*	*241.0-317.0*	*67.0-248.0*	
108	*Ec*	*85.6-98.1*	*0.2-8.8*	*1.5-8.5*	*4.8-5.2*	*0.7-5.8*	*2514.0-3732.0*	*3.0-20.0*	*62.0-93.0*	*1135.0-1777.0*	*285.0-294.0*	*0.0-0.0*	*2.0-4.0*	
123	*Ec*	*48.3-94.6*	*1.1-43.6*	*3.0-20.2*	*0.4-5.5*	*0.4-4862.3*	*70.0-372.0*	*1.0-284.0*	*26.0-71.0*	*565.0-1567.0*	*201.0-268.0*	*223.0-402.0*	*40.0-237.0*	*W*
124	*Ec*	*82.2-97.7*	*0.1-9.3*	*2.0-11.2*	*5.0-5.5*	*0.0-0.0*	*54.0-250.0*	*0.0-0.0*	*262.0-297.0*	*10000.0-10000.0*	*151.0-151.0*	*396.0-453.0*	*67.0-247.0*	
127	*Ec*	*63.7-92.9*	*0.2-30.4*	*3.8-20.5*	*5.0-5.5*	*8.2-49.5*	*190.0-481.0*	*8.0-44.0*	*68.0-147.0*	*2501.0-5553.0*	*147.0-149.0*	*327.0-374.0*	*32.0-66.0*	*W*
128	*Ec*	*78.2-97.2*	*0.1-12.4*	*2.5-18.4*	*5.0-5.5*	*0.0-0.0*	*71.0-343.0*	*0.0-0.0*	*125.0-198.0*	*4439.0-9168.0*	*99.0-149.0*	*139.0-349.0*	*65.0-249.0*	
129	*Ec*	*77.6-96.9*	*0.1-12.8*	*2.5-11.7*	*4.9-5.4*	*2.9-57.0*	*167.0-440.0*	*4.0-44.0*	*69.0-136.0*	*2380.0-4918.0*	*145.0-148.0*	*311.0-431.0*	*46.0-85.0*	
131	*Ec*	*72.6-97.3*	*0.1-19.4*	*2.2-11.9*	*5.0-5.5*	*0.0-0.0*	*79.0-383.0*	*0.0-0.0*	*109.0-152.0*	*4138.0-5679.0*	*149.0-149.0*	*250.0-310.0*	*64.0-222.0*	
132	*Ec*	*49.8-98.3*	*0.1-44.8*	*1.5-15.1*	*1.0-5.3*	*2.1-4326.6*	*60.0-242.0*	*6.0-675.0*	*26.0-60.0*	*965.0-2182.0*	*148.0-150.0*	*262.0-434.0*	*22.0-224.0*	*W*
133	*Ec*	*63.7-98.0*	*0.1-29.8*	*1.6-11.1*	*5.0-5.6*	*0.0-52.0*	*49.0-273.0*	*0.0-40.0*	*52.0-206.0*	*1053.0-4227.0*	*264.0-290.0*	*299.0-498.0*	*43.0-355.0*	*W*
134	*Ec*	*54.7-97.8*	*0.1-39.1*	*2.0-13.3*	*5.0-5.5*	*0.0-0.5*	*80.0-313.0*	*0.0-1.0*	*55.0-101.0*	*2037.0-3608.0*	*146.0-149.0*	*227.0-348.0*	*60.0-248.0*	
135	*Ec*	*79.9-97.7*	*0.1-11.8*	*1.8-11.1*	*5.0-5.5*	*0.0-0.0*	*64.0-279.0*	*0.0-0.0*	*81.0-263.0*	*2978.0-9663.0*	*151.0-151.0*	*287.0-387.0*	*64.0-327.0*	
136	*Ec*	*72.0-97.2*	*0.1-20.2*	*2.3-11.2*	*5.0-5.5*	*0.0-0.0*	*64.0-289.0*	*0.0-0.0*	*148.0-303.0*	*5302.0-10744.0*	*147.0-149.0*	*301.0-388.0*	*66.0-319.0*	
138	*Ec*	*56.5-98.6*	*0.0-38.9*	*1.3-18.7*	*5.0-5.5*	*0.0-0.0*	*70.0-419.0*	*0.0-0.0*	*298.0-446.0*	*10640.0-16060.0*	*149.0-151.0*	*0.0-406.0*	*66.0-319.0*	
139	*Ec*	*86.2-95.1*	*0.1-4.4*	*4.6-12.3*	*5.0-5.5*	*0.0-20.8*	*110.0-295.0*	*0.0-9.0*	*63.0-141.0*	*2171.0-5183.0*	*148.0-150.0*	*368.0-404.0*	*63.0-146.0*	*W*
153	*Ec*	*49.6-97.2*	*0.1-44.5*	*2.6-14.3*	*4.9-5.4*	*0.0-312.6*	*89.0-344.0*	*0.0-132.0*	*39.0-66.0*	*1427.0-2493.0*	*141.0-147.0*	*227.0-278.0*	*63.0-248.0*	*W*
187	*Ec*	*75.2-98.1*	*0.1-17.8*	*1.6-10.7*	*4.5-5.5*	*0.0-727.4*	*63.0-417.0*	*0.0-37.0*	*31.0-91.0*	*1077.0-3352.0*	*142.0-147.0*	*196.0-408.0*	*67.0-224.0*	*W*
222	*Ec*	*88.0-97.5*	*0.4-5.4*	*1.8-8.7*	*4.9-5.4*	*0.0-18.9*	*1026.0-3042.0*	*0.0-68.0*	*125.0-225.0*	*2339.0-3696.0*	*272.0-322.0*	*5.0-14.0*	*2.0-9.0*	

**: indicative QC range; Ec: E. coli; W: warnings were noted in the submitted sequences (see Annex 10).*

3.5 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 STEC EQA scheme. The questionnaire contained questions related to accreditation, information on the individual report, actions taken if errors were detected, the usefulness of the QC evaluation of the participant-sequenced data, the relevance of including low-quality data, and suggestions for improvements. The survey response rate was 58% (15/26). The survey results are summarised in Table 11.

Based on the feedback-survey, we conclude that the assessment of the QC of the participants submitted sequences is being appreciated. However, one laboratory emphasised the significance of the EQA provider to optimise analyses also to IonTorrent data in addition to the standardised Illumina data. Two laboratories had feedback regarding reporting of data; streamline the reporting form for ease of use, particularly in virulence gene determination such as *stx* subtyping, while maintaining the option to save progress'; also, 'explore the possibility of sending isolates for various EQAs simultaneously during planned sequencing runs'.

Table 11. Results of evaluation of the EQA scheme

Questions	Response (Yes)	Comments /actions
1) Used for accreditation/licensing purposes?	12/15 (80%)	One laboratory reported using the results as documentation for accreditation
		One laboratory reported using the results to obtain accreditation for NGS in 2024
2) Satisfied with the format/comments?	15/15 (100%)	
3) Differed any of your analytical test results (*) with the expected results. Can you specify which corrective action(s), if any, was/were/will be taken	6/15 (40%)	One laboratory reported that they will review their Standard Operating Procedures (SOPs) and formats regarding registration and species detections
		One laboratory reported that attention will be given to improving reporting in the future, and they have implemented a novel SNPs pipeline to avoid misidentification
		One laboratory reported that they will work on improving diagnostics in the coming period
		One laboratory has set a threshold of 5 SNPs to define a cluster. Consequently, some isolates were erroneously considered outside the cluster when, in fact, they were part of it. As a result, they will adopt a less stringent approach when defining clusters
4) Usefulness of the manipulated sequences?	13/14 (93%)	
5) Usefulness of the QC-status of your submitted sequences?	12/14 (86%)	One laboratory reported producing IonTorrent data and emphasised the need for increased attention by the EQA provider in the analysis of such data
6) Improvements/remarks		One laboratory reported that the process of filling in the reporting form is lengthy and challenging, and a considerable amount of the provided information, such as analytical procedures, is often overlooked in the reports. Additionally, the reporting form for *stx* subtyping is confusing and has led us into errors
		One laboratory suggested making the form more user-friendly, especially when entering results for virulence gene determination. Otherwise, the ability to save progress and return later was appreciated
		One laboratory suggested sending the isolates from different EQAs simultaneously (due to plan sequencing runs)

N=15 for main questions (1-3+6), N=14 for WGS related questions (4–5).

4 Discussion

Based on the completed evaluation, the majority of the participants were satisfied with the format of the individual report and the additional feedback from the EQA provider. However, as the evaluation is bases on anonymised responses, it is not possible to make a follow-up, but all the EQA documents will be discussed during the planning of the next round. Also, the inclusion of the modified sequences in the cluster analysis and the QC feedback of the uploaded sequences was well received by most of the participants. The suggestions are listed in the Section 6 'Recommendations'.

4.1 Serotyping

In the EQA-12, 96% of the laboratories took part in the serotyping component. Out of these, 24% provided phenotypic serotyping results (6/25), while 76% provided molecular serotyping results (two through PCR and 17 through WGS). Compared to EQA-11, there was a decrease in the use of phenotypic serotyping (24% versus 36%).

Encouragingly, in EQA-12, 22 laboratories engaged in complete O:H serotyping, which marked an increase from EQA-11 where there were 19 participants. Among these, 73% (16 out of 22) correctly identified the serotype for all 12 test strains. This represents a slight decrease from EQA-11, where 84% (16 out of 19) accurately assigned the serotypes for all 12 test strains for both O and H.

4.1.1 O group

When looking at the O group participation in previous EQAs we observed an overall decrease from EQA-4 through EQA-10 (26/28; 26/29; 26/29; 27/30; 23/25; 20/24 to 21/26 [93%]) however, in both EQA-11 and EQA-12 we have seen an increase to 25/26 participants.

The O grouping performance in EQA-12 mirrored that of EQA-11, with 17 of 25 participants (68%) accurately O-typing all 12 test strains. This result aligns closely with the 71% achievement in EQA-10, demonstrating consistency, whereas EQA-9 had a lower success rate of only 50%. Similar to EQA-11, but unlike EQA-10, not all the incorrect O group results were reported by laboratories using phenotypic methods. Laboratories 129, 130, 132, and 136 used WGS-based and PCR-based methods and did not determine several O groups (Annex 3). Ten out of the 32 (31%) incorrect results were reported as an incorrect O group, while the remaining (69%) were reported as non-typable/rough or not done. This marks an improvement from EQA-11, where 46 incorrect results were reported, including 12 incorrect assignments of an O group. It's noteworthy that the majority (78%, 25/32) of the incorrect results were reported by three laboratories.

The inclusion of O group O80 underscores the importance of correctly identifying a non-O157 and a newly emerging strain causing human infections [17]. Four participating laboratories reported it as either O86, O180, or non-typable. Half of these laboratories employed phenotypic methods for the O grouping. As such, not all laboratories using WGS-based methods correctly identified O80. The EQA provider has no knowledge of any known cross-reaction between O80 and O86, as well as the other mis-typed O-groups.

Some of the more common O groups exhibited high performances, except for O146 (O157: 100%, O145: 96%, O26: 96%, O103: 92%, and O111: 92%). Among all 12 strains, O-grouping strain11 proved to be the most challenging, with 80% (20 out of 25) correctly identifying the O8 O-type. However, three of the five participants used phenotypical methods. The average score was higher (89%) than the previous EQA-11 (69%) than the previous EQA-10 (86%), EQA-9 (85%), and EQA-8 (79%). Over the past years, there has been a shift from phenotypic serotyping towards WGS-based analysis, which reflects the percentage of participants using WGS (EQA-8 26%, EQA-11 60%. A likely explanation for some of the erroneous O-group performance in EQA-12 is attributed to three laboratories, two of which are using phenotypic serotyping, collectively accounting for 78% of the incorrect results.

4.1.2 H type

Unlike the previous EQAs (EQA-11 84%, EQA-10 94%, EQA-9 94%, and EQA-8 92%) the average performance for correctly H-typing the 12 tests strains was lower (91%) but slightly higher from EQA-11. However, there was an increase in H-typing participation (22 laboratories) compared to EQA-11 (19 participants). The general performance for correctly reporting the H type, of all 12 test strains, was higher (91%) than the O grouping (68%). This might be explained by fewer participating laboratories and that the majority (20/22) used WGS-based methods. Overall, there were three erroneous reported H-typing results; one participant using WGS-based method reported an incorrect H type (H9 in strain5 instead of H28) and one participant used phenotypic methods incorrectly reported H- instead of H31 (strain8) and H14 instead of H2 (strain10). The EQA provider verified the absence of cross-reactions between H9 and H28 or H14 and H2 by testing the reference strains.

4.1.3 OH serotyping

Complete O:H serotyping was performed by 22 (88%) participants with an average score of 97%, and for each strain the score ranged from 91% (20/22) for Strain8 (O154:H31) and strain 11 (O8:H4) to 100% (22/22) for Strain1 (O146:H28), Strain3 (O111:H21), Strain4 (O80:2), Strain6 (O157:H-/H7), and Strain9 (O26:H11). The correct serotype of all 12 strains were reported by 84% (16/22) of the participants who performed the O:H serotyping (Figure 3, Annex 3).

The average percentage O:H serotyping in this EQA was, higher (97%) compared to EQA-11 (95%), EQA-10 (94%), EQA-9 (92%), EQA-8 (86%), EQA-7 (71%), and EQA-6 (78%). In general, the less common European serotypes, such as O8:H4, O80:H2, and O154:H31 proved more difficult to identify particular if participants used phenotypic methods.

In addition to O grouping, H typing plays a crucial role in outbreak detection, epidemiological surveillance, taxonomic differentiation of *E. coli*, and the identification of pathogenic serotypes. Consequently, facilitating the capability of more NPHRLs to conduct thorough and dependable O:H serotyping, especially H typing, remains a significant challenge. However, the adoption of WGS might make this more achievable for some countries in the future.

4.2 Virulence profile determination

Twenty-five laboratories participated in the detection of the virulence profile with the participation rate and performance varying substantially between the different tests. The participation of the genotypical detection was higher compared to EQA-11; as such, the highest rate was as follows; *stx1* (97%), *stx2* (98%), *eae* (96%), *aggR* (98%), *esta* (98%), and the average score of the combined *stx* subtyping (93%).

4.2.1 Detection of *aggR* and *esta*

The performance in detection of the EAEC *aggR* gene was high, with 98% of the participants correctly identifying *aggR* (23/24). This is comparable to EQA-11, where 95% correctly identified *aggR*. The average performance for *esta* was higher (98%) than in EQA-11 (89%). This performance was attributed to four laboratories (108, 125, 128, and 187) that couldn't identify the gene in strain11. All laboratories, except one, utilised a WGS-based method.

4.2.2 Detection of *eae*

Genotyping of *eae* had a high participation rate (96%) but a lower performance than EQA-11; 19 (76%) laboratories obtained a 100% score, giving an average score of 96%. The lower performance was attributed to two laboratories than reported 62% (eight out of 13) of the incorrect results. The average correct score has been fairly unchanged through the EQAs (EQA-4 to EQA-11, 96%-99%).

4.2.3 Detection of *stx1* and *stx2*

Both the participation (96%) and performance rates were high for genotyping of *stx1* (97%) and *stx2* genes (97%), similar to previous EQAs. As seen in previous EQAs the majority of the incorrect results were reported for *stx2*.

4.2.4 Subtyping of *stx1* and *stx2*

Comparable to EQA-11, the average score of laboratories that correctly performed the *stx* subtyping were; 95% for *stx1*, 94% for *stx2*, and 93% combined *stx1* and *stx2*. Though not as high, as last year's EQA-11 (*stx1* 99%), this year's EQA still showed an increase compared to both EQA-9 (93% and 92%) and EQA-8 (84% and 87%) and all previous EQAs. The unexpected reporting of 'not done' results, which was an issue in EQA-8, was only reported by one laboratory for the subtyping of *stx1*. The EQA-provider specified in the invitation letter and in the submission protocol that when a participant signs up for a test and subsequently participates, all strains must be analysed using this test.

In the current EQA, there were 17 true errors in stx2 subtyping results, which is almost twice as many as in EQA-11. All errors were reported by four laboratories, and one laboratory incorrectly reported stx2 subtyping for 10 of the 12 test strains, accounting for 59% of the erroneous results. The EQA provider included the *stx2e* variant this year, which was correctly identified by 96% (23/24) of the participants. The incorrect result was reported by one laboratory using other methods than WGS.

Since the establishment of the currently accepted Stx subtype taxonomy in 2012, six additional Stx subtypes have been proposed, Stx1e, Stx2h, Stx2i, Stx2k, Stx2l, Stx2m, Stx2n and Stx2o [18], some of which have already been discussed by the EFSA BIOHAZ Panel in the EFSA report [7]. The EQA provider has developed a new protocol for detecting all new *stx* subtypes (unpublished).

4.3 Molecular typing-based cluster analysis

Twenty-three of the 26 laboratories (88%) conducted cluster analysis, all utilising WGS-derived data, with no laboratories submitting PFGE-derived data. This represents an increase from EQA-11, where 77% participated in cluster analysis.

4.3.1 WGS-derived data

Only one laboratory reported the use of external assistance for sequencing, and the majority (18/23) reported using an Illumina platform. All reported using commercial kits for preparing the library.

The EQA provider's QC evaluation of the raw reads submitted by the participants showed good-quality data; however, nine of 23 received warnings from the Bifrost QC pipeline. The contamination assessment part of Bifrost is based on Kraken [12], however, the EQA provider has disregarded any contamination warnings for strain3 or strain11, as similar results were observed for the EQA provider. *S. flexneri* and *E. coli* are highly genetically related and species identification using WGS is not always sufficient. Six laboratories received warnings as the average coverage was below 50. However, some argue that a threshold of 50 is too strict; suggesting that 30-40 would suffice, depending on the analysis [16].

As in previous years, the main QC parameters reported used by the participants in EQA-12 were a threshold of coverage and the checking of genus/species confirmation. The percentage of participants using assessment of the genome size has been above 71% since EQA-9 and confirmation of genus as a QC parameter above 91%.

In general, the performance of the cluster analysis was high, with 15 (65%) laboratories correctly identifying the cluster of closely related strains, which is lower compared to last year (80%). However, the analysis in EQA-12 was complicated by the introduction of the manipulated cluster sequences, strain20, contaminated by 8% *S. sonnei*, and a sequence of cluster strain19, which some participants found to be of lower quality than expected. Both were borderline QC issues that impacted the cluster identification.

Of the 23 laboratories, 20 (87%) reported using an allele-based method as the main analysis, and three (13%) reported using SNP analysis. None of the laboratories that used SNP-based analysis as the main method identified the predetermined cluster, for different reasons. From the additional analyses reported by other participants, the distances reported inside the cluster using SNP-based analyses (and identifying the correct cluster) were 0 or 0–4 (Reference-based method) or 0–46 (Assembly-based method, only one participant), showing a substantial variation depending on the method. Most (60%) of the laboratories using allele differences by cgMLST reported 0–3 inside the correct cluster.

When assessing the reported allele difference or SNP distances, the cgMLST approach showed more comparable results and, for most participants, a clear separation of the cluster and non-cluster strains. An exception was noted in the results from laboratories 135 and 138, which utilised allelic-based analysis with cgMLST only assembly-based. In these cases, strain19 exhibited high number of allelic differences in the participants' analysis (27AD or 30AD), reaching a similar level as the non-cluster strain16. Notably, one participant (135) employed the same tool, method, assembler, and schema as another participant (Laboratory 100). Similar results were observed for the other strains; however, they reported allelic differences for strain19 of only 7AD. The EQA-provider suspects that the reporting from Laboratory 135 may be a typing error. Laboratory 138 reported generally higher allele differences and was the only laboratory using a Shovill assembler.

For the laboratories able to identify the correct cluster, high similarity was observed for the reported cgMLST results based on Enterobase Scheme. Participants using SKESA as an assembler reported a lower number of allelic differences (0AD) compared to SPAdes (0–1 allelic differences) when employing only assembly-based allele calling.

SNP analyses can provide valid cluster detection, however, three laboratories utilising SNP as the main analysis reported different results. One laboratory did not use strain15 but, instead, employed an in-house standard as the reference for the SNP analysis. Interestingly, this same laboratory also included strain18 (with very low coverage) in the analysis. Two laboratories (138 and 187) have only recently started using WGS-derived data, and EQAs are a good way to test the progress of this transition. Laboratory 138 provided good-quality data; however, the analysis and evaluation need some adjustments. This emphasises the importance of understanding the pipeline and carefully evaluating the data. From the data visualised in Figure 9/10, there is a clear separation only for the cluster strains and the remaining strains for Laboratory 138, which successfully identified the cluster. Meanwhile, Laboratory 187 provided quality data with an average coverage below 50 but above 40, and their reporting of an incorrect cluster was a typographical mistake.

The submitted raw data showed that when employing a standardised cgMLST analysis, it is not uncommon to observe a random variation of one allele, even with high coverage (Figure 12). As observed in previous years, two participants (108 and 222) consistently deviated. These laboratories provided Ion Torrent data for which the EQA provider's analysis is not optimised, making correct assembly challenging. Therefore, the observed allelic differences (AD) may be artefacts of the method; however, the use of Ion Torrent data can complicate

communication and investigation of multi-country outbreaks when relying solely on the allelic method. None of them identified the pre-defined cluster, both for selecting a too low cut-off.

Strain17, a non- cluster sequence with low quality of R2. The provider's QC pipeline only accepts paired-end reads. Consequently, the quality of the R2 read is too low for proper analysis. Therefore, a single-end assembly (using R1) was employed in the cluster analysis by the EQA-provider.

The majority of the participants (18/23, 78%) reported the manipulated strain17 sequence as either QC-status B or C and 5 participants reported the sequence as having an acceptable quality (QC-status A) for further analysis. Strain17 only had low quality of R2 therefore a single-end assembly of R1 was useful for the cluster analysis. As many participants use assembly-based allele calling, this might be useful in the future.

All of the participants (100%) correctly reported quality issues of strain18, (a non-cluster sequence with reduced coverage and removal of genes), however four used the sequence for analysis and two of them reported it to be a part of the cluster,illustrating the importance of assessing the QC status of each genome before analysis. Conversely, in Strain20, a cluster sequence (strain14) contaminated with approximately 8% *S. sonnei*, only 48% (11/23) of participants reported it as either QC-status B or C, while the remaining 12 participants reported the sequence to have acceptable quality (QC-status A). The EQA provider anticipated that all participants would have recognised the contamination, however a low contamination does not influence the analysis much. In comparison, in EQA-11, where 85% (17/20) of participants correctly observed contamination (a non-cluster sequence contaminated with approximately 14% *E. albertii),* it is likely more straightforward for laboratories to identify 14% contamination with *E. albertii* compared to 8% *S. sonnei*. However, strain20 was a cluster strain, therefore influenced the reported cluster.

Since the EQA provider has included more difficult sequences compared to EQA-10, most laboratories took more time to assess the modified genomes. In general, participants described in detail what they observed, rather than merely following the previous suggestion to re-run the strain. It appears that the participants accepted the challenge, as advised by the contractor, and utilised the time to analyse the more questionable data, suggesting whether it was a cluster strain or not. However, the EQA provider acknowledges that the modified genomes included in this year's EQA influenced the identification of the cluster.

5 Conclusions

Twenty-six laboratories participated in the EQA-12 scheme, with 25 (96%) performing the serotyping part, 25 (96%) determining the virulence profile, and 23 (88%) engaging in cluster identification. Participation in the cluster analysis increased from EQA-11 (20/26, 77%). The serotyping part remained the same as in EQA-11. Similar to EQA-11, this EQA incorporated cluster analysis based on molecular typing, utilising exclusively WGS-derived data since no participants submitted PFGE data this year. The last instance of PFGE reporting was in EQA-10, indicating a permanent shift in STEC 'finger-printing' from PFGE to WGS among Member States.

O:H serotyping was performed by 88% (22/25) of the participants, achieving an average score of 97%. Similar to previous EQAs, participation in O grouping exceeded that in H typing. Consistent with prior EQAs, not all laboratories exhibited the ability to determine all O groups and H types. Generally, the more prevalent European serotypes generated higher scores compared to the less common ones, such as O154:H31 and O8:H4, which posed greater challenges in identification, especially when participants utilised phenotypic methods.

Once again, this year, the EQA provider included two other DEC pathotypes, EAEC (*aggR* gene), and ETEC (*esta* gene), testing the participating laboratories on their ability to detect STEC hybrid strains. The performance in detecting the *aggR* genes was high (23/24, 98%), surpassing EQA-11 where 95% correctly identified *aggR*. Similarly, the average performance score for *esta* was higher (98%) than in EQA-11 (89%). This performance discrepancy was attributed to four laboratories (108, 125, 128, and 187) that couldn't identify the gene in strain11. All laboratories except one utilised a WGS-based method.

Detection of the *eae* gene had high participation rates, and average scores through the EQAs has always been 96% or above (EQA-4: 96%; EQA-5: 98%; EQA-6: 97%; EQA-7: 98%; EQA-8: 96%; EQA-9: 99%, EQA-10: 98%, and EQA-11: 97%; and EQA-12: 96%).

Similarly to previous EQAs, the participation and average scores for *stx1* and *stx2* gene detection were high, with an average score of 97% for *stx1* and 98% for *stx2.* Subtyping of *stx1* and *stx2* is valuable since specific subtypes (*stx*2a) have been associated with increased risk of HUS, hospitalisation, or bloody diarrhoea respectively [8]. The high participation rate of 83% (20/24) is similar to EQA-11 (85%) which is still encouraging. The average score of laboratories that correctly performed the *stx* subtyping were 95% for *stx1*, 94% for *stx2*, and 93% combined *stx1* and *stx2*.

The incorporation of molecular typing-based cluster analysis in this EQA is up-to-date with the development of surveillance methods used by NPHRLs in Europe. Twenty-three laboratories performed the cluster analysis, which is three more than EQA-11, and all 23 used WGS-derived data. Notably, no laboratory employed PFGE for cluster analysis while participating in this EQA.

Modifying genomes have been the practice by the EQA-provider since EQA-10. As such, the strain sequence data were made accessible by the EQA provider, and participants were instructed to incorporate them into the cluster analysis while reporting characteristics and quality issues. Note, contaminations with a different species can be more challenging to identify than low-quality sequences. Unlike EQA-11, where most participants identified contamination (quality issues), in this EQA, only 48% of the participants reported issues with the sequence's quality. In general, the performance was high, with 15 (65%) laboratories correctly identifying the cluster of closely related strains. Two of the nine laboratories did not identify the correct cluster; they have recently started using WGS data. Additionally, two laboratories mistakenly overlooked the very low coverage of strain18. One laboratory failed to notice the 8% contamination in strain20, which might have influenced their decision to accept it despite detecting 8 SNPs in their analysis. Another laboratory's report appears to contain a typing error, and the last two were affected by the Ion Torrent user, potentially negatively impacting the complexity of the analysis. All in all, results are encouraging.

Further, of the 23 laboratories, 20 (87%) reported using an allele-based method as the main analysis and three (13%) reported using SNP analysis. The use of a standard cgMLST scheme (e.g. Enterobase) gives a very high degree of homogeneity in the results, and allele-based methods seem to be useful for inter-laboratory comparability and communication about cluster definitions. SNP analyses can also provide valid cluster detection at the national level; however, the analysis pipeline needs to be carefully assessed.

The current EQA scheme for typing STEC is the 12th EQA organised for laboratories in FWD-Net. The molecular surveillance system implemented as part of TESSy relies on the capacity of FWD-Net laboratories to produce analysable and comparable typing results into a central database. WGS-based typing for surveillance is increasingly used in the EU. Member States are asked to submit STEC WGS data in real-time to be accompanied by isolate metadata. ECDC coordinates centralised analysis of WGS STEC data when needed to support multi-country outbreak investigations.

6 Recommendations

6.1 Laboratories

Participants are encouraged to assign sufficient resources to repeat failed analysis if required to meet the deadline of submission.

Laboratories are expected to use each method as a stand-alone test, regardless of the results obtained in screening, detection, or any other test. Consequently, when a participant enrols in a test and actively participates, all strains must undergo testing using the specified method, such as the subtyping of *stx.*

6.2 ECDC and FWD-Net

ECDC is working actively with FWD-Net to improve the quality of sequence data generation and analysis through appropriate means like EQA schemes, expert exchange visits and workshops. ECDC encourages more participants to take part in the new molecular typing-based cluster analysis.

6.3 EQA provider

The assessment of the provided genome sequences yielded positive results, with almost all participants successfully identifying the modifications introduced by the EQA provider, particular for strain18 with low coverage. The exception was the contamination with only 8% *S. sonnei* in strain20 where 52% of the participants accepted the sequence as good quality. Consequently, in subsequent EQA rounds, the EQA provider will increase the contamination load, following the approach employed in previous EQAs (e.g. introducing 14% contamination with *E. albertii* in EQA-11). Furthermore, the EQA provider plans to continue and expand this modification aspect of the EQA to challenge participants in evaluating poor-quality genomes and those with contamination.

This expanded approach aims to underscore the importance of assessing genomes even in the presence of low-level contamination or other quality issues. However, it is important to approach such assessments with the utmost caution.

The EQA provider suggests an open 'cut-off' discussion of STEC clusters for WGS analyses with the FWD-Network

Based on the feedback survey, it is evident that participants appreciate the assessment of the quality control (QC) of the submitted sequences. However, it is recommended that the EQA provider places special emphasis on scrutinising IonTorrent-produced data, as indicated by one laboratory. In terms of data reporting, two laboratories provided feedback, suggesting the streamlining of the reporting form for improved user-friendliness, especially in tasks such as virulence gene determination, such as *stx* subtyping. Additionally, it is recommended to explore the feasibility of sending isolates for multiple EQAs simultaneously during planned sequencing runs.

References

1. European Centre for Disease Control Prevention and Control (ECDC). Surveillance report. STEC Infection, Annual Epidemiology Report for 2022. Available at: https://www.ecdc.europa.eu/en/publications-data/stec-infection-annual-epidemiological-report-2022
2. European Parliament and 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. Strasbourg: European Parliament and European Council; 2004. Available at: http://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:32004R0851
3. European Parliament and European Council. Decision No 1082/2013/EU of the European Parliament and of the Council of 23 November 2022 on serious cross-border threats to health and repealing Decision No 1082/2013/EC (Text with EEA relevance). Strasbourg: European Parliament and European Council; 2013. Available at: http://publications.europa.eu/en/publication-detail/-/publication/8d817a1f-45fa-11e3-ae03-01aa75ed71a1
4. Bielaszewska M, Friedrich AW, Aldick T, Schürk-Bulgrin R, Karch H. Shiga Toxin Activatable by Intestinal Mucus in *Escherichia coli* Isolated from Humans: Predictor for a Severe Clinical Outcome. Clin Infect Dis. 2006 Nov 1;43(9):1160-7.
5. Friedrich AW, Bielaszewska M, Zhang WL, Pulz M, Kuczius T, Ammon A, et al. *Escherichia coli* Harboring Shiga Toxin 2 Gene Variants: Frequency and Association with Clinical Symptoms. J Infect Dis. 2002 Jan 1;185(1):74-84.
6. Persson S, Olsen KE, Ethelberg S, Scheutz F. Subtyping Method for *Escherichia coli* Shiga Toxin (Verocytotoxin) 2 Variants and Correlations to Clinical Manifestations. J Clin Microbiol. 2007 Jun;45(6):2020-4.
7. EFSA BIOHAZ Panel, Koutsoumanis K, Allende A, Alvarez-Ordóñez A, Bover-Cid S, Chemaly M, et al. Pathogenicity assessment of Shiga toxin-producing *Escherichia coli* (STEC) and the public health risk posed by contamination of food with STEC. EFSA Journal. 2020;18(1):5967, 105 pp. https://doi.org/10.2903/j.efsa.2020.5967
8. Scheutz F, Teel LD, Beutin L, Piérard D, Buvens G, Karch H, et al. Multicenter Evaluation of a Sequence-Based Protocol for Subtyping Shiga Toxins and Standardizing Stx Nomenclature. J Clin Microbiol. 2012 Sep;50(9):2951-63.
9. Boisen N, Østerlund M T, Joensen K G, Santiago A E, Mandomando I, Cravioto A el at. 2020, Redefining enteroaggregative *Escherichia coli* (EAEC): Genomic characterization of epidemiological EAEC strains, PloS Negl Trop Dis. 2020 Sep 8;14(9):e0008613. Doi: 10.1371/journal.pntd.0008613. eCollection 2020 Sep.
10. International Organization for Standardization (ISO). ISO/IEC 17043:2010 – Conformity assessment – General requirements for proficiency testing. Vernier: ISO; 2010. Available at: http://www.iso.org/iso/catalogue_detail.htm?csnumber=29366
11. Warwick Medical School. EnteroBase. Coventry: University of Warwick; 2018. Available at: http://enterobase.warwick.ac.uk
12. 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.
13. Scheutz F, Fruth A, Cheasty T, Tschäpe H. Appendix 1 – O Grouping: Standard Operation Procedure (O SOP) and Appendix 2: and H Determination: Standard Operation Procedure (H SOP) – *Escherichia coli* O antigen grouping and H antigen determination. Copenhagen: Statens Serum Institut; 2002. Available at: http://www.ssi.dk/English/HealthdataandICT/National%20Reference%20Laboratories/Bacteria/~/media/49802860CB5E44D6A373E6116ABBDC0D.ashx
14. Scheutz F, Morabito S, Tozzoli R, Caprioli A. Identification of three vtx1 and seven vtx2 subtypes of verocytotoxin encoding genes of *Escherichia coli* by conventional PCR amplification. Copenhagen: Statens Serum Institut; 2002.
15. European Centre for Disease Prevention and Control (ECDC). Surveillance of National Reference Laboratory (NRL) capacity for six food- and waterborne diseases in EU/EEA countries – Campylobacteriosis, listeriosis, salmonellosis, Shiga toxin/verocytotoxin–producing *Escherichia coli* (STEC/VTEC), shigellosis and yersiniosis. Stockholm: ECDC; 2012. Available at: http://ecdc.europa.eu/publications-data/survey-national-reference-laboratory-capacity-six-fwd-eueea-countries
16. Statens Serum Institut (SSI). Bifrost_QC [Internet; software package]. Copenhagen: SSI; 2019. Available at: https://github.com/ssi-dk/bifrost
17. Projahn M, Lamparter MC, Ganas P, Goehler A, Lorenz-Wright SC, Maede D, Fruth A, Lang C, Schuh E. Genetic diversity and pathogenic potential of Shiga toxin-producing *Escherichia coli* (STEC) derived from German flour. Int J Food Microbiol. 2021 Jun 2;347:109197. Doi: 10.1016/j.ijfoodmicro.2021.109197. Epub 2021 Apr 20.PMID: 33895597.
18. Boisen N, Østerlund M T, Joensen K G, Santiago A E, Mandomando I, Cravioto A el at. 2020, Redefining enteroaggregative *Escherichia coli* (EAEC): Genomic characterization of epidemiological EAEC strains, PloS Negl Trop Dis. 2020 Sep 8;14(9):e0008613. Doi: 10.1371/journal.pntd.0008613. eCollection 2020 Sep.

Annex 1. List of participants

Country	Laboratory	National institute
Austria	Austrian Reference Center for *Escherichia coli* including VTEC	AGES, Institute for Medical Microbiology and Hygiene, Graz
Belgium	National Reference Centre STEC	UZ Brussel
Bulgaria	NRL for Enteric Diseases	National Center of Infectious and Parasitic Diseases
Croatia	NRL for Salmonellae	Croatian Institute of Public Health
Czechia	National Reference Laboratory for *Escherichia coli* and Shigellae	National Institute of Public Health
Denmark	Laboratory of Gastrointestinal Bacteria	Statens Serum Institut
Estonia	Laboratory of Communicable diseases	Health Board
Finland	Expert Microbiology Unit	Finnish Institute for Health and Welfare
France	CNR E. coli – Institut Pasteur – APHP	Institut Pasteur – CHU Robert Debré -
Germany	NRC Salmonella and other Bacterial Enterics	Robert Koch Institute
Greece	Reference centre for salmonella, shigella, listeria, VTEC	University of West Attica
Hungary	FWD Reference Laboratory	National Center for Public Health and Pharmacy
Iceland	Department of Microbiology	Landspitali University Hospital
Ireland	Public Health Laboratory Dublin	HSE
Italy	Microbiological food Safety and Foodborne Disease Unit	Istituto Superiore di Sanitá
Latvia	Laboratory Service, National Microbiology Reference Laboratory	Riga East University Hospital
Luxembourg	Epidemiology and Microbial Genomics (EPIGEM)	Laboratoire National de Santé
Montenegro	Centre for Medical Microbiology-Department of Sanitary Microbiology	Institute for Public Health of Montenegro
The Netherlands	IDS	RIVM
Norway	National Reference Laboratory for Enteropathogenic Bacteria	Norwegian Institute of Public Health
Poland	Laboratory of Bacteriology and Biocontamination Control	National Institute of Public Health NIH-NRI
Portugal	URGI	National Institute of Health Dr. Ricardo Jorge
Romania	Molecular Epidemiology for Communicable Diseases	Cantacuzino National Military Medical Institute for Research and Development
Slovenia	Department for Public Health Microbiology	National Laboratory of Health, Environment and Food
Spain	Laboratorio de Referencia e Investigación en Enfermedades Transmitidas por Agua y Alimentos	Instituto de Salud Carlos III
Sweden	Unit for Laboratory Surveillance of Bacterial Pathogens	Public Health Agency of Sweden

Annex 2. Participation overview EQA-11/-12

		2021-2022 (EQA-11)			2022-2023 (EQA-12)			
				Cluster				Cluster
Lab. number	Participation (min. 1 part)	Serotyping	Virulence	WGS	Participation (min. 1 part)	Serotyping	Virulence	WGS
19	x	x	x	x	x	x	x	x
34	x	x	x	x	x	x	x	x
80	x	x	x	x	x	x	x	x
88	X	x	x	x	x	x	x	x
90	X		x	x	x		x	x
100	X	x	x	x	x	x	x	x
108	x	x	x	x	x	x	x	x
123	x	x	x	x	x	x	x	x
124	x	x	x	x	x	x	x	x
127	x	x	x		x	x	x	x
128	x	x	x		x	x	x	x
129	x	x	x	x	x	x	x	x
130	x	x	x		x	x	x	
131	x	x	x	x	x	x	x	x
132	x	x	x	x	x	x	x	x
133	x	x	x	x	x	x	x	x
134	x	x	x	x	x	x	x	x
135	x	x	x	x	x	x	x	x
136	x	x	x	x	x	x	x	x
138	x	x	x	x	x	x	x	x
139	x	x	x	x	x	x	x	x
145*	x	x	x					
153	x	x	x	x	x	x	x	x
187#					x	x	x	x
222	x	x	x	x	x	x	x	x
230	x	x	x		x	x	x	
240	x	x			x	x		
Number of participants	**26**	**25**	**25**	**20**	**26**	**25**	**25**	**23**

** = Laboratory did not participate in EQA-12*
#= Laboratory did not participate in EQA-11

Annex 3. Serotyping result scores

O group

	Strain number												
Lab. number	1	2	3	4	5	6	7	8	9	10	11	12	
EQA	O146	O103	O111	O80	O145	O157	O128	O154	O26	O80	O8	O80	Method
19	146	103	111	80	145	157	128	154	26	80	8	80	A
34	146	103	111	80	145	157	128	154	26	80	8	80	C
80	146	103	111	80	145	157	128	154	26	80	8	80	C
88	146	103	111	80	145	157	128	154	26	80	8	80	C
100	146	103	111	80	145	157	128	154	26	80	8	80	A
108	146	103	111	80	145	157	128	154	26	80	8	80	C
123	146	103	111	80	145	157	128	154	26	80	8	80	C
124	146	103	111	80	145	157	128	154	26	80	8	80	C
125	NT	103	111	R	145	157	128	R	26	55	127	86	A
127	146	103	111	80	145	157	128	NT	26	80	8	80	A
128	146	103	111	80	145	157	128	154	26	80	8	80	B
129	146	103	111	80	145	157	128	154	26	80	8	180	C
130	ND	ND	ND	ND	ND	157	ND	ND	ND	ND	ND	ND	B
131	146	103	111	80	145	157	128	154	26	91	NT	80	A
132	146	103	111	80	145	157	128	154	26	80	80	80	C
133	146	103	111	80	145	157	128	154	26	80	8	80	C
134	146	103	111	80	145	157	128	154	26	80	8	80	C
135	146	103	111	80	145	157	128	154	26	80	8	80	C
136	146	128	111	80	145	157	103	154	26	80	8	80	C
138	146	103	111	80	145	157	128	154	26	80	8	80	C
139	146	103	111	80	145	157	128	154	26	80	8	80	C
153	146	103	111	80	145	157	128	154	26	80	8	80	A
187	146	103	111	80	145	157	128	154	26	80	8	80	C
222	146	103	111	80	145	157	128	154	26	80	8	80	C
240	ND	103	157	ND	145	157	ND	ND	26	ND	103	ND	A

n=25 participants
Purple shading: incorrect result
ND: not done
A: phenotypic serotyping, B: PCR-based serotyping, C: WGS-based serotyping
NT: non-typable
R:Rough:

H type

	Strain number												
Lab. number	1	2	3	4	5	6	7	8	9	10	11	12	
EQA	H28	H2	H21	H2	H-/H28	H-/H7	H2	H31	H11	H2	H4	H2	Method
19	28	2	21	2	H-	H-	2	31	11	2	4	2	A
34	28	2	21	2	28	7	2	31	11	2	4	2	C
80	28	2	21	2	28	7	2	31	11	2	4	2	C
88	28	2	21	2	28	7	2	31	11	2	4	2	C
100	28	2	21	2	28	7	2	31	11	2	4	2	C
108	28	2	21	2	28	7	2	31	11	2	4	2	C
123	28	2	21	2	28	7	2	31	11	2	4	2	C
124	28	2	21	2	28	7	2	31	11	2	4	2	C
127	28	2	21	2	28	7	2	31	11	2	4	2	C
128	28	2	21	2	28	7	2	31	11	2	4	2	C
129	28	2	21	2	28	7	2	31	11	2	4	2	C
131	28	2	21	2	H-	7	2	H-	11	14	4	2	A
132	28	2	21	2	28	7	2	31	11	2	4	2	C
133	28	2	21	2	28	7	2	31	11	2	4	2	C
134	28	2	21	2	28	7	2	31	11	2	4	2	C
135	28	2	21	2	28	7	2	31	11	2	4	2	C
136	28	2	21	2	28	7	2	31	11	2	4	2	C
138	28	2	21	2	H-	7	2	31	11	2	4	2	C
139	28	2	21	2	28	7	2	31	11	2	4	2	C
153	28	2	21	2	9	7	2	31	11	2	4	2	C
187	28	2	21	2	28	7	2	31	11	2	4	2	C
222	28	2	21	2	28	7	2	31	11	2	4	2	C

n=22 participants

Purple shading: incorrect result

A: phenotypic serotyping, B: PCR-based serotyping, C: WGS-based serotyping

Some H- results was accepted as correct results (Strain5, Strain6), when the EQA provider observed a tendency to be H- more than one during testing.

Annex 4. Virulence profiles result scores

Detection of *aggR*

	Strain number												
Lab. number	1	2	3	4	5	6	7	8	9	10	11	12	
EQA	-	-	+	-	-	-	-	-	-	-	-	-	Method
19	-	-	+	-	-	-	-	-	-	-	-	-	A
34	-	-	+	-	-	-	-	-	-	-	-	-	B
80	-	-	+	-	-	-	-	-	-	-	-	-	B
88	-	-	+	-	-	-	-	-	-	-	-	-	A
90	-	-	+	-	-	-	-	-	-	-	-	-	A
100	-	-	+	-	-	-	-	-	-	-	-	-	A
108	-	-	+	-	-	-	-	-	-	-	-	-	B
123	-	-	+	-	-	-	-	-	-	-	-	-	B
124	-	-	+	-	-	-	-	-	-	-	-	-	B
125	-	-	-	-	+	-	-	-	+	-	-	-	A
127	-	-	+	-	-	-	-	-	-	-	-	-	A
128	-	-	+	-	-	-	-	-	-	-	-	-	B
129	-	-	+	-	-	-	-	-	-	-	-	-	B
130	-	-	-	-	-	-	-	+	-	-	-	-	A
131	-	-	+	-	-	-	-	-	-	-	-	-	A
132	-	-	+	-	-	-	-	-	-	-	-	-	B
133	-	-	+	-	-	-	-	-	-	-	-	-	B
134	-	-	+	-	-	-	-	-	-	-	-	-	B
135	-	-	+	-	-	-	-	-	-	-	-	-	B
136	-	-	+	-	-	-	-	-	-	-	-	-	B
138	-	-	+	-	-	-	-	-	-	-	-	-	A
139	-	-	+	-	-	-	-	-	-	-	-	-	B
153	-	-	+	-	-	-	-	-	-	-	-	-	B
187	-	-	+	-	-	-	-	-	-	-	-	-	B
222	-	-	+	-	-	-	-	-	-	-	-	-	B

n=25 participants *A: Other than WGS, B: WGS-based*

Purple shading: incorrect result

Detection of *eae*

	Strain number												
Lab. number	1	2	3	4	5	6	7	8	9	10	11	12	
EQA	-	+	-	+	+	+	-	-	+	+	-	+	Method
19	-	+	-	+	+	+	-	-	+	+	-	+	A
34	-	+	-	+	+	+	-	-	+	+	-	+	B
80	-	+	-	+	+	+	-	-	+	+	-	+	B
88	-	+	-	+	+	+	-	-	+	+	-	+	A
90	-	+	-	+	+	+	-	-	+	+	-	+	A
100	-	+	-	+	+	+	-	-	+	+	-	+	A
108	-	+	-	+	+	+	-	-	+	+	-	+	B
123	-	+	-	+	+	+	-	-	+	+	-	+	B
124	-	+	-	+	+	+	-	-	+	+	-	+	B
125	-	+	-	-	+	-	-	-	-	-	-	-	A
127	-	+	-	+	+	+	-	-	+	+	-	+	A
128	-	+	-	+	+	+	-	-	+	+	-	+	B
129	-	+	-	+	+	+	-	-	+	+	-	+	B
130	-	+	-	-	+	+	-	-	+	-	-	-	A
131	-	+	-	-	+	+	-	-	+	+	-	+	A
132	-	+	-	+	+	+	-	-	+	+	-	+	B
133	-	+	-	+	+	+	-	-	+	+	-	+	B
134	-	+	-	+	+	+	-	-	+	+	-	+	B
135	-	+	-	+	+	+	-	-	+	+	-	+	B
136	-	-	-	+	+	+	+	-	+	+	-	+	B
138	+	+	-	+	+	+	-	-	+	+	-	+	A
139	-	+	-	+	+	+	-	-	+	+	-	+	B
153	-	-	-	+	+	+	-	-	+	+	-	+	B
187	-	+	-	+	+	+	-	-	+	+	-	+	B
222	-	+	-	+	+	+	-	-	+	+	-	+	B

n=25 participants *A: Other than WGS, B: WGS-based*

Purple shading: incorrect result

Detection of *esta*

	Strain number												
Lab. number	1	2	3	4	5	6	7	8	9	10	11	12	
EQA	-	-	-	-	-	-	-	-	-	-	+	-	Method
19	-	-	-	-	-	-	-	-	-	-	+	-	A
34	-	-	-	-	-	-	-	-	-	-	+	-	B
80	-	-	-	-	-	-	-	-	-	-	+	-	B
88	-	-	-	-	-	-	-	-	-	-	+	-	A
90	-	-	-	-	-	-	-	-	-	-	+	-	A
100	-	-	-	-	-	-	-	-	-	-	+	-	A
108	-	-	-	-	-	-	-	-	-	-	-	-	B
123	-	-	-	-	-	-	-	-	-	-	+	-	B
124	-	-	-	-	-	-	-	-	-	-	+	-	B
125	-	-	-	-	-	-	-	-	-	-	-	-	A
127	-	-	-	-	-	-	-	-	-	-	+	-	A
128	-	-	-	-	-	-	-	-	-	-	-	-	B
129	-	-	-	-	-	-	-	-	-	-	+	-	B
131	-	-	-	-	-	-	-	-	-	-	+	-	A
132	-	-	-	-	-	-	-	-	-	-	+	-	B
133	-	-	-	-	-	-	-	-	-	-	+	-	B
134	-	-	-	-	-	-	-	-	-	-	+	-	B
135	-	-	-	-	-	-	-	-	-	-	+	-	B
136	-	-	-	-	-	-	-	-	-	-	+	-	B
138	-	-	-	-	-	-	-	-	-	-	+	-	A
139	-	-	-	-	-	-	-	-	-	-	+	-	B
153	-	-	-	-	-	-	-	-	-	-	+	-	B
187	+	-	+	-	-	+	-	-	-	-	-	-	B
222	-	-	-	-	-	-	-	-	-	-	+	-	B

n=24 participants *A: Other than WGS, B: WGS-based*
Purple shading: incorrect result

Detection of *stx1*

	Strain number												
Lab. number	1	2	3	4	5	6	7	8	9	10	11	12	
EQA	-	+	-	-	-	-	+	+	+	-	-	-	Method
19	-	+	-	-	-	-	+	+	+	-	-	-	A
34	-	+	-	-	-	-	+	+	+	-	-	-	B
80	-	+	-	-	-	-	+	+	+	-	-	-	B
88	-	+	-	-	-	-	+	+	+	-	-	-	A
90	-	+	-	-	-	-	+	+	+	-	-	-	A
100	-	+	-	-	-	-	+	+	+	-	-	-	A
108	-	+	-	-	-	-	+	+	+	-	-	-	B
123	-	+	-	-	-	-	+	+	+	-	-	-	B
124	-	+	-	-	-	-	+	+	+	-	-	-	B
125	+	+	-	+	+	+	+	+	-	-	+	+	A
127	-	+	-	-	-	-	+	+	+	-	-	-	A
128	-	+	-	-	-	-	+	+	+	-	-	-	B
129	-	+	-	-	-	-	+	+	+	-	-	-	B
130	-	+	-	-	-	-	+	-	+	-	-	-	A
131	-	+	-	-	-	-	+	+	+	-	-	-	A
132	-	+	-	-	-	-	+	+	+	-	-	-	B
133	-	+	-	-	-	-	+	+	+	-	-	-	B
134	-	+	-	-	-	-	+	+	+	-	-	-	B
135	-	+	-	-	-	-	+	+	+	-	-	-	B
136	-	+	-	-	-	-	+	+	+	-	-	-	B
138	-	+	-	-	-	-	+	+	+	-	-	-	A
139	-	+	-	-	-	-	+	+	+	-	-	-	B
153	-	+	-	-	-	-	+	+	+	-	-	-	B
187	-	+	-	-	-	-	+	+	+	-	-	-	B
222	-	+	-	-	-	-	+	+	+	-	-	-	B

n=25 participants *A: Other than WGS*
Purple shading: incorrect result *B: WGS-based*

Detection of *stx2*

	Strain number												
Lab. number	1	2	3	4	5	6	7	8	9	10	11	12	
EQA	+	+	+	+	+	+	+	-	+	+	+	+	Method
19	+	+	+	+	+	+	+	-	+	+	+	+	A
34	+	+	+	+	+	+	+	-	+	+	+	+	B
80	+	+	+	+	+	+	+	-	+	+	+	+	B
88	+	+	+	+	+	+	+	-	+	+	+	+	A
90	+	+	+	+	+	+	+	-	+	+	+	+	A
100	+	+	+	+	+	+	+	-	+	+	+	+	A
108	+	+	+	+	+	+	+	-	+	+	+	+	B
123	+	+	+	+	+	+	+	-	+	+	+	+	B
124	+	+	+	+	+	+	+	-	+	+	+	+	B
125	+	-	-	-	+	+	+	+	+	-	+	-	A
127	+	+	+	+	+	+	+	-	+	+	+	+	A
128	+	+	+	+	+	+	+	-	+	+	+	+	B
129	+	+	+	+	+	+	+	-	+	+	+	+	B
130	+	+	+	+	+	+	+	-	+	+	+	+	A
131	+	+	+	+	+	+	+	-	+	+	+	+	A
132	+	+	+	+	+	+	+	-	+	+	+	+	B
133	+	+	+	+	+	+	+	-	+	+	+	+	B
134	+	+	+	+	+	+	+	-	+	+	+	+	B
135	+	+	+	+	+	+	+	-	+	+	+	+	B
136	+	+	+	+	+	-	+	-	+	+	+	+	B
138	+	+	+	+	+	+	+	-	+	+	+	+	A
139	+	+	+	+	+	+	+	-	+	+	+	+	B
153	+	+	+	+	+	+	+	-	+	+	+	+	B
187	+	+	+	+	+	+	+	-	+	+	+	+	B
222	+	+	+	+	+	+	+	-	+	+	+	+	B

n=25 participants
Purple shading: incorrect result
A: Other than WGS
B: WGS-based

stx subtyping

stx1

	Strain number												
Lab. number	1	2	3	4	5	6	7	8	9	10	11	12	
EQA	-	stx1a	-	-	-	-	stx1c	stx1d	stx1a	-	-	-	Method
19	-	stx1a	-	-	-	-	stx1c	stx1d	stx1a	-	-	-	A
34	-	stx1a	-	-	-	-	stx1c	stx1d	stx1a	-	-	-	B
80	-	stx1a	-	-	-	-	stx1c	stx1d	stx1a	-	-	-	B
88	-	stx1a	-	-	-	-	stx1c	stx1d	stx1a	-	-	-	A
90	-	stx1a	-	-	-	-	stx1c	stx1d	stx1a	-	-	-	A
100	-	stx1a	-	-	-	-	stx1c	stx1d	stx1a	-	-	-	A
108	-	stx1a	-	-	-	-	stx1c	stx1d	stx1a	-	-	-	B
123	-	stx1a	-	-	-	-	stx1c	stx1d	stx1a	-	-	-	B
124	-	stx1a	-	-	-	-	stx1c	stx1d	stx1a	-	-	-	B
125	ND	ND	ND	ND	ND	ND	ND	ND	ND	-	ND	ND	A
127	-	stx1a	-	-	-	-	stx1c	stx1d	stx1a	-	-	-	A
128	-	stx1a	-	-	-	-	stx1c	stx1d	stx1a	-	-	-	B
129	-	stx1a	-	-	-	-	stx1c	stx1d	stx1a	-	-	-	B
131	-	stx1a	-	-	-	-	stx1c	stx1d	stx1a	-	-	-	A
132	-	stx1a	-	-	-	-	stx1c	stx1d	stx1a	-	-	-	B
133	-	stx1a	-	-	-	-	stx1c	stx1d	stx1a	-	-	-	B
134	-	stx1a	-	-	-	-	stx1c	stx1d	stx1a	-	-	-	B
135	-	stx1a	-	-	-	-	stx1c	stx1d	stx1a	-	-	-	B
136	-	stx1c	-	-	-	-	stx1a	stx1d	stx1a	-	-	-	B
138	-	stx1a	-	-	-	-	stx1c	stx1d	stx1a	-	-	-	A
139	-	stx1a	-	-	-	-	stx1c	stx1d	stx1a	-	-	-	B
153	-	stx1a	-	-	-	-	stx1c	stx1d	stx1a	-	-	-	B
187	-	stx1a	-	-	-	-	stx1c	stx1d	stx1a	-	-	-	B
222	-	stx1a	-	-	-	-	stx1c	stx1d	stx1a	-	-	-	B

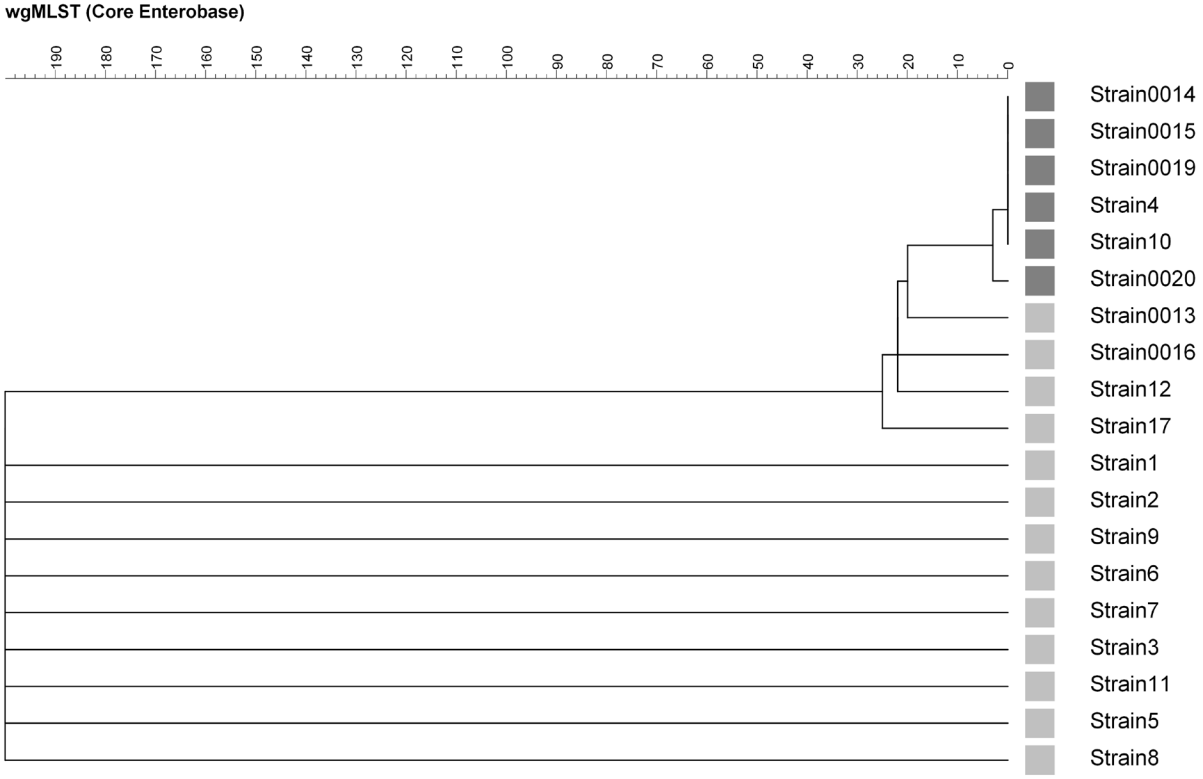
n=24 participants
Purple shading: incorrect result
ND: not done
A: Other than WGS
B: WGS-based

stx2

	Strain number												
Lab. number	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	
EQA	**stx2b**	**stx2d**	**stx2a**	**stx2a**	**stx2a**	**stx2c**	**stx2b**	**-**	**stx2a**	**stx2a**	**stx2e**	**stx2a**	**Method**
19	stx2b	stx2d	stx2a	stx2a	stx2a	stx2c	stx2b	-	stx2a	stx2a	stx2e	stx2a	A
34	stx2b	stx2d	stx2a	stx2a	stx2a	stx2c	stx2b	-	stx2a	stx2a	stx2e	stx2a	B
80	stx2b	stx2d	stx2a	stx2a	stx2a	stx2c	stx2b	-	stx2a	stx2a	stx2e	stx2a	B
88	stx2b	stx2d	stx2a	stx2a	stx2a	stx2c	stx2b	-	stx2a	stx2a	stx2e	stx2a	A
90	stx2b	stx2d	stx2a	stx2a	stx2a	stx2c	stx2b	-	stx2a	stx2a	stx2e	stx2a	A
100	stx2b	stx2d	stx2a	stx2a	stx2a	stx2c	stx2b	-	stx2a	stx2a	stx2e	stx2a	A
108	stx2b	stx2d	stx2a	stx2a	stx2a	stx2c	stx2b	-	stx2a	stx2a	stx2e	stx2a	B
123	stx2b	stx2d	stx2a	stx2a	stx2a	stx2c	stx2b	-	stx2a	stx2a	stx2e	stx2a	B
124	stx2b	stx2d	stx2a	stx2a	stx2a	stx2c	stx2b	-	stx2a	stx2a	stx2e	stx2a	B
125	stx2c; stx2d	stx2d	stx2g	stx2d	stx2g	stx2d	stx2d	-	stx2d	-	stx2d	-	A
127	stx2b	stx2d	stx2a	stx2a	stx2a	stx2c	stx2b	-	stx2a	stx2a	stx2e	stx2a	A
128	stx2b	stx2c; stx2d	stx2a	stx2a; stx2c	stx2a	stx2c	stx2b	-	stx2a	stx2a; stx2c	stx2e	stx2a	B
129	stx2b	stx2d	stx2a	stx2a	stx2a	stx2c	stx2b	-	stx2a	stx2a	stx2e	stx2a	B
131	stx2b	stx2d	stx2a	stx2a	stx2a	stx2c	stx2b	-	stx2a	stx2a	stx2e	stx2a	A
132	stx2b	stx2d	stx2a	stx2a	stx2a	stx2c	stx2b	-	stx2a	stx2a	stx2e	stx2a	B
133	stx2b	stx2d	stx2a	stx2a	stx2a	stx2c	stx2b	-	stx2a	stx2a	stx2e	stx2a	B
134	stx2b	stx2d	stx2a	stx2a	stx2a	stx2c	stx2b	-	stx2a	stx2a	stx2e	stx2a	B
135	stx2b	stx2d	stx2a	stx2a	stx2a	stx2c	stx2b	-	stx2a	stx2a	stx2e	stx2a	B
136	stx2b	stx2b	stx2a	stx2a	stx2a	-	stx2d	-	stx2a	stx2a	stx2e	stx2a	B
138	stx2b	stx2d	stx2a	stx2a	stx2a	stx2c	stx2b	-	stx2a	stx2a	stx2e	stx2a	A
139	stx2b	stx2d	stx2a	stx2a	stx2a	stx2c	stx2b	-	stx2a	stx2a	stx2e	stx2a	B
153	stx2b	stx2d	stx2a	stx2a	stx2a	stx2c	stx2b	-	stx2a	stx2a	stx2e	stx2a	B
187	stx2b	stx2d	stx2a	stx2a	stx2a	stx2c	stx2b	-	stx2a	stx2a	stx2e	stx2a	B
222	stx2b	stx2c	stx2a	stx2a	stx2a	stx2c	stx2b	-	stx2a	stx2a	stx2e	stx2a	B

n=24 participants *A: Other than WGS, B: WGS-based*
Purple shading: incorrect result *-: negative for stx2*

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



Single linked dendrogram of core genome multilocus sequence typing (cgMLST) profiles of STEC EQA-12 strains (cgMLST, EnteroBase, http://enterobase.warwick.ac.uk).
Analysed in BioNumerics: maximum distance of 200 exceeded, results clipped.
Cluster strains: dark grey, outside cluster strains: light grey.
Strain4 and Strain10 are technical duplicates.

Annex 6. Reported sequencing details

Laboratory	Sequencing performed	Protocol (library prep)	Commercial kit	Sequencing platform
19	In own laboratory	Commercial kits	Nextera XT Kit (Illumina)	NextSeq
34	In own laboratory	Commercial kits	NEBNext ULTRA II FS	NextSeq
80	In own laboratory	Commercial kits	Kapa HyperPlus, Kapa Biosystems	NextSeq
88	In own laboratory	Commercial kits	Nextera XT DNA Library Preparation kit (Illumina)	NextSeq
90	In own laboratory	Commercial kits	Nextera XT DNA Library Preparation Kit	MiSeq
100	In own laboratory	Commercial kits	Illumina DNA Prep	MiSeq
108	In own laboratory	Commercial kits	Ion Xpress™ Plus Fragment Library Kit for AB Library Builder™ System*	Ion S5 XL system
123	In own laboratory	Commercial kits	Nextera XT Library Prep Kit (Illumina)**	MiSeq
124	In own laboratory	Commercial kits	KAPA HyperPlus Kit	NovaSeq 6000
127	In own laboratory	Commercial kits	Nextera XT DNA Library preparation kit	MiSeq
128	Externally	Commercial kits	Illumina DNA Prep	NextSeq
129	In own laboratory	Commercial kits	Illumina Nextera XT**	MiSeq
131	In own laboratory	Commercial kits	DNA Prep	NextSeq
132	In own laboratory	Commercial kits	Illumina DNA prep	MiSeq
133	In own laboratory	Commercial kits	Illumina DNA prep	MiSeq
134	In own laboratory	Commercial kits	DNA Prep Illumina**	MiniSeq Illumina
135	In own laboratory	Commercial kits	Illumina DNA prep	NextSeq
136	In own laboratory	Commercial kits	Illumina DNA PREP	NextSeq
138	In own laboratory	Commercial kits	Illumina DNAPrep	NextSeq
139	In own laboratory	Commercial kits	Nextera XT	NextSeq
153	In own laboratory	Commercial kits	Illumina DNA Prep	MiSeq
187	In own laboratory	Commercial kits	Illumina dna prep	NextSeq
222	In own laboratory	Commercial kits	NEBNext® Fast DNA Fragmentation & Library Prep Set for Ion Torrent, New England Biolabs***	Ion GeneStudio S5 System

**: No PCR amplification of libraries prior template preparation on Ion Chef instrument*
***: Adjusted volume of reagents.*
****: Decreased shearing time*

5

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

Laboratory	Reported cluster	Corresponding to EQA provider strains	Correct
Provider		Strain4, Strain10, Strain14, Strain15, Strain19, (Strain20)	Yes
19	9842, 9796, 0014, 0015, 0019, 0020	Strain4, Strain10, Strain14, Strain15, Strain19, Strain20	Yes
34	9471, 9882, 0014, 0015, 0019, 0020	Strain4, Strain10, Strain14, Strain15, Strain19, Strain20	Yes
80	9913, 9526, 0014, 0015, 0019, 0020	Strain4, Strain10, Strain14, Strain15, Strain19, Strain20	Yes
88	9865, 9759, 0014, 0015, 0019, 0020	Strain4, Strain10, Strain14, Strain15, Strain19, Strain20	Yes
90	9168, 9791, 0014, 0015, 0019, 0020	Strain4, Strain10, Strain14, Strain15, Strain19, Strain20	Yes
100	9230, 9289, 0014, 0015, 0019, 0020	Strain4, Strain10, Strain14, Strain15, Strain19, Strain20	Yes
108	9347, 9054, 0015, 0019	Strain4, Strain10, Strain15, Strain19	No
123	9258, 9983, 0014, 0015, 0018, 0019, 0020	Strain4, Strain10, Strain14, Strain15, Strain18, Strain19, Strain20	No
124	9067, 9830, 0014, 0015, 0019	Strain4, Strain10, Strain14, Strain15, Strain19	(Yes)
127	9006, 9878, 0014, 0015, 0019	Strain4, Strain10, Strain14, Strain15, Strain19	(Yes)
128	9338, 9467, 0014, 0015, 0019, 0020	Strain4, Strain10, Strain14, Strain15, Strain19, Strain20	Yes
129	9091, 9110, 0014, 0015, 0019, 0020	Strain4, Strain10, Strain14, Strain15, Strain19, Strain20	Yes
131	9008, 9615, 0014, 0015, 0019, 0020	Strain4, Strain10, Strain14, Strain15, Strain19, Strain20	Yes
132	9014, 9595, 0014, 0015, 0018, 0019, 0020	Strain4, Strain10, Strain14, Strain15, Strain18, Strain19, Strain20	No
133	9036, 9162, 0014, 0015, 0019, 0020	Strain4, Strain10, Strain14, Strain15, Strain19, Strain20	Yes
134	9727, 9443, 0014, 0015, 0019, 0020	Strain4, Strain10, Strain14, Strain15, Strain19, Strain20	Yes
135	9044, 9477, 0014, 0015, 0020	Strain4, Strain10, Strain14, Strain15, Strain20	No
136	9684, 9313, 0014, 0015, 0019, 0020	Strain4, Strain10, Strain14, Strain15, Strain19, Strain20	Yes
138	9871, 9546, 0014	Strain4, Strain10, Strain14	No
139	9282, 9900, 0014, 0015, 0019, 0020	Strain4, Strain10, Strain14, Strain15, Strain19, Strain20	Yes
153	9375, 9273, 0014, 0015, 0019	Strain4, Strain10, Strain14, Strain15, Strain19	No
187	9155, 9981, 9724, 0014, 0015, 0019, 0020	Strain2, Strain4, Strain10, Strain14, Strain15, Strain19, Strain20	No
222	9286, 0014, 0015, 0019, 0020	Strain4, Strain14, Strain15, Strain19, Strain20	No

Strain4 and Strain10 are technical duplicates.

Annex 8. Reported results

SNP distances

			Laboratory No.					
Strain ID	ST	Provider	19*	88*	108	128*	132	153
Strain1	738	NA	NA	NA	101211	38339	18479	NA
Strain2	17	NA	NA	NA	60599	23166	9655	NA
Strain3	40	NA	NA	NA	58591	23090	9699	NA
Strain4#‡	301	0	0	4	3	9	0	0
Strain5	32	NA	NA	NA	70960	28588	11660	NA
Strain6	11	NA	NA	NA	74730	29385	9357	NA
Strain7	4748	NA	NA	NA	61047	23326	9747	NA
Strain8	1892	NA	NA	NA	73894	28824	11601	NA
Strain9	21	NA	NA	NA	57123	22985	9534	NA
Strain10#‡	301	0	0	4	5	9	0	0
Strain11	88	NA	NA	NA	56506	21920	9227	NA
Strain12	301	197	93	NA	248	48	14	127
Strain13	301	121	128	NA	207	55	14	111
Strain14‡	301	0	0	4	8	12	0	0
Strain15¤‡	301	0	0	0	0	0	0	0
Strain16	301	155	100	NA	319	47	12	98
Strain17	301	514	NA	NA	551	61	17	28
Strain18	NA	NA	NA	NA	NA	293	0	NA
Strain19‡	301	0	0	4	2	46	0	0
Strain20‡	301	0	0	4	8	12	4	8

Allelic differences

			Laboratory No.																					
Strain ID	ST	EQA	19	34	80	88	90	90*	100	123	124	124*	127	128	129	131	133	134	135	136	138	139	187	222
Strain1	738	2346	2350	2336	2290	2766	2335	2359	2350	2235	2350	2391	2393	1780	1437	2336	200	2347	2341	2301	2318	2389	2328	2235
Strain2	17	2164	2170	2148	2102	2589	2146	2168	5668	2143	2160	2193	2196	2155	1303	2147	200	2157	2151	3948	2133	2193	2138	2059
Strain3	40	2152	2150	2132	2090	2540	2134	2157	7345	2136	2140	2183	2185	3437	1298	2134	200	2144	2138	7305	2126	2182	2125	2041
Strain4#‡	301	0	0	0	0	4	0	1	0	0	0	1	3	1	0	0	0	0	1	1	8	2	0	16
Strain5	32	2302	2310	2262	2226	2716	2268	2292	2288	2269	2280	2325	2326	2355	1394	2270	200	2279	2272	2245	2256	2324	2257	2176
Strain6	11	2320	2330	2278	2244	2735	2294	2315	6689	2291	2310	2347	2376	2354	1408	2292	200	2302	2295	6657	2282	2346	2282	2199
Strain7	4748	2162	2170	2141	2101	2582	2141	2165	3997	2141	2150	2193	2195	2199	1298	2142	200	2152	2147	5623	2131	2190	2134	2048
Strain8	1892	2296	2300	2275	2227	2733	2275	2297	4488	2275	2280	2323	2326	2334	1395	2275	200	2285	2277	4451	2262	2322	2263	2179
Strain9	21	2184	2190	2161	2115	2590	2161	2183	7393	2159	2170	2211	2212	2226	1310	2162	200	2171	2165	7353	2144	2211	2140	2057
Strain10#‡	301	0	0	0	0	3	0	1	0	0	1	2	2	1	0	0	0	0	1	1	10	0	0	17
Strain11	88	2116	2120	2097	2053	2520	2096	2118	2131	2098	2110	2142	2144	2167	1273	2096	200	2107	2101	2076	2091	2141	2087	2005
Strain12	301	24	24	27	24	38	27	28	27	26	25	30	31	19	18	27	12	26	28	28	36	29	26	45
Strain13	301	24	23	25	22	35	25	26	25	22	24	28	27	26	13	25	15	25	25	26	30	27	23	27
Strain14‡	301	0	0	0	0	3	0	1	0	0	0	1	1	1	0	0	0	0	1	1	13	0	0	2
Strain15¤‡	301	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Strain16	301	24	21	25	23	34	25	26	25	24	24	28	28	27	15	25	14	25	26	26	36	27	24	23
Strain17	301	27	25	25	21	41	25	ND	27	25	24	30	NA	NA	22	25	NA	26	NA	25	NA	NA	NA	33
Strain18	NA	NA	NA	NA	NA	NA	NA	ND	88	6	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
Strain19‡	301	0	0	0	0	13	0	2	7	0	0	4	4	1	0	0	0	1	27	3	30	5	0	8
Strain20‡	301	3	3	0	0	3	0	1	0	0	NA	NA	NA	1	0	0	1	0	1	1	16	0	0	3

ST: sequence type
‡: closely related strains (in grey)
#: technical duplicate
¤: strain used as cluster representative by participant
NA: Not analysed
**: Additional analysis*

Annex 9. Reported QC parameters

Lab no.	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	No. of unidentified bases (N) or ambiguous sites	Available from QC analysis but no threshold		
80	N50	>30 000	Total number of contigs	<1 000				
88	Inter- and intra-species contamination	INNUca (using default kraken parameters) and ConFindr (using default parameters for E. coli).	Number of contigs	INNUca v4.2.2 default parameters				
90	N50	>20 000	contamination check	<5% other species	No.contigs >=200 bases	<1 000 contigs		
100	N50	40 000	contig count	500	contamination check with KmerFinder	most reads classified as E. coli	read length	corresponds to expected length of sequencing platform and kit
108	Toxin genes	Coverage >20x	Resistence genes (like ESBL)	Coverage >20x				
123	Average coverage	>50	assembly length	>5 000	N50	>50 000		
124	GC%	E.coli GC% +- 51%	N50	Threshold set in the quality control window of BioNumerics > 52 100	non-ACGT bases	Scatterplot (length vs non-ACGT)	Nr BAFPerfect	Scatterplot (length vs BAFPerfect)
127	N50 value	>20kb	Number of contigs	<=800	Proportion of scaffolding placeholders (N?s)	<3%	Species assignment using Kraken	> 70% contigs are assigned
128	total read length	expected length	KmerFinder contamination check	most species are E.coli	contig number	500	N50	50 000
129	contig count	200-2000						
131	N50	>100 kb	contig count	<500				
133	Average Quality	>=30	N50	>=70 000	N contigs	<500	NonAGCT	<2500
135	number of contigs	<= 650	GC%	Between 49.5 and 51.0%	N50	>=30 000	completeness (CheckM)	>96%
136	N50	>30 000 bp	Total number of contigs	Less than 550				
138	N50	x > 72 925	GC%	50.3 < x ≤ 50.9	number of contigs >=0bp	x ≤ 605	rMLST_Support_%	>90% of alleles
139	General read quality control	fastq_info v2.0	Inter-species contamination	Kraken2 (PlusPF-16 database), threshold 2%	N50	>20kb		
153	N50>30000							
222	Assembly quality (N50)	N50 >30 000 was used as a threshold for acceptable quality of the assembled contigs.						

In addition: Laboratory 128 also used L50 < 20; Laboratory 133: Contamination by species ID BioNumerics & PubMLST; Laboratory 135: Contamination (CheckM) <4%.

Annex 10. Calculated qualitative/quantitative parameters

Quality Assessment made by the SSI in-house quality control pipeline https://github.com/ssi-dk/bifrost [15]

Warning* (strain3 and strain11) was discarded by the EQA provider as the strains have been sub cultured 10 times, and similar results were observed for the EQA provider. *S. flexneri* and *E. coli* are highly genetically related and species identification using WGS is not always sufficient.

		Laboratory 19											
Parameter	**Ranges***	9071	9159	9359	9548	9603	9610	9668	9697	9713	9796	9842	9863
Detected species	{Ec} or {Sf}	Ec	Ec	Ec	Ec, Sf	Ec	Ec, Sf	Ec	Ec	Ec	Ec	Ec	Ec
% Species 1		91.0	94.7	95.5	88.6	95.4	68.8	84.4	88.6	93.7	94.2	94.4	95.7
% Species 2		1.1	0.7	0.2	5.4	0.4	21.7	0.5	1.3	0.5	0.4	0.4	0.1
Unclassified reads (%)	{<100}	7.0	4.2	4.1	5.1	4.0	5.0	14.0	7.9	4.4	4.5	4.2	4.1
Length at >25 x min. coverage (Mbp)	{>4.5 ^ <5.8}	5.3	5.5	5.1	5.0	5.3	5.2	5.5	5.3	5.4	5.3	5.3	5.2
Length [1-25] x min. coverage (kbp)	{<250}	14.8	1.9	9.1	4.2	19.1	11.9	9.3	8.2	4.1	9.5	19.9	33.1
No. of contigs at 25 x min. coverage	{>0}	268	364	412	203	353	338	389	189	369	334	419	380
No. of contigs [1-25] x min. coverage	{<1000}	14	2	12	5	23	15	10	10	6	15	24	41
Average coverage	{>50}	87	84	134	125	94	105	61	116	102	114	115	111
No. of reads (x 1 000)		3267	3299	4953	4464	3587	3929	2485	4484	4029	4422	4392	4145
Average read length		144	144	141	141	142	142	141	140	142	141	142	141
Average insert size		285	297	250	263	272	266	279	254	262	265	272	265
N50 (kbp)		48	62	40	65	48	39	54	81	55	54	39	39
QC-status (Bifrost)		OK	OK	OK	warning*	OK	warning *	OK	OK	OK	OK	OK	OK

All strains passed the QC.

		Laboratory 34										
Parameter	**Ranges***	9115	9343	9371	9466	9471	9478	9493	9560	9706	9745	9853	9882
Detected species	{Ec} or {Sf}	Ec	Ec	Ec, Sf	Ec	Ec	Ec	Ec	Ec	Ec	Ec	Ec	Ec
% Species 1		92.0	90.3	45.9	87.3	91.5	94.0	93.7	86.9	89.3	82.5	94.1	93.1
% Species 2		0.4	0.6	43.9	0.6	0.4	0.0	0.2	4.8	1.0	0.3	0.2	0.4
Unclassified reads (%)	{<100}	7.2	7.4	7.6	11.4	7.4	5.9	5.8	7.3	7.6	16.1	5.5	5.7
Length at >25 x min. coverage (Mbp)	{>4.5 ^ <5.8}	5.5	5.4	5.1	5.3	5.4	5.2	5.2	5.0	5.3	5.5	5.3	5.3
Length [1-25] x min. coverage (kbp)	{<250}	0.0	0.0	0.0	0.0	0.0	0.9	0.4	0.0	0.0	0.0	0.0	0.0
No. of contigs at 25 x min. coverage	{>0}	363	368	243	229	295	234	288	98	84	377	252	249
No. of contigs [1-	{<1000}	0	0	0	0	0	2	1	0	0	0	0	0

25] x min. coverage													
Average coverage	{>50}	98	63	88	115	111	47	56	105	97	120	116	116
No. of reads (x 1 000)		4220	2768	3559	4927	4804	1936	2281	4028	3647	5239	4722	4669
Average read length		131	127	128	126	127	130	131	132	143	131	133	134
Average insert size		222	205	211	203	208	222	226	233	278	219	240	236
N50 (kbp)		97	71	64	136	110	98	116	226	248	115	136	110
QC-status (Bifrost)		OK	OK	warning *	OK	OK	warning	OK	warning	OK	OK	OK	OK

Ten strains passed the QC, strain 9478 displays warning 'Average coverage' is just below 50, strain 9560 displays warning as '% Species 1' + '% unclassified' below 95%.

		Laboratory 80											
Parameter	**Ranges***	9150	9256	9293	9307	9333	9415	9526	9631	9639	9653	9913	9939
Detected species	{Ec} or {Sf}	Ec	Ec, Sf	Ec	Ec	Ec	Ec	Ec	Ec, Sf	Ec	Ec	Ec	Ec
% Species 1		94.0	62.8	92.7	88.2	87.2	90.5	89.5	86.4	81.8	93.4	92.3	92.3
% Species 2		0.7	25.8	1.1	0.9	1.3	1.0	0.8	5.0	0.9	0.6	0.7	0.8
Unclassified reads (%)	{<100}	5.0	6.1	5.6	9.0	8.7	5.9	7.9	6.1	15.7	5.3	5.5	6.1
Length at >25 x min. coverage (Mbp)	{>4.5 ^ <5.8}	5.2	5.2	5.3	5.3	5.3	5.5	5.4	5.0	5.5	5.2	5.4	5.5
Length [1-25] x min. coverage (kbp)	{<250}	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
No. of contigs at 25 x min. coverage	{>0}	213	225	265	200	92	308	259	101	342	262	250	312
No. of contigs [1-25] x min. coverage	{<1000}	0	0	0	0	0	0	0	0	0	0	0	0
Average coverage	{>50}	153	149	159	120	132	184	151	235	140	151	120	119
No. of reads (x 1 000)		5 681	5 510	6 133	4 548	4 924	7 246	5 876	8 488	5 576	5 596	4 510	4 585
Average read length		151	151	151	151	151	151	151	151	151	151	151	151
Average insert size		259	265	237	266	271	244	266	233	260	258	286	289
N50 (kbp)		124	60	119	111	221	88	110	171	130	160	108	95
QC-status (Bifrost)		OK	warning *	OK	OK	OK	OK	OK	warning *	OK	OK	OK	OK

All strains passed the QC.

		Laboratory 88											
Parameter	**Ranges***	9227	9494	9540	9556	9601	9686	9759	9789	9865	9966	9972	9990
Detected species	{Ec} or {Sf}	Ec	Sf, Ec	Ec	Ec	Ec	Ec	Ec	Ec	Ec	Ec, Sf	Ec	Ec
% Species 1		97.3	48.6	76.4	97.8	94.8	94.3	96.6	97.6	96.6	83.4	90.6	92.8
% Species 2		0.2	45.5	2.5	0.0	0.2	1.6	0.2	0.1	0.3	11.0	1.5	0.8
Unclassified reads (%)	{<100}	2.1	2.9	19.9	2.0	4.4	2.4	2.5	2.2	2.4	4.6	5.6	5.4
Length at >25 x min. coverage (Mbp)	{>4.5 ^ <5.8}	5.2	5.2	5.5	5.2	5.5	5.4	5.4	5.3	5.4	5.0	5.3	5.3
Length [1-25] x min. coverage (kbp)	{<250}	9.2	6.0	8.9	1.5	7.9	3.1	4.5	0.0	0.6	0.0	1.0	6.7
No. of contigs at 25 x min. coverage	{>0}	284	243	324	207	374	322	288	208	281	90	180	238

No. of contigs [1-25] x min. coverage	{<1000}	10	6	9	1	11	3	4	0	1	0	1	9
Average coverage	{>50}	114	109	103	128	116	115	124	133	120	133	158	115
No. of reads (x 1 000)		4 095	3 887	3 984	4 670	4 385	4 342	4 635	4 912	4 468	4 544	5 743	4 201
Average read length		148	148	147	146	148	147	147	147	147	148	148	148
Average insert size		380	405	360	323	365	369	348	336	354	364	373	398
N50 (kbp)		65	55	70	96	54	62	67	119	66	110	83	58
QC-status (Bifrost)		OK	warning *	OK	OK	OK	OK	OK	OK	OK	warning*	OK	OK

All strains passed the QC.

		Laboratory 90											
Parameter	**Ranges***	9081	9127	9168	9220	9301	9399	9522	9530	9629	9791	9876	9946
Detected species	{Ec} or {Sf}	Ec	Ec	Ec	Ec	Ec	Ec	Ec	Ec	Ec	Ec	Ec, Sf	Ec
% Species 1		97.0	93.0	95.2	90.8	94.5	93.4	86.7	96.9	96.6	96.3	77.2	98.0
% Species 2		0.4	0.8	1.1	1.4	1.0	1.8	0.6	0.4	0.6	0.7	15.7	0.1
Unclassified reads (%)	{<100}	2.2	5.0	2.2	5.2	2.8	3.5	11.3	2.3	2.2	1.6	1.7	1.7
Length at >25 x min. coverage (Mbp)	{>4.5 ^ <5.8}	5.3	5.3	5.4	5.3	5.5	5.0	5.5	5.2	5.5	5.4	5.2	5.2
Length [1-25] x min. coverage (kbp)	{<250}	4.3	8.0	6.3	17.4	8.7	0.3	12.8	48.8	27.6	16.9	29.1	16.3
No. of contigs at 25 x min. coverage	{>0}	233	218	297	139	316	111	335	286	334	245	242	193
No. of contigs [1-25] x min. coverage	{<1000}	4	7	6	11	10	1	10	42	30	21	30	12
Average coverage	{>50}	104	107	109	103	106	152	138	50	77	84	79	100
No. of reads (x 1 000)		2478	2559	2602	2436	2574	3677	3555	1103	1821	1942	1751	2274
Average read length		230	227	230	230	230	211	224	240	240	237	238	236
Average insert size		325	304	317	318	321	262	321	408	400	370	381	369
N50 (kbp)		94	85	63	110	61	141	77	55	57	74	57	132
QC-status (Bifrost)		OK	OK	OK	OK	OK	OK	OK	warning	OK	OK	warning*	OK

Eleven strains passed the QC, strain 9539 displays warning as 'Average coverage' is just below 50.

		Laboratory 100											
Parameter	**Ranges***	9034	9066	9117	9149	9158	9230	9289	9340	9451	9455	9672	9683
Detected species	{Ea}, {Ec}, {Pt} or {Sf}	Ec	Ec	Ec, Sf	Ec	Ec	Ec	Ec	Ec	Ec	Ec	Ec	Ec
% Species 1		92.0	87.3	51.8	95.2	97.9	95.8	95.7	95.6	97.9	97.8	92.5	93.2
% Species 2		1.2	0.6	43.5	1.1	0.3	0.7	0.9	1.2	0.3	0.1	1.1	3.1
Unclassified reads (%)	{<100}	4.4	10.8	1.8	1.5	1.4	1.7	1.7	2.6	1.4	1.8	5.1	2.5
Length at >25 x min. coverage (Mbp)	{>4.5 ^ <5.8}	5.3	5.5	5.1	5.2	5.2	5.1	5.3	5.4	5.3	5.2	5.3	5.0
Length [1-25] x min. coverage (kbp)	{<250}	0.0	0.0	73.9	258.8	0.6	286.6	113.1	86.4	0.0	81.6	19.9	0.0
No. of contigs at 25	{>0}	102	441	211	199	311	199	234	437	300	174	157	74

x min. coverage													
No. of contigs [1-25] x min. coverage	{<1000}	0	0	15	70	2	52	16	10	0	11	2	0
Average coverage	{>50}	89	82	78	77	87	66	83	84	86	77	76	76
No. of reads (x 1 000)		2085	1979	1592	1693	1994	1361	1822	2036	1994	1568	1599	1491
Average read length		236	244	266	259	238	270	255	237	239	268	263	266
Average insert size		244	252	304	285	245	317	279	241	248	316	297	311
N50 (kbp)		248	132	67	88	168	110	121	92	135	146	152	245
QC-status (Bifrost)		OK	OK	warning *	OK	OK	OK	OK	OK	OK	OK	OK	OK

All strains passed the QC.

		Laboratory 108											
Parameter	**Ranges***	9054	9347	9352	9393	9410	9553	9554	9643	9753	9768	9899	9905
Detected species	{Ea}, {Ec}, {Pt} or {Sf}	Ec	Ec	Ec	Ec	Ec	Ec	Ec	Ec	Ec, Sf	Ec	Ec	Ec
% Species 1		95.9	96.1	91.1	95.8	96.5	96.1	94.3	93.5	85.6	97.6	98.1	90.0
% Species 2		1.0	1.0	1.6	0.8	0.7	1.1	1.9	1.0	8.8	0.5	0.2	0.3
Unclassified reads (%)	{<100}	2.1	1.9	5.1	2.2	2.7	2.4	2.9	4.7	2.2	1.6	1.5	8.5
Length at >25 x min. coverage (Mbp)	{>4.5 ^ <5.8}	5.1	5.1	5.1	5.1	5.1	5.1	4.8	5.1	4.9	4.9	4.9	5.2
Length [1-25] x min. coverage (kbp)	{<250}	0.9	1.6	3.8	1.9	2.1	5.8	0.7	2.5	3.1	2.1	3.3	2.9
No. of contigs at 25 x min. coverage	{>0}	2948	3034	3206	3313	2802	3732	2677	2514	3010	3158	3338	3411
No. of contigs [1-25] x min. coverage	{<1000}	4	7	16	8	9	20	3	11	13	9	14	12
Average coverage	{>50}	70	82	67	70	93	63	78	76	63	62	64	63
No. of reads (x 1 000)		1339	1569	1264	1334	1777	1245	1375	1462	1135	1145	1176	1225
Average read length		285	287	287	291	285	288	290	286	292	289	291	294
Average insert size		0	0	0	0	0	0	0	0	0	0	0	0
N50 (kbp)		3	3	3	3	3	2	3	4	3	3	3	3
QC-status (Bifrost)		OK	OK	OK	OK	OK	OK	OK	OK	warning*	OK	OK	OK

All strains passed the QC. Some QC values is unreliable due to assembly issues for Ion Torrent data (Contigs, Average insert size, N50)

		Laboratory 123											
Parameter	**Ranges***	9042	9070	9193	9258	9337	9459	9705	9712	9904	9911	9918	9983
Detected species	{Ea}, {Ec}, {Pt} or {Sf}	Ec	Ec	Ec	Ec	Ec	Ec	Ec	Ec	Ec	Sf, Ec	Ec	Ec
% Species 1		91.6	73.6	92.6	92.8	89.6	92.9	90.5	88.2	94.6	48.3	94.6	92.8
% Species 2		3.7	2.6	1.3	1.5	1.3	1.1	1.1	5.0	1.8	43.6	1.6	1.4
Unclassified reads (%)	{<100}	4.1	20.2	3.2	3.8	5.4	4.9	6.8	4.6	3.2	5.0	3.0	4.0
Length at >25 x min. coverage (Mbp)	{>4.5 ^ <5.8}	0.4	4.1	5.5	5.4	5.3	5.5	5.3	5.0	5.2	5.2	5.2	5.4
Length [1-25] x min. coverage (kbp)	{<250}	4862.3	1397.7	0.4	6.0	7.5	1.2	7.6	0.7	23.9	1.0	9.1	15.7

No. of contigs at 25 x min. coverage	{>0}	134	258	264	275	95	372	225	70	252	254	241	301
No. of contigs [1-25] x min. coverage	{<1000}	284	107	1	7	5	2	8	1	31	2	18	14
Average coverage	{>50}	26	30	69	58	70	68	49	71	38	58	42	45
No. of reads (x 1 000)		565	683	1513	1326	1476	1503	1015	1392	822	1567	988	1004
Average read length		260	264	259	245	263	259	268	264	255	201	236	254
Average insert size		402	390	323	313	333	319	347	346	324	223	282	341
N50 (kbp)		40	52	69	74	203	60	73	237	59	60	122	52
QC-status (Bifrost)		warning	warning	OK	OK	OK	OK	warning	warning	warning	warning*	warning	warning

Five strains passed the QC. Strain 9042, 9070, 9705, 9904, 9918 and 9983 displays warnings as 'Average coverage' is just below 50. Strain 9712 displays warning as '% Species 1' + '% unclassified' is below 95%.

		Laboratory 124											
Parameter	**Ranges***	9067	9085	9093	9124	9348	9417	9541	9707	9754	9779	9800	9830
Detected species	{Ea}, {Ec}, {Pt} or {Sf}	Ec	Ec	Ec	Ec	Ec	Ec	Ec	Ec	Ec, Sf	Ec	Ec	Ec
% Species 1		95.8	92.8	97.4	95.4	94.7	87.4	97.7	96.3	82.2	89.8	96.9	95.1
% Species 2		0.5	1.1	0.2	0.6	1.0	0.3	0.1	0.9	9.3	1.2	0.4	0.8
Unclassified reads (%)	{<100}	2.7	5.0	2.1	2.7	3.2	11.2	2.0	2.5	3.4	6.8	2.4	3.0
Length at >25 x min. coverage (Mbp)	{>4.5 ^ <5.8}	5.4	5.3	5.2	5.5	5.0	5.5	5.2	5.5	5.2	5.3	5.3	5.4
Length [1-25] x min. coverage (kbp)	{<250}	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
No. of contigs at 25 x min. coverage	{>0}	208	145	186	199	54	228	166	250	186	64	177	193
No. of contigs [1-25] x min. coverage	{<1000}	0	0	0	0	0	0	0	0	0	0	0	0
Average coverage	{>50}	274	278	285	269	297	262	282	267	284	278	276	274
No. of reads (x 1 000)		10000	10000	10000	10000	10000	10000	10000	10000	10000	10000	10000	10000
Average read length		151	151	151	151	151	151	151	151	151	151	151	151
Average insert size		396	412	416	451	438	432	400	439	450	445	426	453
N50 (kbp)		110	160	166	86	247	136	141	102	67	238	136	110
QC-status (Bifrost)		OK	OK	OK	OK	OK	OK	OK	OK	warning*	OK	OK	OK

All strains passed the QC.

		Laboratory 127											
Parameter	**Ranges***	9006	9040	9130	9156	9195	9226	9318	9439	9801	9861	9878	9927
Detected species	{Ea}, {Ec}, {Pt} or {Sf}	Ec	Ec	Ec	Ec	Sf, Ec	Ec, Sf	Ec	Ec	Ec	Ec	Ec	Ec
% Species 1		89.7	81.5	92.9	89.7	63.7	82.2	75.9	81.2	87.5	87.7	88.6	80.6
% Species 2		0.5	0.6	1.5	0.5	30.4	7.5	2.3	0.4	0.2	1.2	0.6	0.2
Unclassified reads (%)	{<100}	9.0	17.4	3.8	9.1	4.4	9.2	20.5	17.4	12.0	8.8	9.9	18.9
Length at >25 x min. coverage (Mbp)	{>4.5 ^ <5.8}	5.3	5.2	5.4	5.3	5.2	5.0	5.5	5.4	5.1	5.3	5.3	5.3
Length [1-25] x min.	{<250}	15.7	19.5	8.5	27.3	8.2	18.5	17.2	49.5	26.3	12.5	16.1	9.3

coverage (kbp)													
No. of contigs at 25 x min. coverage	{>0}	422	324	454	306	307	190	404	481	360	230	369	336
No. of contigs [1-25] x min. coverage	{<1000}	23	24	11	23	8	9	22	44	33	14	22	13
Average coverage	{>50}	116	91	147	68	108	95	78	86	111	113	100	114
No. of reads (x 1 000)		4501	3628	5553	2501	3868	3319	2992	3461	4190	4126	3814	4672
Average read length		147	148	147	149	148	149	149	149	149	149	149	149
Average insert size		327	332	343	359	348	350	360	373	374	368	357	349
N50 (kbp)		35	59	36	42	42	56	41	32	40	66	42	48
QC-status (Bifrost)		OK	OK	OK	OK	warning *	warning *	warning	OK	OK	OK	OK	OK

Eleven strains passed the QC, strain 9318 displays warning as '% unclassified above 20%'.

		Laboratory 128											
Parameter	**Ranges***	9101	9147	9157	9172	9338	9446	9467	9561	9741	9869	9888	9974
Detected species	{Ea}, {Ec}, {Pt} or {Sf}	Ec, Sf	Ec	Ec	Ec	Ec	Ec	Ec	Ec	Ec	Ec	Ec	Ec
% Species 1		78.2	95.0	94.6	94.7	80.5	87.2	94.7	96.9	92.0	89.9	96.5	97.2
% Species 2		12.4	1.0	1.2	0.6	0.4	0.3	0.6	0.2	1.0	1.3	0.5	0.1
Unclassified reads (%)	{<100}	3.4	3.6	3.1	3.3	18.4	11.5	3.6	2.5	5.9	6.6	2.7	2.6
Length at >25 x min. coverage (Mbp)	{>4.5 ^ <5.8}	5.2	5.5	5.0	5.5	5.3	5.5	5.4	5.2	5.3	5.3	5.3	5.2
Length [1-25] x min. coverage (kbp)	{<250}	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
No. of contigs at 25 x min. coverage	{>0}	208	288	71	253	343	284	224	205	166	73	212	188
No. of contigs [1-25] x min. coverage	{<1000}	0	0	0	0	0	0	0	0	0	0	0	0
Average coverage	{>50}	150	148	193	125	163	127	169	126	158	192	137	198
No. of reads (x 1 000)		5313	5608	6514	4669	9168	4872	6216	4439	5736	6950	4974	7075
Average read length		149	149	148	149	99	149	149	149	149	149	149	149
Average insert size		348	332	326	342	139	346	347	349	335	338	336	333
N50 (kbp)		65	102	234	88	94	133	107	166	156	249	134	141
QC-status (Bifrost)		warning *	OK	OK	OK	OK	OK	OK	OK	OK	OK	OK	OK

All strains passed the QC.

		Laboratory 129											
Parameter	**Ranges***	9016	9053	9091	9110	9187	9281	9346	9402	9438	9637	9827	9973
Detected species	{Ea}, {Ec}, {Pt} or {Sf}	Ec, Sf	Ec	Ec	Ec	Ec, Sf	Ec	Ec	Ec	Ec	Ec	Ec	Ec
% Species 1		77.6	96.9	95.0	95.4	88.5	95.5	86.5	96.5	91.7	95.5	96.6	90.0
% Species 2		12.8	0.2	0.7	0.6	5.2	0.7	0.3	0.1	1.0	0.8	0.3	1.2
Unclassified reads (%)	{<100}	5.1	2.6	3.4	3.0	5.2	2.5	11.7	3.3	5.9	3.2	2.8	6.3
Length at >25 x min. coverage (Mbp)	{>4.5 ^ <5.8}	5.1	5.1	5.3	5.3	4.9	5.4	5.3	5.2	5.3	5.4	5.1	5.3
Length [1-25] x min.	{<250}	4.8	2.9	5.1	13.3	33.9	4.0	23.0	4.9	11.6	8.8	57.0	6.0

coverage (kbp)													
No. of contigs at 25 x min. coverage	{>0}	314	297	315	347	214	395	328	252	278	440	327	167
No. of contigs [1-25] x min. coverage	{<1000}	6	5	4	13	30	5	22	5	16	10	44	6
Average coverage	{>50}	128	136	129	94	69	126	76	126	88	107	83	115
No. of reads (x 1 000)		4636	4918	4793	3463	2380	4734	2962	4560	3270	4179	3082	4246
Average read length		147	148	147	148	146	148	148	148	145	146	146	146
Average insert size		344	327	328	383	426	323	431	333	325	311	419	355
N50 (kbp)		49	85	71	53	54	56	59	77	59	54	46	84
QC-status (Bifrost)		warning *	OK	OK	OK	warning *	OK	OK	OK	OK	OK	OK	OK

All strains passed the QC.

		Laboratory 131											
Parameter	**Ranges***	9008	9030	9074	9146	9152	9208	9290	9376	9391	9615	9665	9823
Detected species	{Ea}, {Ec}, {Pt} or {Sf}	Ec	Ec	Ec, Sf	Ec	Ec	Ec	Ec	Ec	Ec	Ec	Ec	Ec
% Species 1		95.9	94.2	72.6	90.3	94.2	94.8	97.0	86.5	97.3	95.6	92.6	97.3
% Species 2		0.3	0.6	19.4	1.2	1.8	0.4	0.3	0.3	0.1	0.3	1.2	0.2
Unclassified reads (%)	{<100}	2.9	4.7	3.0	6.2	3.0	3.1	2.5	11.9	2.4	2.9	5.1	2.2
Length at >25 x min. coverage (Mbp)	{>4.5 ^ <5.8}	5.4	5.5	5.2	5.3	5.0	5.5	5.3	5.5	5.2	5.4	5.3	5.2
Length [1-25] x min. coverage (kbp)	{<250}	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
No. of contigs at 25 x min. coverage	{>0}	274	311	236	79	91	271	239	383	204	290	209	263
No. of contigs [1-25] x min. coverage	{<1000}	0	0	0	0	0	0	0	0	0	0	0	0
Average coverage	{>50}	121	109	141	119	145	129	122	139	141	152	132	134
No. of reads (x 1 000)		4499	4138	5074	4368	4986	4840	4492	5422	5110	5679	4835	4835
Average read length		149	149	149	149	149	149	149	149	149	149	149	149
Average insert size		278	294	262	290	281	310	285	250	289	250	264	268
N50 (kbp)		93	97	64	222	189	84	130	130	124	108	113	166
QC-status (Bifrost)		OK	OK	warning *	OK	OK	OK	OK	OK	OK	OK	OK	OK

All strains passed the QC.

		Laboratory 132											
Parameter	**Ranges***	9014	9114	9144	9189	9275	9353	9365	9483	9595	9784	9893	9947
Detected species	{Ea}, {Ec}, {Pt} or {Sf}	Ec	Ec	Ec	Ec	Ec	Ec	Ec, Sf	Ec	Ec	Ec, Sf	Ec	Ec
% Species 1		97.3	82.9	95.3	96.1	98.0	98.3	49.8	94.6	97.1	84.1	92.2	93.0
% Species 2		0.2	0.8	0.1	0.6	0.3	0.1	44.8	0.3	0.3	10.5	1.1	0.4
Unclassified reads (%)	{<100}	1.7	15.1	4.4	2.2	1.6	1.5	2.2	4.7	2.0	4.2	4.7	6.0
Length at >25 x min. coverage (Mbp)	{>4.5 ^ <5.8}	5.0	3.3	2.7	4.6	5.3	4.0	4.1	1.0	3.0	4.2	5.2	4.5
Length [1-25] x min.	{<250}	331.9	2130.1	2483.7	821.9	2.1	1188.8	1085.7	4326.6	2335.2	780.4	94.9	844.8

coverage (kbp)													
No. of contigs at 25 x min. coverage	{>0}	222	242	121	218	226	182	161	196	158	60	75	156
No. of contigs [1-25] x min. coverage	{<1000}	45	347	109	65	6	66	80	675	190	79	8	41
Average coverage	{>50}	44	29	27	40	60	39	37	26	31	40	49	37
No. of reads (x 1 000)		1621	1106	971	1482	2182	1402	1309	965	1128	1351	1765	1340
Average read length		149	149	150	149	149	149	149	148	150	149	149	149
Average insert size		338	290	427	361	298	331	340	262	434	361	359	386
N50 (kbp)		97	69	90	72	129	116	56	22	63	183	224	111
QC-status (Bifrost)		warning	warning	warning	warning	OK	warning	warning *	warning	warning	warning*	warning	warning

One strain passed the QC. Eleven strains display warnings as the 'Average coverage' is just below 50.

		Laboratory 133											
Parameter	**Ranges***	9036	9052	9089	9126	9162	9180	9186	9197	9246	9647	9772	9868
Detected species	{Ea}, {Ec}, {Pt} or {Sf}	Ec	Ec	Ec	Ec	Ec	Ec	Ec	Ec	Ec	Ec, Sf	Ec	Ec
% Species 1		95.6	89.6	90.7	95.3	95.1	98.0	97.7	94.2	87.1	63.7	96.1	97.5
% Species 2		0.6	3.1	1.1	0.7	0.7	0.1	0.2	0.6	0.4	29.8	0.5	0.2
Unclassified reads (%)	{<100}	2.3	5.8	5.8	2.3	2.8	1.7	1.8	4.3	11.1	1.6	2.9	1.9
Length at >25 x min. coverage (Mbp)	{>4.5 ^ <5.8}	5.4	5.0	5.3	5.5	5.4	5.2	5.3	5.3	5.6	5.1	5.5	5.2
Length [1-25] x min. coverage (kbp)	{<250}	0.0	0.0	0.0	0.0	0.5	0.0	0.0	0.0	0.0	52.0	0.0	2.2
No. of contigs at 25 x min. coverage	{>0}	192	49	54	218	193	163	197	144	273	273	267	180
No. of contigs [1-25] x min. coverage	{<1000}	0	0	0	0	1	0	0	0	0	40	0	4
Average coverage	{>50}	99	72	108	120	141	176	206	153	107	52	196	81
No. of reads (x 1 000)		1936	1300	2080	2404	2744	3446	4227	3100	2261	1053	4053	1583
Average read length		286	290	284	285	289	275	268	272	276	264	277	276
Average insert size		392	498	404	376	410	343	323	332	339	299	358	351
N50 (kbp)		110	304	355	110	121	147	136	156	134	43	104	160
QC-status (Bifrost)		OK	OK	OK	OK	OK	OK	OK	OK	warning	warning*	OK	OK

Eleven strains passed the QC. Strain 9246 genome size is higher than expected.

		Laboratory 134											
Parameter	**Ranges***	9083	9210	9225	9279	9443	9458	9727	9816	9880	9884	9897	9929
Detected species	{Ea}, {Ec}, {Pt} or {Sf}	Ec	Ec	Ec	Ec	Ec	Ec	Ec	Ec	Ec	Ec, Sf	Ec, Sf	Ec
% Species 1		97.5	91.8	94.8	90.2	96.0	97.8	95.9	85.1	95.1	54.7	89.4	97.4
% Species 2		0.1	0.8	0.7	1.2	0.3	0.1	0.4	0.5	0.4	39.1	5.7	0.2
Unclassified reads (%)	{<100}	2.1	6.6	2.8	6.4	2.8	2.0	2.9	13.3	4.1	2.9	3.8	2.3
Length at >25 x min. coverage (Mbp)	{>4.5 ^ <5.8}	5.2	5.3	5.4	5.3	5.4	5.2	5.4	5.5	5.5	5.2	5.0	5.3
Length [1-25] x min.	{<250}	0.0	0.0	0.0	0.0	0.0	0.0	0.3	0.0	0.5	0.0	0.0	0.0

coverage (kbp)													
No. of contigs at 25 x min. coverage	{>0}	216	183	313	80	293	199	239	286	299	229	91	222
No. of contigs [1-25] x min. coverage	{<1000}	0	0	0	0	0	0	1	0	1	0	0	0
Average coverage	{>50}	101	59	59	67	85	80	55	61	58	79	91	74
No. of reads (x 1 000)		3608	2147	2213	2447	3166	2887	2037	2339	2205	2803	3127	2722
Average read length		149	148	147	149	146	148	149	149	149	148	147	148
Average insert size		323	314	245	316	227	306	328	348	321	300	255	314
N50 (kbp)		160	109	82	248	88	139	93	130	99	60	165	129
QC-status (Bifrost)		OK	OK	OK	OK	OK	OK	OK	OK	OK	warning*	warning*	OK

All strains passed the QC.

		Laboratory 135											
Parameter	**Ranges***	9044	9207	9241	9319	9345	9477	9591	9720	9722	9765	9803	9956
Detected species	{Ea}, {Ec}, {Pt} or {Sf}	Ec	Ec	Ec, Sf	Ec	Ec	Ec	Ec	Ec	Ec	Ec	Ec	Ec
% Species 1		96.4	97.7	79.9	96.3	92.6	95.9	89.8	87.4	94.7	95.7	97.7	96.9
% Species 2		0.3	0.2	11.8	0.9	0.9	0.4	1.2	0.3	0.8	0.4	0.1	0.4
Unclassified reads (%)	{<100}	2.2	1.8	2.9	2.4	5.0	2.7	6.5	11.1	3.3	2.4	2.0	2.4
Length at >25 x min. coverage (Mbp)	{>4.5 ^ <5.8}	5.4	5.2	5.2	5.5	5.3	5.4	5.3	5.5	5.0	5.5	5.2	5.3
Length [1-25] x min. coverage (kbp)	{<250}	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
No. of contigs at 25 x min. coverage	{>0}	239	224	195	279	160	216	67	265	64	235	179	190
No. of contigs [1-25] x min. coverage	{<1000}	0	0	0	0	0	0	0	0	0	0	0	0
Average coverage	{>50}	263	148	93	199	181	81	165	176	124	195	180	113
No. of reads (x 1 000)		9663	5242	3310	7506	6534	2978	5993	6755	4193	7292	6421	4083
Average read length		151	151	151	151	151	151	151	151	151	151	151	151
Average insert size		287	311	369	356	371	387	362	376	382	361	372	374
N50 (kbp)		110	180	64	97	136	110	327	134	234	88	139	119
QC-status (Bifrost)		OK	OK	warning *	OK	OK	OK	OK	OK	OK	OK	OK	OK

All strains passed the QC.

		Laboratory 136											
Parameter	**Ranges***	9121	9194	9313	9369	9488	9571	9590	9596	9649	9684	9698	9840
Detected species	{Ea}, {Ec}, {Pt} or {Sf}	Ec	Ec	Ec	Ec	Ec	Ec	Ec	Ec	Ec	Ec	Ec	Ec, Sf
% Species 1		96.1	87.4	95.8	90.2	95.2	96.9	94.8	93.0	97.0	95.6	97.2	72.0
% Species 2		0.8	0.3	0.3	1.3	0.3	0.1	1.2	0.7	0.3	0.3	0.2	20.2
Unclassified reads (%)	{<100}	2.7	11.2	3.0	6.3	3.3	2.9	3.0	5.4	2.5	3.1	2.3	3.0
Length at >25 x min. coverage (Mbp)	{>4.5 ^ <5.8}	5.5	5.5	5.4	5.3	5.5	5.2	5.0	5.3	5.3	5.4	5.2	5.2
Length [1-25] x min.	{<250}	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0

coverage (kbp)													
No. of contigs at 25 x min. coverage	{>0}	289	277	227	64	239	156	86	165	211	224	212	206
No. of contigs [1-25] x min. coverage	{<1000}	0	0	0	0	0	0	0	0	0	0	0	0
Average coverage	{>50}	248	198	289	230	195	148	303	208	221	244	272	295
No. of reads (x 1 000)		9409	7624	10744	8366	7389	5302	10352	7594	8120	9027	9782	10522
Average read length		149	148	147	148	147	148	148	148	148	149	149	148
Average insert size		342	349	320	357	354	388	301	332	337	343	354	324
N50 (kbp)		97	134	107	319	86	141	226	156	129	110	166	66
QC-status (Bifrost)		OK	OK	OK	OK	OK	OK	OK	OK	OK	OK	OK	warning*

All strains passed the QC.

		Laboratory 138											
Parameter	**Ranges***	9028	9235	9509	9529	9546	9581	9604	9640	9839	9871	9955	9997
Detected species	{Ea}, {Ec}, {Pt} or {Sf}	Ec	Ec	Ec	Ec	Ec	Sf, Ec	Ec	Ec	Ec, Sf	Ec	Ec	Ec
% Species 1		95.3	91.4	97.1	98.6	97.4	56.5	97.8	94.1	89.6	96.7	78.8	91.5
% Species 2		0.8	0.4	0.1	0.0	0.3	38.9	0.1	0.4	6.3	0.3	1.4	1.3
Unclassified reads (%)	{<100}	2.2	7.7	2.6	1.3	1.6	2.5	1.8	5.1	3.1	2.3	18.7	5.0
Length at >25 x min. coverage (Mbp)	{>4.5 ^ <5.8}	5.5	5.3	5.2	5.2	5.3	5.2	5.2	5.5	5.0	5.3	5.5	5.3
Length [1-25] x min. coverage (kbp)	{<250}	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
No. of contigs at 25 x min. coverage	{>0}	242	163	386	184	400	191	212	279	70	419	279	71
No. of contigs [1-25] x min. coverage	{<1000}	0	0	0	0	0	0	0	0	0	0	0	0
Average coverage	{>50}	382	315	336	413	361	325	298	334	384	373	383	446
No. of reads (x 1 000)		14207	11378	12190	14668	13556	11453	10640	12551	12951	14027	14648	16060
Average read length		150	150	151	150	149	150	150	150	150	151	150	150
Average insert size		356	348	0	339	0	406	368	352	345	0	340	339
N50 (kbp)		86	143	126	141	93	66	166	101	226	90	132	319
QC-status (Bifrost)		OK	OK	OK	OK	OK	warning *	OK	OK	warning*	OK	OK	OK

All strains passed the QC.

		Laboratory 139											
Parameter	**Ranges***	9181	9282	9295	9335	9381	9498	9557	9645	9674	9715	9825	9900
Detected species	{Ea}, {Ec}, {Pt} or {Sf}	Ec	Ec	Ec	Ec	Ec	Ec	Ec	Ec	Ec	Ec	Ec	Ec
% Species 1		88.6	93.0	92.3	94.3	91.5	93.4	90.3	86.2	95.1	94.0	86.2	93.0
% Species 2		1.4	0.8	0.7	0.4	1.1	1.4	1.3	0.3	0.1	0.8	4.4	0.9
Unclassified reads (%)	{<100}	7.6	4.9	5.8	4.8	5.9	4.7	7.2	12.3	4.6	4.8	5.0	5.0
Length at >25 x min. coverage (Mbp)	{>4.5 ^ <5.8}	5.3	5.4	5.0	5.2	5.4	5.5	5.3	5.5	5.2	5.3	5.2	5.4
Length [1-25] x min.	{<250}	4.8	0.0	10.1	20.8	0.8	0.0	0.0	6.2	0.8	3.5	7.8	4.0

coverage (kbp)													
No. of contigs at 25 x min. coverage	{>0}	110	226	149	223	261	295	171	274	179	227	212	236
No. of contigs [1-25] x min. coverage	{<1000}	5	0	7	9	1	0	0	4	1	1	3	1
Average coverage	{>50}	111	141	63	77	74	74	72	94	83	70	95	115
No. of reads (x 1 000)		4031	5183	2171	2788	2811	2807	2606	3619	2978	2554	3364	4268
Average read length		149	150	150	149	149	149	150	150	149	148	149	149
Average insert size		386	404	380	390	393	404	404	390	388	368	389	388
N50 (kbp)		146	88	81	138	74	82	101	119	126	94	63	88
QC-status (Bifrost)		OK	OK	OK	OK	OK	OK	OK	OK	OK	OK	warning	OK

Eleven strains passed the QC. Strain 9825 displays warning as '% Species 1' + '% unclassified' below 95%.

		Laboratory 153											
Parameter	**Ranges***	9024	9041	9132	9209	9228	9273	9320	9375	9387	9495	9566	9646
Detected species	{Ea}, {Ec}, {Pt} or {Sf}	Ec	Ec	Ec	Ec	Ec	Ec	Ec	Ec	Ec	Ec	Ec, Sf	Ec
% Species 1		95.6	94.3	92.2	84.1	96.8	95.8	93.5	95.9	90.3	97.2	49.6	96.8
% Species 2		1.0	1.5	1.0	0.6	0.3	0.4	0.6	0.4	1.3	0.1	44.5	0.4
Unclassified reads (%)	{<100}	3.1	3.2	5.8	14.3	2.6	2.9	4.4	2.8	6.2	2.6	3.2	2.6
Length at >25 x min. coverage (Mbp)	{>4.5 ^ <5.8}	5.4	4.9	5.2	5.4	5.1	5.0	5.1	5.0	5.3	5.2	5.0	5.2
Length [1-25] x min. coverage (kbp)	{<250}	122.5	48.1	128.7	108.7	49.7	312.6	289.6	250.4	0.0	30.3	225.7	68.9
No. of contigs at 25 x min. coverage	{>0}	338	91	196	344	238	239	239	259	89	176	205	245
No. of contigs [1-25] x min. coverage	{<1000}	15	3	13	7	14	132	118	47	0	19	37	7
Average coverage	{>50}	58	52	43	41	46	39	48	66	55	52	40	43
No. of reads (x 1 000)		2212	1809	1571	1618	1649	1457	1873	2493	2025	1908	1427	1574
Average read length		146	145	146	146	147	146	141	143	146	145	146	146
Average insert size		262	271	265	262	272	278	251	227	270	267	273	261
N50 (kbp)		97	189	136	129	140	88	74	94	248	140	63	126
QC-status (Bifrost)		OK	OK	warning	warning	warning	warning	warning	OK	OK	OK	warning*	warning

Six strains passed the QC, six strains displays warnings as 'Average coverage' is below 50.

		Laboratory 187											
Parameter	**Ranges***	9155	9185	9276	9444	9724	9783	9819	9844	9879	9912	9968	9981
Detected species	{Ea}, {Ec}, {Pt} or {Sf}	Ec	Ec	Ec	Ec	Ec	Ec	Ec	Ec, Sf	Ec	Ec	Ec	Ec
% Species 1		97.9	94.7	90.8	98.1	96.3	88.0	96.6	75.2	93.3	95.4	98.0	96.2
% Species 2		0.3	1.8	1.1	0.1	0.3	0.3	0.9	17.8	1.0	0.5	0.1	0.4
Unclassified reads (%)	{<100}	1.7	2.6	5.9	1.8	2.5	10.7	2.3	2.3	4.8	2.6	1.6	2.3
Length at >25 x min. coverage (Mbp)	{>4.5 ^ <5.8}	5.3	4.5	5.3	5.2	5.2	5.5	4.7	5.2	5.3	5.3	5.2	5.4
Length [1-25] x min.	{<250}	0.0	446.3	0.0	0.9	186.8	0.0	727.4	0.0	0.0	116.8	4.8	9.7

coverage (kbp)													
No. of contigs at 25 x min. coverage	{>0}	253	63	67	184	236	417	258	216	201	256	217	228
No. of contigs [1-25] x min. coverage	{<1000}	0	19	0	2	16	0	37	0	0	8	3	2
Average coverage	{>50}	78	31	49	53	50	84	35	91	76	42	37	70
No. of reads (x 1 000)		2898	1077	1776	1913	1850	3352	1337	3260	2784	1586	1336	2607
Average read length		146	147	147	147	147	142	147	147	146	147	147	147
Average insert size		250	404	375	348	335	196	408	328	276	404	371	349
N50 (kbp)		131	171	224	115	95	130	83	67	113	74	136	90
QC-status (Bifrost)		OK	warning	warning	OK	warning	OK	warning	warning *	OK	warning	warning	OK
Six strains passed the QC, six strains (9185, 9276, 9724, 9819, 9912 and 9968) displays warnings as the 'Average coverage' is below 50													

		Laboratory 222											
Parameter	**Ranges***	9031	9264	9286	9314	9520	9589	9644	9858	9872	9952	9958	9977
Detected species	{Ea}, {Ec}, {Pt} or {Sf}	Ec	Ec	Ec	Ec	Ec	Ec	Ec	Ec	Ec	Ec, Sf	Ec	Ec
% Species 1		89.6	94.3	94.9	93.5	95.8	95.9	94.5	95.1	97.5	88.0	90.7	97.0
% Species 2		0.4	1.3	1.1	1.0	0.9	1.4	1.2	1.3	0.4	5.4	1.3	0.7
Unclassified reads (%)	{<100}	8.7	2.4	2.6	3.7	2.4	1.8	2.1	2.0	1.8	2.3	5.8	1.9
Length at >25 x min. coverage (Mbp)	{>4.5 ^ <5.8}	5.4	5.3	5.1	5.2	5.2	5.0	5.1	4.9	5.0	5.0	5.2	5.1
Length [1-25] x min. coverage (kbp)	{<250}	0.0	0.0	0.0	0.0	0.5	0.0	0.0	0.0	0.0	18.9	0.0	0.0
No. of contigs at 25 x min. coverage	{>0}	1488	1302	3042	1026	1394	3037	2811	1363	2018	2299	1404	1616
No. of contigs [1-25] x min. coverage	{<1000}	0	0	0	0	2	0	0	0	0	68	0	0
Average coverage	{>50}	193	198	161	187	125	171	169	225	177	156	162	182
No. of reads (x 1 000)		3696	3591	2835	3138	2339	3276	2952	3569	2985	2784	3243	3396
Average read length		299	306	309	322	290	295	313	319	315	299	272	284
Average insert size		7	9	10	5	14	6	8	8	9	5	12	9
N50 (kbp)		7	7	3	9	7	2	3	6	4	4	6	6
QC-status (Bifrost)		OK	OK	OK	OK	OK	OK	OK	OK	OK	warning*	OK	OK
All strains passed the QC. Some QC values is unreliable due to assembly issues for Ion Torrent data (Contigs, Average insert size, N50)													

Quality assessment made by the EQA-provider in-house quality control pipeline.
**: indicative QC ranges; Ec. E. coli, Sf: S.flexneri (listed if >5%).*

Annex 11. Accessing provided sequences

Lab ID	Sero/Stx sub	ST	Cluster	QC Status	Description Strain17
EQA provider		**301**	**No**	**B**	**A nonCluster sequence with low quality of R2**
19	O80:H2/stx2f	301	No	B	The sequence quality is only accepted for outbreak situations. Raw read files are not of same size and is likely a result of incorrect file transfer. A denovo assembly was made of a single read file and the cgMLST analysis is only based on assembly-based calls. A supporting SNP analysis was made but the quality of the analysis was affected and strain0017 was excluded from the SNP analysis. The file transfer has to be redone or strain0017 has to be resequenced for cluster analysis.
34	O80:H2	301	No	A	Good QC, good coverage
80	O80:H2/stx2f	301	No	B	97,7% good targets (cgMLST) -Top species match SeqSphere E. coli 1.0 - FastQC per base sequence quality (reverse reads) – failed! (Phred 4) -Raw reads: -Avg. read length: 139 - No. of reads: 4365754 -Coverage: 132 -Assembled reads -Size: 5,1 Mb -N50: 18776 (< 30 000, less good quality!) -Tot. no of contigs: 836
88	O80:H2/stx2f	301	No	B	QC failed due to absence of quality data in the R2 sequencing file. Still, an assembly was made with R1, passing all downstream QC criteria (good coverage and >95% loci called).
90	O80:H2	301	No	B	FastQC per base sequence quality check for reverse reads failed (Q score 4). Consequently, only the forward reads were used for the assembly by SKESA. QC parameters: average genome coverage -114x size of assembled genome – 5.1 Mbp N50 – 18776 total no. of contigs >=200 bases – 836 cgMLST good targets – 97.7% contamination check (<5% other species) – no evidence of contamination.
100	O80:H2/stx2f	301	No	B	FastQC parameters are OK for read 1 (R1). Read 2 (R2) seems to be corrupted or something actually went wrong during sequencing. All reads in R2 are below Q5. GC content for R1 is OK. Assembly was attempted and successfully done- however, assembly stats are lower. N50 is 48795. Total assembled genome length is 5221769 and the number of contigs is 336. Genome coverage is 110x. Kmerfinder results show no significant contamination (2% of reads are possibly attributed to Salmonella). % of good cgMLST targets is 99%.
108	O80:H2/stx2f	301	No	A	Strain 0017 – QC status Ok.
123	O80:H2	301	No	A	% good targets E. coli cgMLST: 98,8 (our threshold: 98%) species match: E. coli (no evidence for contamination) GC content: 50,6 Genome size: 5,4 Av.Coverage: 64 (our threshold : 50) N50: only 21532 (our threshold: 50000).
124	O80:H2/stx2f	301	No	B	The de novo assembly could initially not be performed in BioNumerics as the quality of the R2-file was low (Q<30). Therefore, only the R1 file was analyzed. Strain 17 has a genome size of 5.2 Mb, a high number of N bases, a relatively low N50 (27735), a relatively high number of contigs (657) and an average read coverage of 59. 96% of alleles were called.
127				C	Mixture of *Salmonella* sp. And *Escherichia* sp.
128	O80:H2/stx2f	301	No	B	FASTQC parameters for read 1 has OK QC, has quality score around Q30. Read 2 has bad QC and very low quality scores <4. GC% for raw data is 52 for both reads, which is good. Assembly was successful using SPAdes despite low quality of R2. However, it was not possible to assemble the sequences in Enterobase, that is why the allelic difference could not be determined. Assembled genome has 336 contigs. Total assembly length is 5221769 bp. N50 is 48795 and L50 is 35. Assembly statistics are lower than for previous strains. GC% for assembled genome is 50.47. KmerFinder shows contamination: *Salmonella* enterica 2%, E.albertii 1% and E.marmotae 1%.
129		301	No	A	Good quality (percentage of good targets 97.4%, contig count 1616, average coverage 67).
131	O80:H2	301	No	A	cgMLST Perc. Good Targets: 97,7 % Avg. Coverage (Assembled): 114x Approximated Genome Size: 5,1 Mb N50 (Assembled): 18776 Top Species (Match Identity): *Escherichia coli* (1.0) Contig Count (Assembled): 836.
132			No	C	Was not a part of the cluster with 17 SNP difference. Other things were not done.

133				C	Very Poor Average Quality: >= 30 =FALSE, 18 Average Coverage: >=30 =TRUE N50: >=70,000 =FALSE, ? Ncontigs:<500 =FALSE, ? NonAGCT:<2500 =FALSE, ? Length: 5Mb-5.8Mb =FALSE, ? CorePercent: >96% =FALSE, 94% This strain failed 6/7 QC parameters. Therefore, this strain is not acceptable for analysis.
134		301	No	B	Coverage = 111 Countig count = 1888 (value >1000 but all other criteria are in the good range) Targets found = 98.6 Genome size = 5.4 Statistics indicated that FastQC for reverse reads failed.
135				C	fastq files could not be assembled, file 17_R2 rejected in fastqc report based on per base sequence quality and per sequence quality score.
136	O80:H2	301	No	B	This sample lost all of its reads during pre-processing because the R2 reads were all of poor quality. We used only R1 (single-end sample) as approach for the outbreak investigation.
138				C	QC – Fail: Not enough reads and very low Q30 q30_rate':0.395606' Possible issued during library preparation.
139				C	Average phred score for the R2 file below acceptable ranges. Impossible to assemble.
153	O80:H2	301	No	B	R2 file has low PHRED score (=4). QC for only R1 file is acceptable. Just for outbreak situation only R1 file was used.
187				C	Too many contigs.
222	O80:H2/stx2f	301	No	B	The R2 file caused an error in the analytical pipeline, due to incorrect Phred values. The R1 file was then analysed as single end. 2328/2360 loci of cgMLST scheme correctly mapped, satisfying our quality threshold set at 80% for reliability of cluster analysis. The coverage was 61x. Moreover, the 7 genes of conventional MLST (Warwick scheme) were all found 100% in length. N50 calculated on the assembled contigs was 28992 bp, which is <30000, used as a threshold for A of the assembly. The sequence could be accepted only for outbreak investigation.

Lab ID	Sero/Stx sub	ST	Cluster	QC Status	Description Strain18
EQA provider	**-**	**-**	**No**	**C**	**A nonCluster sequence with low coverage**
19				C	The sequence quality is not accepted for cluster analysis because of low read coverage, resulting in a low cgMLST core%. The strain should be resequenced for higher read coverage.
34				C	Bad quality, only 15x coverage, only 1120 cgMLST targets found (out of 2513).
80				C	44.6% good targets (cgMLST) (not acceptable! < 90% good targets) -Raw reads: -Avg. read length: 139 - No. of reads: 568286 -Coverage: 17 (failed! <50x) -Assembled reads -N50: 1508 (< 30000, not acceptable!) -Tot. no of contigs: 4186 (> 1000, not acceptable!).
88				C	QC failed due to low coverage. It was not possible to assemble the genome of this sample as the depth of coverage was lower than allowed by INNUca (<15x for first estimated coverage).
90				C	QC parameters: average genome coverage -15x size of assembled genome - 4.4 Mbp N50 - 1508 total no. of contigs >=200 bases - 4186 cgMLST good targets - 44.6% contamination check (<5% other species) - no evidence of contamination.
100	O80:H2/stx2a	301	No	B	FastQC parameters are relatively bad. GC content is OK. N50 is 11248. Total assembled genome length is 5006224 and the number of contigs is 789. Genome coverage is only 15x. Kmerfinder results show no contamination. % of good cgMLST targets is 95.1%.
108				C	Strain 0018 - Overall low coverage, the QC status was not Ok.
123	O80:H2	301	Yes	B	% good targets E. coli cgMLST: only 93,1 (our threshold: 98%) species match: E. coli (no evidence for contamination) GC content: 50.7 Genome size: only 5,0 Av.Coverage: only 14 (our threshold : 50) N50: only 9692 (our threshold: 50000) Although this strain shows 6 AD to strain 0015 it presumably belongs to the outbreak because the 6 genes difference could be due to the missing 173 values in the analysis (only 93.1 % good targets). Additionally this strain has the same complex type as strain 0015 (CT 23482).

124				C	Strain 18 has a genome size of 5.1 Mb, 52.08 GC%, a low N50 (9728), a high number of contigs (160869), a high number of N bases and an average read coverage of 15 (<30). 61% of alleles were called (<80%).
127				C	Coverage- poor N50-bad No of contigs- too high.
128	O80:H2 / stx2a, stx2d	301	No	B	FASTQC parameters for both reads have OK QC, but QC in Enterobase failed (was not possible to do cgMLST analysis because of that, and coverage was also not determined). GC% for raw data is 52 for both reads, which is good. Most sequences have a quality score around Q30, but the quality is a little lower than, for example, strains 13 and 14 have. Assembled genome has 789 contigs. Total assembly length is 5006224 bp, which is lower than other strains' assemblies. N50 is 11248 (a bit low) and L50 is 132 (too high). GC% for assembled genome is 50.87. KmerFinder shows no contamination. Coverage is 16x.
129				C	Percentage of good targets too low (59.3%), contig count too high (3181), average coverage too low (14).
131				C	cgMLST Perc. Good Targets: 44,6 % (too low) Avg. Coverage (Assembled): 15 x (too low) Approximated Genome Size: 4,4 Mb (smaller than expected) N50 (Assembled): 1508 (too short) Top Species (Match Identity): *Escherichia coli* (0.99) Contig Count (Assembled): 4186 (too high).
132			Yes	C	Was a part of the cluster with 0 SNP difference. Other things were not done.
133				C	Poor Average Quality: >= 30 =TRUE Average Coverage: >=30 =FALSE, 15 N50: >=70,000 =FALSE, 9728 Ncontigs:<500 =FALSE, 1164 NonAGCT:<2500 =FALSE, 3715 Length: 5Mb-5.8Mb =TRUE CorePercent: >96% =FALSE, 62% Therefore, this strain is not acceptable for analysis. It has a very low core (could be a different species), low coverage and it also has a high no of contigs could be an indicator of contamination, since it could be the sum of contigs of two or more organisms. Furthermore, no ST, O group or stx typing could be retrieved by the analysis.
134				C	Coverage = 15 (min = 50) Countig count = 1643 (N= 1000 max) Targets found = 60% (>98 % min) Genome size = 5.2 Coverage is too low and % of targets found not enough to analyse the data.
135				C	Too many contigs, N50 too low, GC% above threshold, coverage too low, contamination too high, % cgMLST alleles too low.
136				C	Low genome size (4 Mb), low genome fraction (70.8%), Very low N50 (5.7 Kbp), High conting number (1336). Indicates DNA fragmentation.
138				C	QC - Fail: 1) contigs >=0bp - 1564 (x ? 605) 2) average coverage - 15.23 (x > 50) 3) GC% - 51.15 (50.3 < x ? 50.9) 4) N50 - 4697 (x > 72925) 5) assembly length - 4878530 (4909000 < x ? 5493000) Insufficient read count which was the reason why 7 gene MLST and 252 of cgMLST alleles could not be determined. Also contamination with *Shigella boydii* was detected based on rMLST results. Based on metadata, reculturing might be required.
139				C	Average genome coverage below acceptable ranges.
153				C	Not enough coverage (15,5x) and to low N50 value. Also many low quality bases detected.
187				C	Low coverage.
222				C	The coverage was 16x, not acceptable because lower than 50x. N50 calculated on the assembled contigs was 9722 bp, which is <30000, used as a threshold for acceptable quality of the assembly.

-: no reported data/analysis performed

Lab ID	Sero	ST	Cluster	QC Status	Description Strain20
EQA provider			**Yes**	**B**	**A Cluster sequence contaminated with approx. 8% *S. sonnei***
19	O80:H2/stx2a	301	Yes	B	The sequence quality is only accepted for outbreak. The strain is slightly contaminated with Shigella, however the strain is still accepted for outbreak. The Bracken analysis report 8 % Shigella. The genome is slightly larger than normal, as is the cgMLST multiple consensus calls and the number of contigs. This also supports that the strain is likely contaminated, however it is a borderline judgement, since it is not unusual that a certain % of Shigella is determined because of the close genetic relatedness between E coli and Shigella. Strain0020 should be restreaked for pure culture and resequenced for confirmation of cluster analysis.
34	O80:H2	301	Yes	A	Good QC, good coverage.

80	O80:H2/stx2a	301	Yes	B	96.1% good targets (cgMLST) -Kraken: E. coli -Raw reads: -Avg. read length: 141 - No. of reads: 2 922 400 -Coverage: 89 -Assembled reads -Size: 5.3 Mb -N50: 22 079 (< 30 000, less good quality!) - Tot. no of contigs: 820.
	O80:H2/stx2a	301	Yes	B	QC failed due to the detection of an intraspecies contamination. INNUca pipeline was able to assemble the reads and remove the contaminated contigs. The final assembly passed all the dowstream QC criteria (coverage and >95% loci called), and, for this reason, was used for outbreak investigation.
90	O80:H2	301	Yes	A	QC parameters: average genome coverage -74x size of assembled genome - 5.3 Mbp N50 - 22079 total no. of contigs >=200 bases - 820 cgMLST good targets - 96.1% contamination check (<5% other species) - no evidence of contamination.
100	O80:H2/stx2a	301	Yes	B	FastQC parameters are relatively OK. Phred score is OK. GC content is OK. N50 is 78726. Total assembled genome length is 5694270 and the number of contigs is 530. Genome coverage is 68x. Kmerfinder results show possible contamination with Shigella (around 2%). This would explain the larger assembled genome size. % of good cgMLST targets is 99,1%.
108	O80:H2/H16/stx2a	301	No	B	Strain 0020 - Overall QC status was Ok aside from the fact that there was contamination of Shigella sonnei
123	O80:H2	301	Yes	A	% good targets E. coli cgMLST: 99,0 (our threshold: 98%) species match: E. coli (no evidence for contamination) GC content: 50,5 Genome size: 5,9 Av.Coverage: 62 (our threshold : 50) N50: 59784 (our threshold: 50000).
124				C	Strain 20 has a genome size of 5.7 Mb, 52.31 GC%, a relatively high number of contigs (722), a high number of N bases and an average read coverage of 79. 99% of alleles were called. The ipaH gene was detected meaning contamination of the sample with Shigella/EIEC (which was confirmed after running Kmer Finder). Both germs are from the same genus which makes analysis complex. The samples cannot be used for cgMLST analysis as both germs are analysed with the same scheme. If no contamination was detected this strain would have been included as part of the identified cluster.
127				C	Probably mixture of STEC (stx2 positive) and Shigella/EIEC (ipaH positive) We need to know if pure colony of E. coli was used for DNA isolation.
128	O80:H2/stx2a, stx2c	301	Yes	B	FASTQC parameters for both reads have OK QC. GC% for raw data is 52 for both reads, which is good. Most sequences have quality score around Q30 or more. Assembled genome has 547 contigs. Total assembly length is 5707211 bp (larger than it should be). N50 is 78726 and L50 is 22. GC% for assembled genome is 50.44. KmerFinder shows 2% contamination with Shigella sonnei. Coverage is 75x.
129		301	Yes	A	Good quality (percentage of good targets 98.8%, contig count 940, average coverage 66)
131	ONT:H2	301	Yes	A	cgMLST Perc. Good Targets: 96,1 % Avg. Coverage (Assembled): 74 x Approximated Genome Size: 5,3 Mb N50 (Assembled): 22079 Top Species (Match Identity): *Escherichia coli* (1.0) Contig Count (Assembled): 820.
132			Yes	C	Was a part of the cluster with four SNP difference. Other things were not done.
133	O80:H2/stx2a	301	Yes	B	Less good quality Average Quality: >= 30 =TRUE Average Coverage: >=30 =TRUE N50: >=70,000 =FALSE, 61,239 Ncontigs:<500 =FALSE, 722 NonAGCT:<2500 =FALSE, 2949 Length: 5Mb-5.8Mb =TRUE CorePercent: >96% =TRUE (100%) This strain failed in 3/6 QC criteria, but for OB analysis core quality is acceptable since clustering is based off of the core genome
134		301	Yes	A	Coverage =70 Contig count = 1 167 (criteria is <1 000, but the other criteria are OK) Targets found = 98.9% Genome size = 5.8
135	O80:H2/stx2a	301	Yes	A	Number of contigs and % contamination elevated but still below thresholds
136	O80:H2	301	Yes	A	All values were in acceptable range
138	O80:H2/stx2a	301	No	B	QC - Warn: 1) contigs >=0bp - 661 (x ? 605) 2) average coverage - 67.59 (x > 50) 3) GC% - 50.52 (50.3 < x ? 50.9) 4) N50 - 61641 (x > 72925) 5) assembly length – 5 767 060 (4909000 < x ? 5493000) Possible contamination with Shigella *sonnei* based on Kraken results. Might be the reason for observed deviations in contig count and assembly length. Based on metadata, reculturing might be required.

139	O80:H2/stx2a	301	Yes	A	Sequence quality within acceptable ranges regarding average phred score of the reads, genome length, N50, number of contigs, and average genome coverage.
153	O80:H2	301	No	A	QC ok.
187	O80:H2/stx2a	301	Yes	A	Complete serotype inferred from known cluster serotypes.
222	O80:H2/stx2a	301	Yes	A	2342/2360 loci of cgMLST scheme correctly mapped, satisfying our quality threshold set at 80% for reliability of cluster analysis. The coverage was 83x. Moreover, the seven genes of conventional MLST (Warwick scheme) were all found 100% in length. N50 calculated on the assembled contigs was >30000, used as a threshold for A of the assembly. All quality criteria were satisfied for this sequence.

-: no reported data/analysis performed

Annex 12. Word format of the online form

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

Please keep in mind that, depending on your answers in the questionnaire, you will not necessarily have to answer all the questions (indicated by the 'Go to').

STEC EQA-12 2023-2024

Dear Participant,

Welcome to the twelfth External Quality Assessment (EQA-12) scheme for typing of STEC in 2023-2024.

NOTE: New virulence gene esta (STa).

If you are using WGS, please read the WGS part of the submission protocol thoroughly before starting your analysis. This year, you are required to use a specific strain/sequence when reporting allele differences/SNP distances.

Please note that most of the fields must be filled in before the submission can be completed. You can write any comments at the end of the form. If you have any questions, please feel free to contact us at ecoli.eqa@ssi.dk.

To begin, please fill in your country, laboratory name, and LAB_ID.

The available options in this participation form include:

- Provide your email to receive a link with your answers. The email containing the link will be sent after pressing 'Finish' on the last slide of the survey.
- Open the windows in full screen for the best survey format.
- If the survey is closed before completion, your answers will be saved, and you can return to the survey using the same link.

Note: After pressing 'Finish' you will not be able to review your results.

1. Country

(State one answer only)

- ❑ Australia
- ❑ Austria
- ❑ Belgium
- ❑ Bulgaria
- ❑ Canada
- ❑ Croatia
- ❑ Czechia
- ❑ Denmark
- ❑ Estonia
- ❑ Finland
- ❑ France
- ❑ Germany
- ❑ Greece
- ❑ Hungary
- ❑ Iceland
- ❑ Ireland
- ❑ Italy
- ❑ Israel
- ❑ Latvia

- ❑ Lithuania
- ❑ Luxembourg
- ❑ Malta
- ❑ México
- ❑ Montenegro
- ❑ New Zealand
- ❑ Norway
- ❑ Paraguay
- ❑ Poland
- ❑ Portugal
- ❑ Romania
- ❑ Scotland, UK
- ❑ Slovakia
- ❑ Slovenia
- ❑ South Africa
- ❑ Spain
- ❑ Sweden
- ❑ The Netherlands
- ❑ Türkiye
- ❑ United Kingdom
- ❑ United States of American

2. Institute name

3. Laboratory name

4. Laboratory ID

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

5. E-mail

6. STEC EQA-12 Strain ID's

Please enter the strain ID (4 digits)
We recommend to print this page out!
To have the overview of strain IDs and strain No. 1-12, it will make the work easier.

STEC
Strain 1 ___
Strain 2 ___
Strain 3 ___
Strain 4 ___
Strain 5 ___
Strain 6 ___
Strain 7 ___
Strain 8 ___
Strain 9 ___
Strain 10 ___

7. Serotyping and virulence gene determination of STEC

8. Submitting results

(State one answer only)

- ❑ Submit serotyping/virulence gene determination results
- ❑ Did not participate in the serotyping nor virulence determination part(s) – Go to 21

9. Submitting results - Serotyping

(State one answer only)

- ❑ Both O group and H type – Go to 10
- ❑ Only O Group – Go to 10
- ❑ Only H type – Go to 12
- ❑ Did not participate in serotyping – Go to 14

10. Results for serotyping (O Group)

Please type the number of O Group by using (1-188)
Non Typable: 7777, Rough: 8888, Not done: 9999

O Group:
Strain 1 ___
Strain 2 ___
Strain 3 ___
Strain 4 ___
Strain 5 ___
Strain 6 ___
Strain 7 ___
Strain 8 ___
Strain 9 ___
Strain 10 ___
Strain 11 ___
Strain 12 ___

11. Please specify the method used:

Phenotypic or molecular (PCR-based, WGS-based)
(State only one answer per question)

Method:

- ❑ Phenotypic
- ❑ PCR-based
- ❑ WGS-based

12. Results for serotyping (H Type)

Please type the number of H Type by using (1-56)
H-: 6666, Non Typable: 7777, Not done: 9999

H type:
Strain 1 ___
Strain 2 ___
Strain 3 ___
Strain 4 ___
Strain 5 ___
Strain 6 ___
Strain 7 ___
Strain 8 ___
Strain 9 ___
Strain 10 ___

Strain 11 ___
Strain 12 ___

13. Please specify the method used:

Phenotypic or molecular (PCR-based, WGS-based)
(State only one answer per question)

Method:

- ❑ Phenotypic
- ❑ PCR-based
- ❑ WGS-based

14. Submitting results – Virulence gene determination

(State only one answer per question)

- ❑ Submit Virulence gene determination data (*eae, aggR, esta* (STa), stx1, stx2 or subtyping)
- ❑ Did not participate in the Virulence gene determination (*eae, aggR, esta* (STa) stx1, *stx2* or subtyping). - Go to 21

15. Please specify the method used for the virulence gene determination (incl. subtyping):

(State only one answer per question)

- ❑ WGS – Go to 17
- ❑ Other – Go to 16

16. If another method is used please describe in detail your method:

17. Results for virulence gene determination

Please use 1 for detected and 0 for not detected, Not done: 9999

	eae	*aaiC*	*aagR*	*stx1*	*stx2*	
Strain 1		___	___	___	___	___
Strain 2		___	___	___	___	___
Strain 3		___	___	___	___	___
Strain 4		___	___	___	___	___
Strain 5		___	___	___	___	___
Strain 6		___	___	___	___	___
Strain 7		___	___	___	___	___
Strain 8		___	___	___	___	___
Strain 9		___	___	___	___	___
Strain 10	___	___	___	___	___	
Strain 11	___	___	___	___	___	
Strain 12	___	___	___	___	___	

18. Submitting results – subtyping results

(State one answer only)

- ❑ Submit subtyping data
- ❑ Did not participate in subtyping – Go to 21

19. Results for subtyping

Subtyping of *stx1,* select variant (*stx1a, stx1c, stx1d*)
All isolates have to be subtyped regardless of the results of the initial screening. 'Not done/ND' will by default be evaluated as an incorrect result.
(State one answer only)

	stx1a	*stx1c*	*stx1d*	*stx1a; stx1c*	*stx1a; stx1d*	*stx1c; stx1d*	Negative	ND
Strain 1	❑	❑	❑	❑	❑	❑	❑	❑
Strain 2	❑	❑	❑	❑	❑	❑	❑	❑
Strain 3	❑	❑	❑	❑	❑	❑	❑	❑
Strain 4	❑	❑	❑	❑	❑	❑	❑	❑
Strain 5	❑	❑	❑	❑	❑	❑	❑	❑
Strain 6	❑	❑	❑	❑	❑	❑	❑	❑
Strain 7	❑	❑	❑	❑	❑	❑	❑	❑
Strain 8	❑	❑	❑	❑	❑	❑	❑	❑
Strain 9	❑	❑	❑	❑	❑	❑	❑	❑
Strain 10	❑	❑	❑	❑	❑	❑	❑	❑
Strain 11	❑	❑	❑	❑	❑	❑	❑	❑
Strain 12	❑	❑	❑	❑	❑	❑	❑	❑

20. Subtyping of *stx2* select variant (stx2a, stx2b, stx2c, stx2d, stx2e, stx2f, stx2g)

All isolates have to be subtyped regardless of the results of the initial screening. 'ND' will by default be evaluated as an incorrect result. (State one answer only)

	stx2a	stx2b	stx2c	stx2d	stx2e	stx2f	stx2g	stx2a; stx2b	stx2a; stx2c	stx2a; stx2d		stx2a; stx2e	stx2a; stx2g	stx2b; stx2c	stx2b; stx2d	stx2b; stx2g	stx2c; stx2d	stx2c; stx2e	stx2c; stx2g	stx2d; stx2e	stx2d; stx2g	stx2e; stx2f	stx2a; stx2b; stx2c	stx2a; stx2c; stx2d	stx2b; stx2c; stx2d	stx2a; stx2b; stx2c; stx2d	Negative	ND
Strain 1	▯	▯	▯	▯	▯	▯	▯	▯	▯	▯		▯	▯	▯	▯	▯	▯	▯	▯	▯	▯	▯	▯	▯	▯	▯	▯	▯
Strain 2	▯	▯	▯	▯	▯	▯	▯	▯	▯	▯		▯	▯	▯	▯	▯	▯	▯	▯	▯	▯	▯	▯	▯	▯	▯	▯	▯
Strain 3	▯	▯	▯	▯	▯	▯	▯	▯	▯	▯		▯	▯	▯	▯	▯	▯	▯	▯	▯	▯	▯	▯	▯	▯	▯	▯	▯
Strain 4	▯	▯	▯	▯	▯	▯	▯	▯	▯	▯		▯	▯	▯	▯	▯	▯	▯	▯	▯	▯	▯	▯	▯	▯	▯	▯	▯
Strain 5	▯	▯	▯	▯	▯	▯	▯	▯	▯	▯		▯	▯	▯	▯	▯	▯	▯	▯	▯	▯	▯	▯	▯	▯	▯	▯	▯
Strain 6	▯	▯	▯	▯	▯	▯	▯	▯	▯	▯		▯	▯	▯	▯	▯	▯	▯	▯	▯	▯	▯	▯	▯	▯	▯	▯	▯
Strain 7	▯	▯	▯	▯	▯	▯	▯	▯	▯	▯		▯	▯	▯	▯	▯	▯	▯	▯	▯	▯	▯	▯	▯	▯	▯	▯	▯
Strain 8	▯	▯	▯	▯	▯	▯	▯	▯	▯	▯		▯	▯	▯	▯	▯	▯	▯	▯	▯	▯	▯	▯	▯	▯	▯	▯	▯

Strain 9																												
Strain 10																												
Strain 11																												
Strain 12																												

21. Submitting Cluster results

(State one answer only)

- ❑ Cluster analyses based on PFGE and/or WGS
- ❑ Did not participate in the Cluster part – Go to 257

22. Submitting Cluster analysis results

(State one answer only)

- ❑ Cluster analysis based on PFGE – Go to 23
- ❑ Do not wish to submit any cluster results based on PFGE analysis – Go to 26

23. Cluster analysis based on PFGE data

24. Please list the ID for the strain included in the cluster of closely related strains detected by PFGE results (bands >33 kb):

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

25. XbaI – Total number of bands (>33kb) in a cluster strain

26. Submitting Cluster results

(State one answer only)

- ❑ Cluster analysis based on WGS data – Go to 27
- ❑ Do not wish to submit any cluster results based on WGS data – Go to 257

27. Cluster analysis based on WGS data

28. Please select the analysis used to detect the cluster using 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 30
- ❑ Allele-based – Go to 37
- ❑ Other – Go to 29

29. If another analysis is used please describe your approach (including: assembler, number of loci, variant caller, read mapper or reference ID, etc.)

– Go to 44

30. Please report the used SNP-pipeline

(reference if publicly available or in-house pipeline)

31. Please select the approach used for the SNP analysis

(State one answer only)

- ❑ Reference-based – Go to 32
- ❑ Assembly-based – Go to 35

32. Reference genome used:

Preferable use EQA strain 0015 (downloaded sequences) as reference. Otherwise indicate Multi-locus Sequence Type (e.g. ST8) and identification of the used reference.

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

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

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

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

37. Please select tools used for the allele analysis

(State one answer only)

- ❑ BioNumerics – Go to 39
- ❑ SeqSphere – Go to 39
- ❑ Enterobase – Go to 39
- ❑ Other – Go to 38

38. If another tool is used please enter here:

39. Please indicate allele calling method:

(State one answer only)

- ❑ Assembly-based and mapping-based – Go to 40
- ❑ Only assembly-based – Go to 40
- ❑ Only mapping-based – Go to 41

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

41. Please select scheme used for the allele analysis

(State one answer only)

- ❑ Applied Maths (wgMLST) – Go to 43
- ❑ Applied Maths (cgMLST/Enterobase) – Go to 43
- ❑ Enterobase (cgMLST) – Go to 43
- ❑ Other – Go to 42

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

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

44. Cluster detected by analysis on data derived from WGS

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

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

45. Strain 1

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

46. (Optional) Report the serotype

47. (Optional) Report Subtype

48. Report the 7-gene MLST
(State value between 0 and 1 000 000)

49. Report if this strain is a part of identified cluster
(State one answer only)

- ❑ Yes
- ❑ No

50. Report the allele difference/SNP distance to the strain 15 (as 0015 downloaded sequence)

Please use 9999 for not analysed.

51. Strain 2
Report the MLST, serotype/group, part of the cluster and SNP distance /allele differences.

52. (Optional) Report the serotype

53. (Optional) Report Subtype

54. Report the 7-gene MLST
(State value between 0 and 1 000 000)

55. Report if this strain is a part of identified cluster
(State one answer only)

- ❑ Yes
- ❑ No

56. Report the allele difference/SNP distance to the strain 15 (as 0015 downloaded sequence)

Please use 9999 for not analysed.

57. Strain 3
Report the MLST, serotype/group, part of the cluster and SNP distance /allele differences.

58. (Optional) Report the serotype

59. (Optional) Report Subtype

60. Report the 7-gene MLST
(State value between 0 and 1 000 000)

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

(State one answer only)

- ❑ Yes
- ❑ No

62. Report the allele difference/SNP distance to the strain 15 (as 0015 downloaded sequence)

Please use 9999 for not analysed.

63. Strain 4

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

64. (Optional) Report the serotype

65. (Optional) Report Subtype

66. Report the 7-gene MLST

(State value between 0 and 1 000 000)

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 15 (as 0015 downloaded sequence)

Please use 9999 for not analysed.

69. Strain 5

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

70. (Optional) Report the serotype

71. (Optional) Report Subtype

72. Report the 7-gene MLST

(State value between 0 and 1 000 000)

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

(State one answer only)

- ❑ Yes
- ❑ No

74. Report the allele difference/SNP distance to the strain 15 (as 0015 downloaded sequence)

Please use 9999 for not analysed.

75. Strain 6

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

76. (Optional) Report the serotype

77. (Optional) Report Subtype

78. Report the 7-gene MLST

(State value between 0 and 1 000 000)

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

(State one answer only)

- ❑ Yes
- ❑ No

80. Report the allele difference/SNP distance to the strain 15 (as 0015 downloaded sequence)

Please use 9999 for not analysed.

81. Strain 7

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

82. (Optional) Report the serotype

83. (Optional) Report Subtype

84. Report the 7-gene MLST

(State value between 0 and 1000000)

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

(State one answer only)

- ❑ Yes
- ❑ No

86. Report the allele difference/SNP distance to the strain 15 (as 0015 downloaded sequence)

Please use 9999 for not analysed.

87. Strain 8

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

88. (Optional) Report the serotype

89. (Optional) Report Subtype

90. Report the 7-gene MLST

(State value between 0 and 1000000)

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

(State one answer only)

- ❑ Yes
- ❑ No

92. Report the allele difference/SNP distance to the strain 15 (as 0015 downloaded sequence)

Please use 9999 for not analysed.

93. Strain 9

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

94. (Optional) Report the serotype

95. (Optional) Report Subtype

96. Report the 7-gene MLST

(State value between 0 and 1 000 000)

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

(State one answer only)

- ❑ Yes
- ❑ No

98. Report the allele difference/SNP distance to the strain 15 (as 0015 downloaded sequence)

Please use 9 999 for not analysed.

99. Strain 10

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

100. (Optional) Report the serotype

101. (Optional) Report Subtype

102. Report the 7-gene MLST
(State value between 0 and 1 000 000)

103. Report if this strain is a part of identified cluster
(State one answer only)

- ❑ Yes
- ❑ No

104. Report the allele difference/SNP distance to the strain 15 (as 0014 downloaded sequence)

Please use 9 999 for not analysed.

105. Strain 11
Report the MLST, serotype/group, part of the cluster and SNP distance /allele differences

106. (Optional) Report the serotype

107. (Optional) Report Subtype

108. Report the 7-gene MLST
(State value between 0 and 1 000 000)

109. Report if this strain is a part of identified cluster
(State one answer only)

- ❑ Yes
- ❑ No

110. Report the allele difference/SNP distance to the strain 15 (as 0015 downloaded sequence)

Please use 9 999 for not analysed.

111. Strain 12
Report the MLST, serotype/group, part of the cluster and SNP distance /allele differences

112. (Optional) Report the serotype

113. (Optional) Report Subtype

114. Report the 7-gene MLST
(State value between 0 and 1 000 000)

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

(State one answer only)

- ❑ Yes
- ❑ No

116. Report the allele difference/SNP distance to the strain 15 (as 0015 downloaded sequence)

Please use 9999 for not analysed.

117. Strain 0013 (as downloaded sequence)

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

118. QC observations

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

119. Please select the QC status that fit 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 126

120. Strain 0013 continue

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

121. (Optional) Report the serotype

122. (Optional) Report Subtype

123. Report the 7-gene MLST

(State value between 0 and 1 000 000)

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

(State one answer only)

- ❑ Yes
- ❑ No

125. Report the allele difference/SNP distance to the strain 15 (as 0015 downloaded sequence)

Please use 9999 for not analysed.

126. Strain 0014 (as downloaded sequence)
Report the MLST, serotype/group, part of the cluster and SNP distance /allele differences.

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

128. Please select the QC status that fit 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 135

129. Strain 0014 continue
Report the MLST, serotype/group, part of the cluster and SNP distance /allele differences

130. (Optional) Report the serotype

131. (Optional) Report Subtype

132. Report the 7-gene MLST
(State value between 0 and 1 000 000)

133. Report if this strain is a part of identified cluster
(State one answer only)

- ❑ Yes
- ❑ No

134. Report the allele difference/SNP distance to the strain 15 (as 0015 downloaded sequence)

Please use 9999 for not analysed.

135. Strain 0015 (as downloaded sequence)
Report the MLST, serotype/group, part of the cluster and SNP distance /allele differences

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

137. Please select the QC status that fit 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 144

138. Strain 0015 continue
Report the MLST, serotype/group, part of the cluster and SNP distance /allele differences.

139. (Optional) Report the serotype

140. (Optional) Report Subtype

141. Report the 7-gene MLST
(State value between 0 and 1 000 000)

142. Report if this strain is a part of identified cluster
(State one answer only)

- ❑ Yes
- ❑ No

143. Report the allele difference/SNP distance to the strain 15 (as 0015 downloaded sequence)

Please use 9999 for not analysed.

144. Strain 0016 (as downloaded sequence)
Report the MLST, serotype/group, part of the cluster and SNP distance /allele differences.

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

146. Please select the QC status that fit 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 153

147. Strain 0016 continue
Report the MLST, serotype/group, part of the cluster and SNP distance /allele differences

148. (Optional) Report the serotype

149. (Optional) Report Subtype

150. Report the 7-gene MLST
(State value between 0 and 1 000 000)

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

(State one answer only)

- ❑ Yes
- ❑ No

152. Report the allele difference/SNP distance to the strain 15 (as 0015 downloaded sequence)

Please use 9999 for not analysed.

153. Strain 0017 (as downloaded sequence)

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

154. QC observations

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

155. Please select the QC status that fit 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 162

156. Strain 0017 continue

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

157. (Optional) Report the serotype

158. (Optional) Report Subtype

159. Report the 7-gene MLST

(State value between 0 and 1000000)

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

(State one answer only)

- ❑ Yes
- ❑ No

161. Report the allele difference/SNP distance to the strain 15 (as 0015 downloaded sequence)

Please use 9999 for not analysed.

162. Strain 0018 (as downloaded sequence)
Report the MLST, serotype/group, part of the cluster and SNP distance /allele differences

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

164. Please select the QC status that fit 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 171

165. Strain 0018 continue
Report the MLST, serotype/group, part of the cluster and SNP distance /allele differences.

166. (Optional) Report the serotype

167. (Optional) Report Subtype

168. Report the 7-gene MLST
(State value between 0 and 1 000 000)

169. Report if this strain is a part of identified cluster
(State one answer only)

- ❑ Yes
- ❑ No

170. Report the allele difference/SNP distance to the strain 15 (as 0015 downloaded sequence)
Please use 9999 for not analysed.

171. Strain 0019 (as downloaded sequence)
Report the MLST, serotype/group, part of the cluster and SNP distance /allele differences.

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

173. Please select the QC status that fit 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 180

174. Strain 0019 continue
Report the MLST, serotype/group, part of the cluster and SNP distance /allele differences

175. (Optional) Report the serotype

176. (Optional) Report Subtype

177. Report the 7-gene MLST
(State value between 0 and 1 000 000)

178. Report if this strain is a part of identified cluster
(State one answer only)

- ❑ Yes
- ❑ No

179. Report the allele difference/SNP distance to the strain 15 (as 0015 downloaded sequence)

Please use 9 999 for not analysed.

180. Strain 0020 (as downloaded sequence)
Report the MLST, serotype/group, part of the cluster and SNP distance /allele differences.

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

182. Please select the QC status that fit 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 189

183. Strain 0020 continue
Report the MLST, serotype/group, part of the cluster and SNP distance /allele differences.

184. (Optional) Report the serotype

185. (Optional) Report Subtype

186. Report the 7-gene MLST
(State value between 0 and 1 000 000)

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

(State one answer only)

- ❑ Yes
- ❑ No

188. Report the allele difference/SNP distance to the strain 15 (as 0015 downloaded sequence)

Please use 9999 for not analysed.

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

e.g. if SNP based results are submitted you can also report allele based results or results from a second SNP analysis (State one answer only)

- ❑ Yes – Go to 190
- ❑ No – Go to 227

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

(State one answer only)

- ❑ SNP-based – Go to 192
- ❑ Allele-based – Go to 199
- ❑ Other – Go to 191

191. If another analysis is used please describe your approach (including: assembler, number of loci, variant caller, read mapper or reference ID ect.)

- Go to 206

192. Please report the used SNP-pipeline

(reference if publicly available or in-house pipeline)

193. Please select the approach used for the SNP analysis

(State one answer only)

- ❑ Reference-based – Go to 194
- ❑ Assembly-based – Go to 197

194. Reference genome used:

(preferable use EQA strain 0015, downloaded sequences as reference). Otherwise indicate Multi-locus Sequence Type (e.g. ST8) and isolate ID

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

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

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

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

199. Please select tool used for the allele analysis

(State one answer only)

- ❑ BioNumerics – Go to 201
- ❑ SeqSphere – Go to 201
- ❑ Enterobase – Go to 201
- ❑ Other – Go to 200

200. If another tool is used please list here:

201. Please indicate allele calling method:

(State one answer only)

- ❑ Assembly-based and mapping-based – Go to 202
- ❑ Only assembly-based – Go to 202
- ❑ Only mapping-based – Go to 203

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

203. Please select scheme used for the allele analysis

(State one answer only)

- ❑ Applied Maths (wgMLST) – Go to 205
- ❑ Applied Maths (cgMLST/Enterobase) – Go to 205
- ❑ Enterobase (cgMLST) – Go to 205
- ❑ Other – Go to 204

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

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

206. Additional analysis on data derived from WGS

207. Results for the additional cluster analysis.

Reporting allele differences/SNP distances to strain 0015 (as downloaded sequence) (e.g. SNP- or Allele-based)
Please use 9999 for not analysed.

	Distance/difference (e.g. SNP/allele) to the strain 0020 (downloaded sequence)
Strain 1	___
Strain 2	___
Strain 3	___
Strain 4	___
Strain 5	___
Strain 6	___
Strain 7	___
Strain 8	___
Strain 9	___
Strain 10	___
Strain 11	___
Strain 12	___
Strain 0013 (as downloaded sequences)	___
Strain 0014 (as downloaded sequences)	___

Strain 0015 (as downloaded sequences) ___
Strain 0016 (as downloaded sequences) ___
Strain 0017 (as downloaded sequences) ___
Strain 0018 (as downloaded sequences) ___
Strain 0019 (as downloaded sequences) ___
Strain 0020 (as downloaded sequences) ___

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

e.g. if SNP-based results are submitted you can also report allele-based results or results from an additional SNP analysis
(State one answer only)

- ❑ Yes – Go to 209
- ❑ No – Go to 227

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

(State one answer only)

- ❑ SNP-based – Go to 211
- ❑ Allele-based – Go to 218
- ❑ Other – Go to 210

210. If another analysis is used please describe your approach:

– Go to 225

211. Please report the used SNP-pipeline

Reference if publicly available or in-house pipeline.

212. Please select the approach used for the SNP analysis

(State one answer only)

- ❑ Reference-based – Go to 213
- ❑ Assembly-based – Go to 216

213. Reference genome used:

(preferable use EQA strain 0015, downloaded sequences as reference). Otherwise indicate Multi-locus Sequence Type (e.g. ST8) and isolate ID

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

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

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

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

218. Please select tool used for the allele analysis

(State one answer only)

- ❑ BioNumerics – Go to 220
- ❑ SeqSphere – Go to 220
- ❑ Enterobase – Go to 220
- ❑ Other – Go to 219

219. If another tool is used please enter here:

220. Please indicate allele calling method:

(State one answer only)

- ❑ Assembly-based and mapping-based – Go to 221
- ❑ Only assembly-based – Go to 221
- ❑ Only mapping-based – Go to 222

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

222. Please select scheme used for the allele analysis

(State one answer only)

- ❑ Applied Maths (wgMLST) – Go to 224
- ❑ Applied Maths (cgMLST/Enterobase) – Go to 224
- ❑ Enterobase (cgMLST) – Go to 224
- ❑ Other – Go to 223

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

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

225. Third analysis on data derived from WGS

226. Results for the third cluster analysis

Reporting allele differences/SNP distances to strain 0015 (as downloaded sequence) (e.g. SNP- or Allele-based)
Please use 9999 for not analysed

	Distance/difference (e.g. SNP/allele) to the strain 0020 (downloaded sequence)
Strain 1	___
Strain 2	___
Strain 3	___
Strain 4	___
Strain 5	___
Strain 6	___
Strain 7	___
Strain 8	___
Strain 9	___
Strain 10	___
Strain 11	___
Strain 12	___
Strain 0013 (as downloaded sequences)	___
Strain 0014 (as downloaded sequences)	___
Strain 0015 (as downloaded sequences)	___
Strain 0016 (as downloaded sequences)	___
Strain 0017 (as downloaded sequences)	___
Strain 0018 (as downloaded sequences)	___
Strain 0019 (as downloaded sequences)	___
Strain 0020 (as downloaded sequences)	___

227. Additional questions to the WGS part

228. Where was the sequencing performed

(State one answer only)

- ❑ In own laboratory
- ❑ Externally

229. Protocol used to prepare the library for sequencing:

(State one answer only)

- ❑ Commercial kits – Go to 230
- ❑ Non-commercial kits – Go to 232

230. Please indicate name of commercial kit:

231. If relevant please list deviation from commercial kit shortly in few bullets:

- Go to 233

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

233. The sequencing platform used

(State one answer only)

- ❑ Ion Torrent PGM – Go to 235
- ❑ Ion Torrent Proton – Go to 235
- ❑ Genome Sequencer Junior System (454) – Go to 235
- ❑ Genome Sequencer FLX System (454) – Go to 235
- ❑ Genome Sequencer FLX+ System (454) – Go to 235
- ❑ PacBio RS – Go to 235
- ❑ PacBio RS II – Go to 235
- ❑ HiScanSQ – Go to 235
- ❑ HiSeq 1000 – Go to 235
- ❑ HiSeq 1500 – Go to 235
- ❑ HiSeq 2000 – Go to 235
- ❑ HiSeq 2500 – Go to 235
- ❑ HiSeq 4000 – Go to 235
- ❑ Genome Analyzer Iix – Go to 235
- ❑ MiSeq – Go to 235
- ❑ MiSeq Dx – Go to 235
- ❑ MiSeq FGx – Go to 235
- ❑ ABI SOLiD – Go to 235
- ❑ NextSeq – Go to 235
- ❑ MinION (ONT) – Go to 235
- ❑ Other – Go to 234

234. If another platform is used please list here:

235. Criteria used to evaluate the quality of sequence data

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

Please first reply on the use of 5 selected criteria, which were the most frequently reported by in previous EQAs.

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

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

236. Did you use confirmation of organism to evaluate the quality of sequence data?

- ❑ Yes
- ❑ No – Go to 238

237. Procedure used to evaluate confirmation of organism:

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

(State one answer only)

- ❑ Yes
- ❑ No – Go to 240

239. Procedure or threshold used for coverage:

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

(State one answer only)

- ❑ Yes
- ❑ No – Go to 242

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

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

(State one answer only)

- ❑ Yes
- ❑ No – Go to 244

243. Procedure or threshold used for genome size:

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

(State one answer only)

- ❑ Yes
- ❑ No – Go to 246

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

246. 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).

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

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

249. Other criteria used to evaluate the quality of sequence data – additional criteria 2:

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

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

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

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

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

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

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

257. Comment(s):

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

258. You have reached the end of the reporting scheme.

Please note that when you select 'Yes' and 'Next', your results will be automatically submitted and the reporting form will be locked.

If you wish to change your answers, use 'Previous' to navigate backwards.
Upon completion, you will receive a link with your answers.

❑ Yes

Thank you for your participation

Thank you for filling out the Submission form for the STEC EQA-12.

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

Remember to press 'Finish' to complete submission.

After submission you will receive a confirmation email with a link to the answers. We highly recommend to save this email.

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

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