ecdc
EUROPEAN CENTRE FOR DISEASE PREVENTION AND CONTROL

TECHNICAL REPORT

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

www.ecdc.europa.eu

ECDC TECHNICAL REPORT

Eleventh 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. Eleventh external quality assessment scheme for typing of Shiga toxin-producing *Escherichia coli*. Stockholm: ECDC; 2023.

Stockholm, October 2023

ISBN 978-92-9498-660-3
DOI 10.2900/738691
Catalogue number TQ-07-23-395-EN-N

Contents

Figures

Tables

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*
EURL	European Union Reference Laboratory
FWD	Food- and waterborne diseases
FWD-Net	Food- and Waterborne Diseases and Zoonoses Network
GFN	Global Foodborne Infections Network, Food Safety
HUS	Haemolytic uraemic syndrome
ND	Not done
NPHRL	National public health reference laboratory
NSF	Non-sorbitol fermenter
NT	Non-typeable
PCR	polymerase chain reaction
PFGE	Pulsed-field gel electrophoresis
QC	Quality control
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 11th round of the external quality assessment (EQA-11) 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-11 contains serotyping, detection of virulence genes, and molecular typing-based cluster analysis.

Human STEC infection is a zoonotic disease. For 2021, 6 534 confirmed cases of STEC infection were reported by 30 EU/EEA countries. Twenty-seven countries reported at least two confirmed cases, and three countries reported no cases. The EU/EEA notification rate was 2.2 cases per 100 000 population, representing a 37.5% increase compared with the previous year. The five most frequently reported serogroups were O157 (15.1%), O26 (14.7%), O103 (8.4%), O145 (4.6%), and O146 (3.7%).

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 either pulsed-field gel electrophoresis (PFGE) and/or whole genome sequencing (WGS)-derived data has been included since EQA-8.

The objectives of the EQA 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. Eight additional strains (sequences) were included for the molecular typing-based cluster analysis. Twenty-seven laboratories registered and 26 completed the exercise, comparable to EQA-10.

The full O:H serotyping was performed by 73% (19/26) of participating laboratories, with an average score of 95%. In general, the more common European serotypes generated the highest scores compared to the less common serotypes, such as O187:H28, O80:H2, and O91:H14, 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 low (19/26), although higher than EQA-10 (16/19), most likely reflecting a shift towards WGS-based methods. This could also be seen in the reported O-grouping results, where 60% (15/22) used WGS-based methods, which is higher than 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 97%, 99%, and 96% for *eae, stx1*, and *stx2*, respectively, similar to previous EQAs.

In EQA-11, two other diarrhoeagenic *E. coli* (DEC) pathotypes were included – EAEC Strain 12 (*aggR* gene) and ETEC Strain 2 (*esta* gene) – testing the participating laboratories in their abilities to detect STEC hybrid strains (Strain12 and Strain2). The performance of detection of the *aggR* genes was high (21/22, 95%) which is comparable to EQA-10, where 94% correctly identified *aggR*. Likewise, the performance for *esta* was also high (89%), where only two laboratories reported a false negative result. Both laboratories used other methods than WGS.

Of the 26 laboratories participating in EQA-11, 20 (77%) performed molecular typing-based cluster analysis using WGS data analysed by different approaches. Notably, all laboratories used WGS in EQA-11 and none chose PFGE, a decrease from EQA-10 (two laboratories) and EQA-9 (eight 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.

Sixteen participants (80%) correctly identified the cluster of five closely related ST11 strains defined by pre-categorisation from the EQA provider among the 12 test strains and eight test strains (genomic sequences).

Also 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; 80% (16/20) used core genome MultiLlocus Sequence Type (cgMLST) compared to 20% (4/20) 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 between laboratories difficult. The latter was reflected in the reported results, as 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 in order to give a realistic view of different quality issues. Most of the participants (17/20) successfully identified the 14% contamination with *E. albertii*, and the poor quality for one genome was observed by all 20 laboratories. Both contamination with a different species and poor quality are important to assess before the analysis of WGS.

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; 15/26 responded. The usefulness of the QC evaluation of the participant-sequenced data and the usefulness of including low quality data were appreciated by 86–93% of respondents.

1 Introduction

1.1 Background

ECDC is a European Union (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 [1].

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 [2] 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 Statens Serum Institut (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-11.

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.

For 2021, 6 534 confirmed cases of STEC infection were reported by 30 EU/EEA countries. Twenty-seven countries reported at least two confirmed cases, and three countries reported no cases. The EU/EEA notification rate was 2.2 cases per 100 000 population, representing a 37.5% increase compared with the previous year [3].The five most frequently reported serogroups were O157 (15.1%), O26 (14.7%), O103 (8.4%), O145 (4.6%), and O146 (3.7%) [3]. These serogroups together accounted for 46.5% of the total confirmed STEC cases with known serogroups in 2021. Among HUS cases, serogroup O26 was most commonly reported (34%) followed by O157 (19.8%). The proportion of cases where no serotype could be retrieved was 25.9%, which was an increase on previous years. Antigen H was reported for 2 496 confirmed cases (38.2%).

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 Member States to The European Surveillance System (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/VTEC. 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 [1] [4] *stx2d* subtype in *eae*-negative STEC appear to be highly associated with the serious sequela HUS [5–7]. 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 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-11 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 represent 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-11 on STEC

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

As a result, and as part of the recommendations in EQA-10, the EQA provider removed the *aaiC* gene in EQA-11. 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-11 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-11 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, stx2a, stx2b, stx2c, stx2d, stx2f, and stx2g*).

1.4.3 Molecular typing-based cluster analysis

The objective of the molecular typing-based cluster analysis of STEC EQA-11 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.

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

2 Study design

2.1 Organisation

STEC EQA-11 was funded by ECDC and arranged by SSI following ISO/IEC 17043:2010 [9]. EQA-11 included serotyping, virulence gene determination, and a molecular typing-based cluster analysis, and was carried out between May to December 2022.

Invitations were emailed by ECDC to the contact points in the FWD-Net (30 countries) by 7 April 2022, with a deadline to respond by 1 May 2022. In addition, invitations were sent to the EU candidate countries Albania, Bosnia and Herzegovina, Kosovo[2], Montenegro, North Macedonia, Serbia, and Türkiye.

Twenty-seven NPHRLs in EU/EEA and EU candidate countries accepted the invitation to participate, and 26 submitted results (Annex 1). EQA test strains were sent to participants from 1 June to 5 June 2022. In Annex 2, participation details in EQA-10 and EQA-11 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 2022 (Annex 12). Six laboratories were asked to submit the missing raw reads/upload them again.

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

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 17 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).

Unlike previous EQAs, we have included two hybrid *E. coli* pathotype test strains; Shiga toxin-producing and enterotoxigenic *E. coli* (STEC/ETEC) 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 present a major challenge for the public health, due to the needs to now implement diagnostic procedures that will identify the most virulent clones.

Based on the WGS-derived data, the selected cluster of closely-related strains consisted of five STEC ST11 strains (including the technical duplicate set strain3/strain7 and provided sequence strain20). Characteristics of all the STEC test strains are listed in Table 1 and Annexes 3–9. The EQA provider found at most two allele differences or five 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 [10]) and SNP analysis (NASP [11]). 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. Eight additional strains (sequences) for cluster analysis were selected to include strains with different or varying relatedness and different sequence types (ST11). A set of triplicates were included in the test strains (strain3 and strain7) and the sequence strain20 (Annexes 5–7, 9–10). Two of the sequences were modified by the EQA provider, one with reduced coverage and one contaminated with 14% *Escherichia albertii.* The characteristics of all the strains and sequences are listed as 'EQA provider' in Annexes 4–10.

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

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	O26:H11	Strains for virulence profile	*stx1a, stx2a, eae*	Strains/sequences for cluster analysis	21	-	
Strain2		O187:H28		*stx2g, esta*		200	-	
Strain3#‡		O157:H-/H7		*stx1a, eae*		11	-	Yes
Strain4		O177:H-/H25		*stx2a, stx2c**		342	-	
Strain5		O91:H14		*stx1a, stx2b*		33	-	
Strain6		O80:H2		*stx2d, eae*		301	-	
Strain7#‡		O157:H-/H7		*stx1a, eae*		11	-	Yes
Strain8		O157:H-/H7		*stx1a, stx2c, eae*		11	-	
Strain9		O128:H-/H2		*stx2f, eae*		20	-	
Strain10		O145:H-/H28		*stx2a, eae*		32	-	
Strain11		O146:H21		*stx1c, stx2b*		442	-	
Strain12		O104:H4		*aggR*		678	-	
Strain13^	-	O157:H7		*stx1a, eae*		11	C	NA
Strain14	-	O157:H7		*stx1a, stx2c, eae*		11	A	
Strain15‡	-	O157:H7		*stx1a, stx2c, eae*		11	A	Yes
Strain16	-	O157:H7		*stx1a, stx2c, eae*		11	A	
Strain17‡	-	O157:H7		*stx1a, stx2c, eae*		11	A	Yes
Strain18	-	O157:H7		*stx1a, stx2c, eae*		11	A	
Strain19^	-	-		-		-	B/C	NA
Strain20#‡	-	O157:H7		*stx1a, eae*		11	A	Yes

‡: closely related strains.
#: technical triplicates strains.
ST: sequence type.
^modified sequences: strain13, a non-cluster sequence with reduced coverage and removal of genes, strain19, a non-cluster sequence contaminated with approx. 14% E. albertii.
NA: Not applicable.
A: Acceptable quality, B: Quality only acceptable for outbreak situations (less good quality) and C: Not acceptable quality – strain not analysed.
** the EQA provider excluded the stx subtyping results as reported by participants. See text for explanation.*

2.3 Distribution of strains and sequences

The 12 test strains were blinded and shipped from 1–9 June 2022 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. The packages were shipped from SSI, labelled 'UN3373 Biological Substance'. Five participants received the strains within three days, 13 within seven and eight days, eight within nine to 10 days, and one after 14 days following shipment, respectively. No participants reported damage to the shipment or errors in the unique strain IDs. We did have one laboratory whose institution had misplaced the strains, but these were found again.

In July 2022, instructions for the submission of results procedure were emailed to the participants. These 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 [12] 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, stx2a, stx2b, stx2c, stx2d, stx2f, and stx2g*) according to suggested protocol [13]. 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-11, 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 (strain20) selected by the EQA provider.

In addition, each participant needed to assess the QC of the provided sequences (two 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 two sequences (strain13 and strain19). See 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-11 BioNumerics (BN) database. The EQA provider contacted six participants in order to ensure sequences were uploaded to the SFTP site.

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 [10] and SNP analysis (NASP) [11]. 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 cluster contained five ST11 strains/sequences: strain3, strain7, strain15, strain17, and strain20 (strain3/7 and strain/sequence20 were technical triplicates). The EQA provider found at most two allele differences or five SNPs between any two strains in the cluster.

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

Individual evaluation reports were distributed to participants in December 2022 and certificates of attendance in February 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. More than three-quarters of the participants (77%; 20/26) completed all three parts of the EQA-11 (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 20 (77%) 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	20
Percent of participants	96*	96*	77*

[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 19 (73%). Almost two-thirds, 16/25 (64%), used molecular-based serotyping (only one reported PCR-based method) (Annex 4). The majority of the participants (96%, 25/26) performed the detection of virulence genes *stx1* and *stx2*. Slightly fewer 92% (24/26) participated in the detection of *eae*. Detection of the enteroaggregative gene, *aggR,* and the ETEC gene, *esta*, were reported by 85% (22/26) and 73% (19/26). Additionally, the *stx* subtyping detection of were reported by 85% (22/26) for *stx1* and stx2 85% (22/26) (Annex 4). The majority of the participants performed the cluster analyses (77%, 20/26), 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		Virulence profile determination					Cluster analysis
	n=25		n=25					n=20
	O group	H type	*aggR*	*eae*	*esta*	*stx1* and *stx2*	*stx* subtyping	WGS
Number of participants	25[#]	19Δ	22	24	19	25	22	20
Percentage of participants^	100%	76%	88%	96%	76%	100%	88%	100%
Percentage of participants *	96%	73%	85%	92%	73%	96%	85%	77%

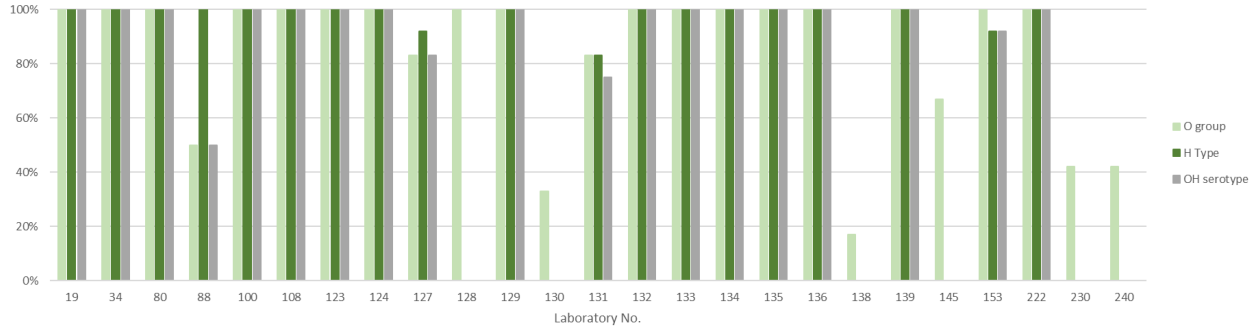
^: percentage of participants in respective part of EQA.
**: percentage of total number of participating laboratories (26).*
[#]: phenotypic (n=9)/PCR-based (n=1)/WGS-based (n=15).
Δ: phenotypic (n=2)/PCR-based (n=1)/WGS-based (n=16).

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 five laboratories had a score of ≤ 50%, giving an average score of 85% (Figure 1). Eighteen laboratories (69%) reported the correct O group for the uncommon O group O187 (strain2) and 19 (73%) correctly reported O80 (Strain6) (Figure 2). The highest performances were obtained for the O26 (96%), O145 (96%), and O157 (96%) positive strains (Figure 2).

Nineteen (73%) laboratories performed H typing. Of the 25 laboratories participating in O grouping, 76% (19/25) also reported H type. The general performance for H typing was higher than O grouping, with the majority (84%; 16/19) of participants correctly H typing all 12 test strains, resulting in an average score of 98% (Figure 1). In six out of the 12 strains reporting H- was accepted as a correct result when using phenotypical H-typing as these strains we non-motile. One laboratory (127) reported H42 instead of H28, laboratory 131 reported two incorrect H types (NT and H- instead of H25 and H14, respectively), likewise, laboratory 153 also reported H- instead of H14 (Annex 3).

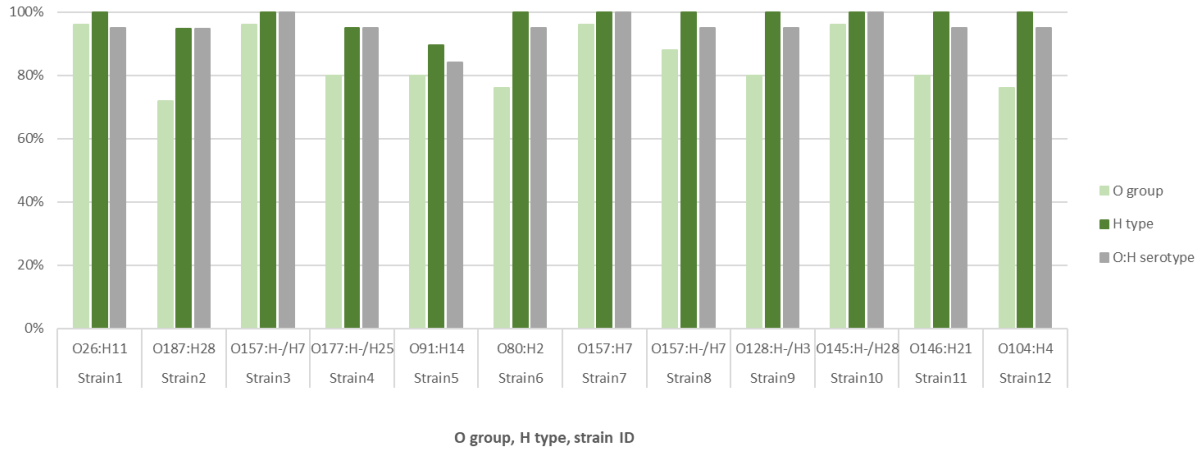
Figure 1. Participant percentage scores for O grouping and H typing



Arbitrary numbers represent participating laboratories.
Bars represent the percentage of correctly assigning O groups (light green), n= 25 participants, H types (dark green), n=19 participants, Combined O:H serotypes (grey), n=19 participants.

Complete O:H serotyping was performed by 19 (73%) participants with an average score of 95%, and for each strain the score ranged from 89% (17/19) for Strain5 (O91:H14) to 100% (19/19) for Strain3 (O157:H-/H7), Strain7 (O157:H-/H7), and Strain10 (O145:H-/H28). The correct serotype of all 12 strains were reported by 84% (16/19) of the participants who performed the O:H serotyping (Figure 2, Annex 3).

Figure 2. 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=19 participants. Combined O:H serotypes (grey): n=19 participants.
Average scores: O group, 85%; H type, 98% and combined O:H serotype, 95%.

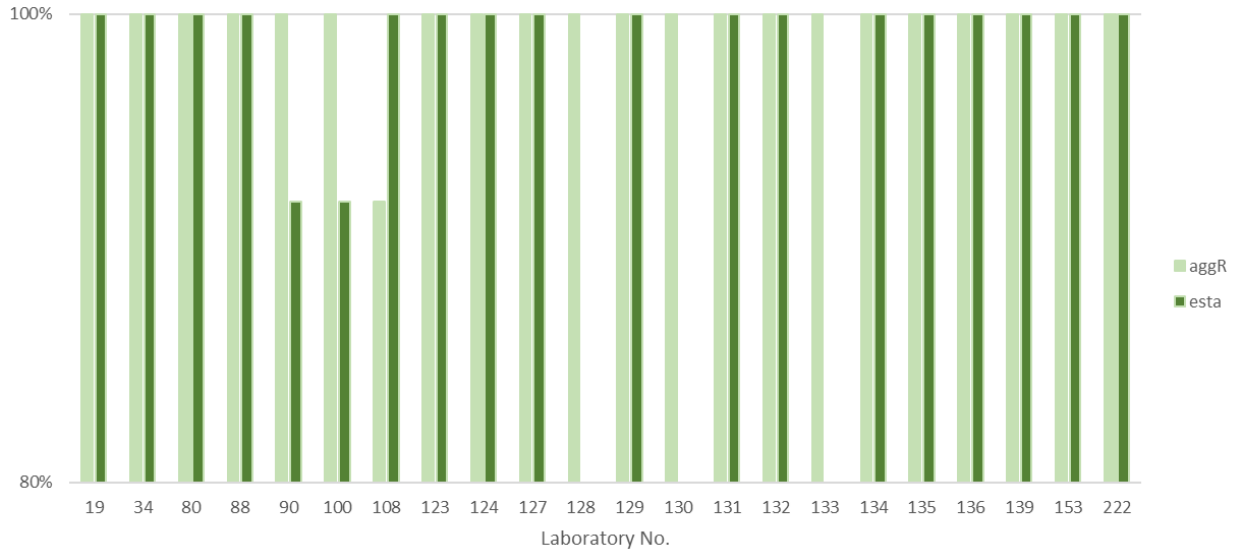
3.3 Virulence profile determination

Between 19 and 25 laboratories submitted results for some, or all, of the following virulence genes; *aggR* (22 participants), *eae* (24 participants), *esta* (19 participants), *stx1* (25 participants), *stx2* (25 participants), and subtyping of *stx1* (22 participants), and *stx2* (22 participants).

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

Among the strains in EQA-11 two test strains harboured other pathotype-defining virulence genes; strain2 harbouring the ETEC associated *esta* gene and strain12 harbouring the EAEC defining gene *aggR*. All laboratories, except for one (108), correctly identified *aggR* in strain12, as such the performance for *aggR* was 95% (21/22). The performance for *esta* was 89% corresponding to two laboratories (90 and 100) that didn't identify the gene in strain2 (Figure 3, Annex 4).

Figure 3. Participant percentage scores for genotyping of *aggR* and *esta*

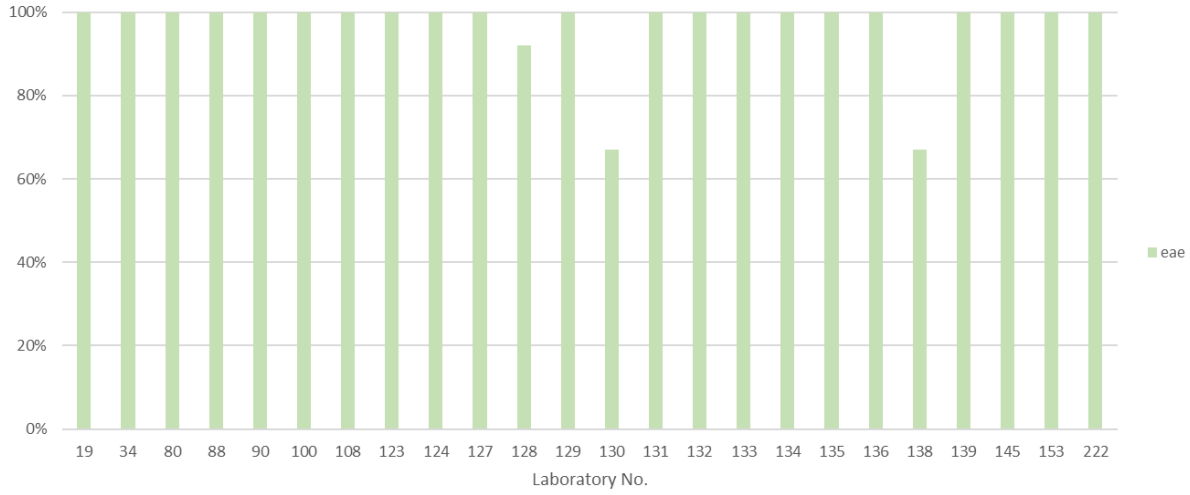


Arbitrary numbers represent participating laboratories.
Bars represent percentage of correct genotyping of aggR (light green) n=22 participants and esta (dark green): n=19 participants.

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

Detection of virulence genes *eae*, *stx1* and *stx2* was performed by 24-25 (92-96%) laboratories with a generally high performance (Figures 4–5). For *eae* detection, 21 (88%) laboratories obtained a 100% score and three laboratories (128, 130, and 138) reported incorrect results for the *eae* gene (Figure 4). Eight of the nine incorrect results were false negative found by laboratories 130 and 138 in multiple strains. Laboratory 128 reported the only false positive in strain12 (Annex 4).

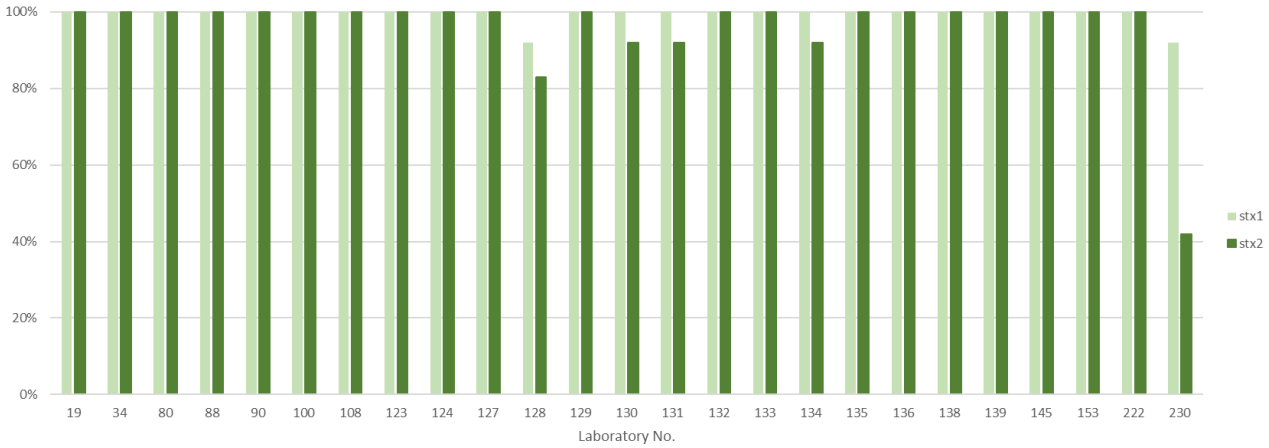
Figure 4. Participant percentage scores for genotyping of *eae*



Arbitrary numbers represent participating laboratories.
Bars represent percentage of correct genotyping of eae *(light green): n=24 participants.*

The performance for the detection of *stx1* and *stx2* genes was high for *stx1*; 23 (92%) laboratories reported 100% correct *stx1* results and 20 (80%) laboratories reported 100% correct *stx2* results (Figure 5). The two incorrect results for *stx1* were reported by laboratories 128 and 230 and were both false positive. There were 12 incorrect results reported for *stx2* were 11 of them were false negative. Mainly laboratory 230 reported seven of the incorrect *stx2* results. Four laboratories (128, 130, 131, and 230) reported a false negative result for strain9. The one false positive result was reported by laboratory 128 in strain12 (Annex 4).

Figure 5. 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, 99%; stx2, 96%.

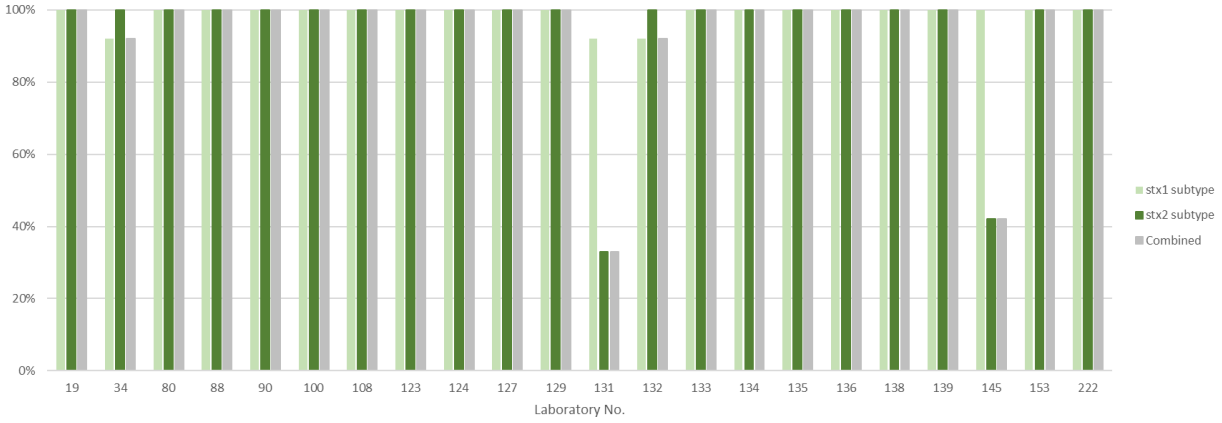
3.3.3 Subtyping of *stx1* and *stx2*

Subtyping of *stx1* and *stx2* was performed by 22 laboratories. The results of strain4 was disregard as the identification of *stx2a* and *stx2c* in general is very difficult to identify and the EQA provider reanalysed the EQA provided sequences and found small changes in the quality of sequences impacted the result. Only 32% (7/22) of the participants found *stx2a* and *stx2c.* As such, the EQA provider has decided to exclude results for this strain. Nineteen laboratories subtyped *stx1* correctly and *stx2* correctly (86%; 19/22) for all 11 test strains (disregarding strain4) (Figure 6; Annex 4). Correctly reported *stx1* and *stx2* subtyping of the 11 strains was reported by 77% of the laboratories (17/22).

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.

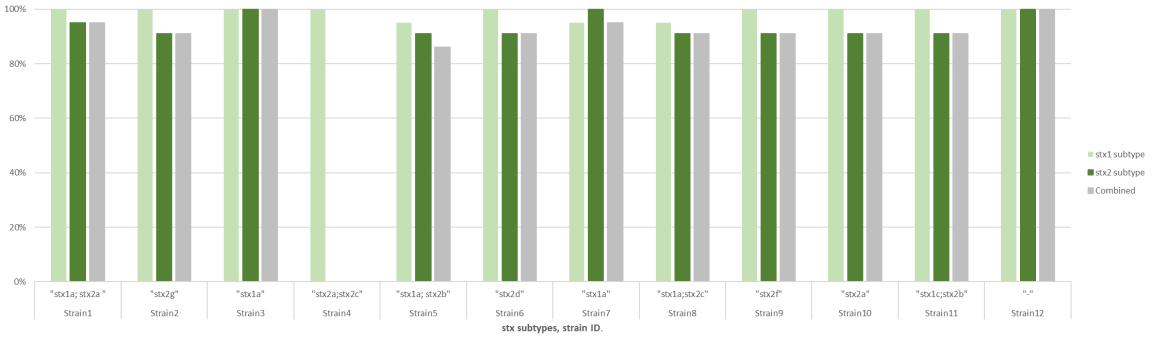
Only four laboratories (18%) reported an incorrect subtyping of either/or both *stx1* and *stx2*. For the *stx1* subtyping one laboratory (34), reported a false negative for strain7 where one laboratory (131) correctly identified the *stx1a* subtype for strain8 but in-correctly also reported an additional stx1 subtype (*stx1c*). As such, the average score of the 11 test strains were 99% for the *stx1* subtyping. The average score for the *stx2* subtyping was 94 % with the majority of the mis-subtyping *stx2* and reported results could be attributed to two laboratories 131 and 145 corresponding to 33% and 42% correct stx2 subtyping, respectively. The average score of the combined subtyping was 93%.

Figure 6. 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=22 participants. Reporting ND (not done) evaluated as incorrect.

Figure 7. 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=22.
Average scores: stx1, 99%; stx2, 94% and combined stx1 and stx2, 93%.

Two laboratories (132 and 145) 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 three categories: false negatives (1/9), incorrect subtype of *stx2* (8/9) or ND (6).

Table 4. Incorrect *stx2* subtype results

		Incorrect subtype results			
Strain ID	**EQA provider**	**False negative**	**Incorrect**	**Total true errors**	**Errors by reporting ND#**
Strain1	*Stx2a*		*stx2a; stx2c* (1)	1	
Strain2	*Stx2g*		*stx2a; stx2b* (1)	1	1
Strain3	-				
Strain4*	*Stx2a; Stx2c*				
Strain5	*Stx2b*		*stx2a* (2)	2	
Strain6	*Stx2d*		*stx2a; stx2c; stx2d* (1)	1	1
Strain7	-				
Strain8	*Stx2c*		*stx2a; stx2c; stx2d (1)*	1	1
Strain9	*Stx2f*	1		1	1
Strain10	*Stx2a*		*stx2a; stx2c* (1)	1	1
Strain11	*Stx2b*		*stx2a* (1)	1	1
Strain12	-				
Total				**9**	**6**

ND#: not done.
** disregarded strain4 – see text for explanation.*

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* ST11, based on WGS-derived data (Tabel1). The EQA provider's cluster analysis of WGS-derived data was based on an allele-based (cgMLST [10]) and SNP analysis (NASP [11]).

The correct cluster based on WGS-derived data contained five ST11 strains: strain3, strain7, strain15, strain17 and strain20 (strain3/strain7 and strain20 (sequence) were technical triplicates). The EQA provider found at most two allele differences or five 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 participants (77%) performed cluster analysis using WGS-derived data. One laboratory reported using external assistance for sequencing. Different sequencing platforms were listed among the participants: 1 MiniSeq, 8 MiSeq, 7 NextSeq, 2 Novaseq, 2 Ion Torrent (Ion GeneStudio S5 System and Ion Torrent S5XL). All laboratories reported using commercial kits for library preparation. Of the 20 participants, 16 (80%) used Illumina's Nextera kit. Four participants reported changes from the manufacturer protocol, three in the volume, and one in the shearing time (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 two sequences (strain13 and strain19), one with contamination and one with massively reduced coverage (Table 5).

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

For strain13,a non-cluster sequence with reduced coverage and removal of genes, 80% (16/20) of the participants correctly identified the poor quality of the sequence, and excluded the sequence from the cluster analysis. Additional four participants accepted the quality for outbreak investigation.

For strain19, a non-cluster sequence contaminated with approx. 14% *E. albertii*, 85% (17/20) correctly observed the contamination of the sequence and reported a QC status of B or C. Three participants reported the sequence to have acceptable quality (QC status A), but one of the three described the contamination. Four participants included the strain/sequences in their cluster analysis and 13 discarded the strain/sequences from the analysis.

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

Genome	Characteristics	Provider	A	B	C
Strain13	A non-cluster sequence with massive reduced coverage and removal of genes	C	0	4	16
Strain19	A non-cluster sequence contaminated with approx. 14 *% E. albertii*	B/C	3	4	13

Raw data available in Annex 11.

3.4.1.3 Cluster analysis

Each participant should use their own produced sequences and the provided sequences (after assessment of QC status) in the cluster analysis and report which strains/sequences were a part of the cluster of closely related strains, thereby mimicking an urgent outbreak situation where it is impossible to rerun the sequence and the sequences must be assessed despite poor quality, etc.

Performance was high in cluster analysis with WGS-derived data. Sixteen participants (80%) correctly identified the cluster of closely related strains 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 strain20 (sequence) as a representative in the cluster for reporting SNP distance or allelic differences. Laboratories could report results from 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	-	+	-	+	-	ND	+	A[a]	+
34	-	-	+	-	-	-	+	-	-	-	-	-	ND	-	+	-	+	-	-	+	A	+
80	-	-	+	-	-	-	+	-	-	-	-	-	ND	-	+	-	+	-	ND	+	A	+
88	-	-	+	-	-	-	+	-	-	-	-	-	-	-	+	-	+	-	-	+	A[c]	+
90	-	-	+	-	-	-	+	-	-	-	-	-	ND	-	+	-	+	-	-	+	A	+
100	-	-	+	-	-	-	+	-	-	-	-	-	ND	-	+	-	+	-	ND	+	A[b]	+
108	-	-	-	-	-	-	-	-	-	-	-	-	ND	-	+	-	-	-	-	+	S	No
123	-	-	+	-	-	-	+	-	-	-	-	-	-	-	+	-	+	-	ND	+	A	+
124	-	-	+	-	-	-	+	-	-	-	-	-	-	-	+	-	+	-	ND	+	A	+
129	-	-	+	-	-	-	+	-	-	-	-	-	ND	-	+	-	+	-	-	+	A	+
131	-	-	+	-	-	-	+	-	-	-	-	-	ND	-	+	-	+	-	-	+	A	+
132	-	-	+	-	-	-	+	+	-	-	-	-	-	+	+	+	+	+	ND	+	S	No
133	-	-	+	-	-	-	+	-	-	-	-	-	ND	-	-	-	-	-	-	+	A	No
134	-	-	+	-	-	-	+	-	-	-	-	-	ND	-	+	-	+	-	ND	+	A	+
135	-	-	+	-	-	-	+	-	-	-	-	-	ND	-	+	-	+	-	ND	+	A	+
136	-	-	+	-	-	-	+	-	-	-	-	-	ND	-	+	-	+	-	ND	+	A	+
138	-	-	+	-	-	-	-	+	-	-	-	-	ND	+	+	+	+	+	ND	+	S	No
139	-	-	+	-	-	-	+	-	-	-	-	-	ND	-	+	-	+	-	ND	+	A	+
153	-	-	+	-	-	-	+	-	-	-	-	-	ND	-	+	-	+	-	ND	+	S	+
222	-	-	+	-	-	-	+	-	-	-	-	-	ND	-	+	-	+	-	ND	+	A	+

‡: closely related strains (in grey). #: technical duplicates strains (in bold). 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 (Annex 7).

Four participants (108, 132, 138, and 153) used SNP as their main analyses, and two laboratories reported SNP as an additional analysis. All except laboratory 132 used a reference-based approach with the EQA strain20 as reference. As the read mapper, one laboratory used Burrows-Wheeler Aligner (BWA), one Bowtie2, one used CLC mapper and the last one CSI Phylogeny. Four different variant callers were used (Table 7). Note that of these four laboratories, only laboratory 153 correctly identified the cluster.

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	Strain20	BWA	GATK	Yes	0–5	30-127
19*	NASP	Rb	Strain20	BWA	GATK	Yes	0-4	37-79
100*	SNV in SepSphere	Rb	Strain20	-	-	Yes	0-44	78-80845
108	In-house pipeline	Rb	Strain20	CLC mapper	CLC find variation	No	0-4	12-80845
132	CSIPhylogeny	Rb	Strain4	BWA	SAMtools	No	0-16	7031-8767
138	CFSAN SNP Pipeline	Rb	Strain20	Bowtie2	GATK	No	0-5	11-53
153	CSI Phylogeny	Rb	Strain20	CSI Phylogeny	CSI Phylogeny	Yes	0-1	19-9325

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

Sixteen participants used allele-based analysis as the main analysis for cluster detection – one reported additional analysis (different number of strains included in the cgMLST) (Table 8). Over half (10/16; 63%), used an assembly-based allele calling method, five laboratories used both mapping- and assembly-based allele calling, and one used only mapping-based allele calling (Table 8).

Table 8. Results of allele-based cluster analysis

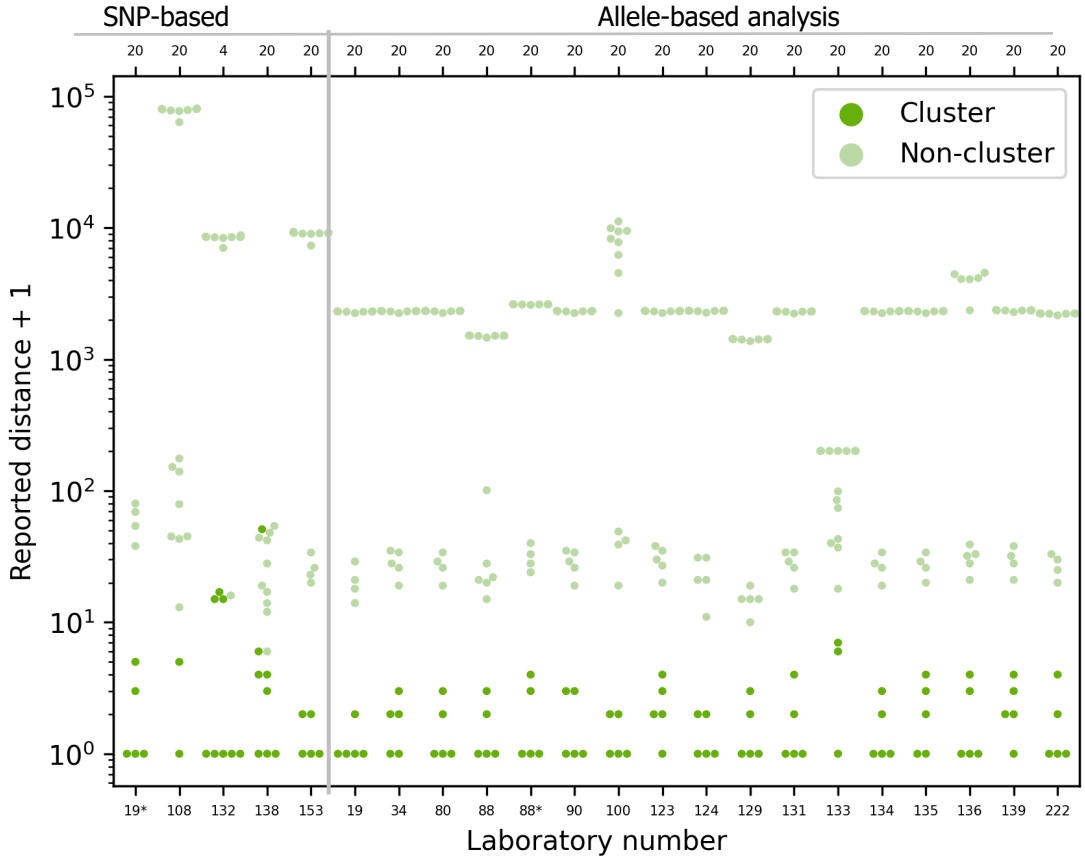
Lab No.	Allele-based analysis							
	Approach	Allelic calling method	Assembler	Scheme	No. of loci	Identified Pre-defined Cluster	Difference within cluster	Difference outside cluster
Provider	BioNumerics	A&M	SPAdes	Applied Maths (cgMLST/Enterobase)	2513	Yes	0–2	17-2356
19	BioNumerics	A&M	SPAdes	Applied Math (cgMLST/Enterobase)	2513	Yes	0-1	13-2315
34	SeqPhere	OAB	SKESA	Enterobase (cgMLST)	2513	Yes	0-2	18-2333
80	SeqPhere	OAB	Skesa	Enterobase (cgMLST)	2513	Yes	0-2	18-2337
88	INNUca/chewBBACA/ReporTree	OAB	SPAdes (version 3.14.0.)	INNUENDO (wgMLST)	7601/ 1592 shared	Yes	0-2	14-1517
88*	INNUca/ chewBBACA /ReporTree	OAB	SPAdes v3.14.0	INNUENDO (wgMLST)	7601	Yes	0-3	23-2622
90	SeqPhere	A&M	EnteroBase QAssembly	Enterobase (cgMLST)	2513	Yes	0-2	18-2332
100	SeqPhere	A&M	SKESA	Enterobase (cgMLST)	2513	Yes	0-1	18-11184
123	SeqPhere	OAB	SPAdes	Enterobase (cgMLST)	2513	Yes	0-3	19-2334
124	BioNumerics	OAB	SPAdes	Applied Math (cgMLST/Enterobase)	2506	Yes	0-1	10-2340
129	SeqPhere	OAB	Velvet	The Ridom§	1516	Yes	0-2	9-1428
131	SeqPhere	OAB	Velvet	Enterobase (cgMLST)	2513	Yes	0-3	17-2317
133	BioNumerics	OAB	Spades	Applied Math (cgMLST/Enterobase)	2513	No	0-6	17-200
134	SeqPhere	A&M	SKESA	Enterobase (cgMLST)	2513	Yes	0-2	18-2333
135	SeqPhere	OAB	SPAdes	Enterobase (cgMLST)	2513	Yes	0-3	19-2333
136	Enterobase	A&M	SPAdes	Enterobase (cgMLST)	2513	Yes	0-3	20-4546
139	Enterobase	OMB	-	Enterobase (cgMLST)	2513	Yes	0-3	20-2369
222	PHANtAsTiC[a]	OAB	SPADES[b]	Innuendo-curated Enterobase scheme	2360	Yes	0-3	19-2232

**: additional analysis. A&M: assembly- and mapping-based. OAB: Only assembly-based. OMB: Only mapping-based.*
[a]: pipeline run through IRIDA-ARIES webserver, using ARIES server as calculation engine.
[b]: SPADES 3.14.1, by using a default Filter SPAdes repeat.
§: SeqSphere in combination with software Target Definer (Annex 8).

Of the 16 laboratories using allele-based methods (main analysis), almost all (94%) identified the correct cluster of four closely related strains (Table 8). Twelve laboratories performed cgMLST using the same scheme as the EQA provider (cgMLST/Enterobase [10]) with 2513 loci. Three laboratories (124, 129, and 222) used a scheme with a slightly lower number of loci (2506, 1516, and 2360). Additional, one laboratory (88) used wgMLST in both the main and the additional analysis and obtained allelic differences within the cluster 0–2 and 0–3 (7601 loci). Fifteen of the 16 laboratories that identified the correct cluster reported allele differences of 0–3 within the cluster of closely related strains in their main analysis (Figure 8, Table 8). Laboratory 133 reported a slightly higher number of allelic differences 0–6 within the triplicated cluster strains (strain3, strain7 and strain20). Strain15 and strain17 were excluded from the cluster by the participant as they found 17 and 98 ADs, much higher than the EQA provider 1–2 ADs. However, the scheme was the same, but different approach for analysing (Only assembly-based (laboratory 133) and assembly- and mapping-based (EQA provider).

Four other test strains (strain8, strain14, strain16, and strain18) were also ST11, but not pre-defined by the EQA provider as part of the cluster. Based on the main analysis of cgMLST, 16 laboratories reported allele differences to the selected cluster strain at 9–73 for this group of strains (Table 8, Annex 8).

Figure 8. Reported SNP distances or allelic differences for each test strain to selected cluster representative strain



SNP: single nucleotide polymorphism.
Participants were instructed to select strain20 as reference (listed as '20' on the top scale). Note Laboratory 132 used strain4 (as reference as shown).
Dark green: reported cluster of closely related strains, Light green: not reported as part of cluster.

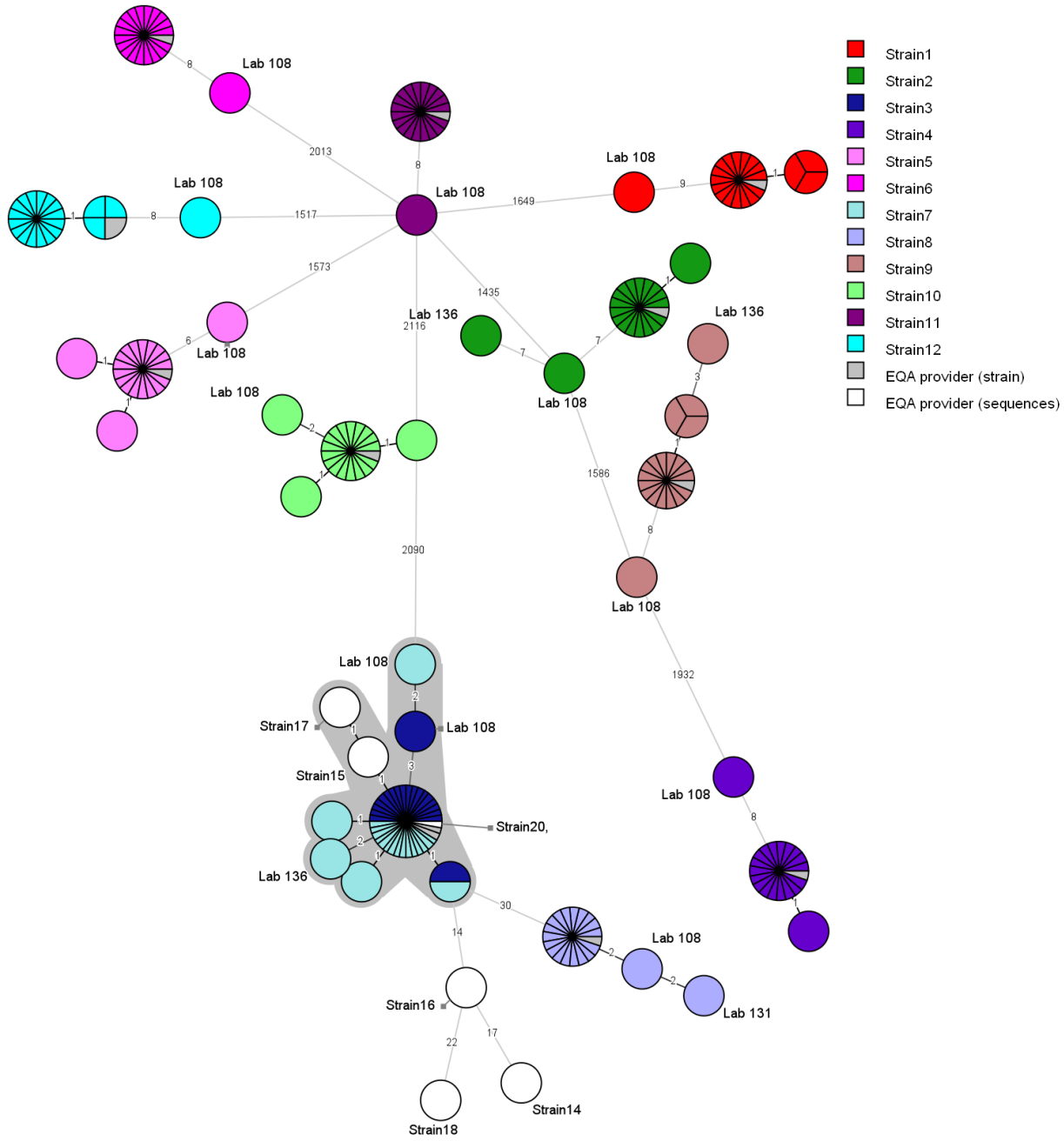
Only one of four laboratories identified the correct cluster when using SNP analyses, reported SNP with a maximum of 0–1 SNP distances (Table 7/Figure 8). Laboratories 108, 132, and 138 did not identify the correct cluster of closely related strains, as 108 only included two strains (15 and 20), laboratory 132 included all strains with ST11, and conducted the analysis using Strain4 as the reference instead of strain20 (sequence). Laboratory 138 included almost all ST11, but discarded one of the cluster strains (strain7) as the distance was 50 SNPs.

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 calculation engine for allele calling (Enterobase) [10] and evaluated by the EQA provider's in-house quality control (QC) pipeline [15].

The overall cgMLST analysis, shown in the minimum spanning tree (MST) based on submitted raw reads (FASTQ files) from the 20 laboratories revealed a clear clustering of the results for each test strain (Figure 9). Laboratory 108 did fall slightly outside of the clusters from each of the test strains (2-9 alleles) this is likely due to artefacts from comparing Ion Torrent generated data with illumina data. Also, the EQA provider strain (shown in grey) fall outside the main cluster of strain12 with one allele differences. The EQA provider strain fall inside the cluster of the other test strains (Figure 9).

Figure 9. Minimum spanning tree of core genome multilocus sequence typing participant FASTQ files

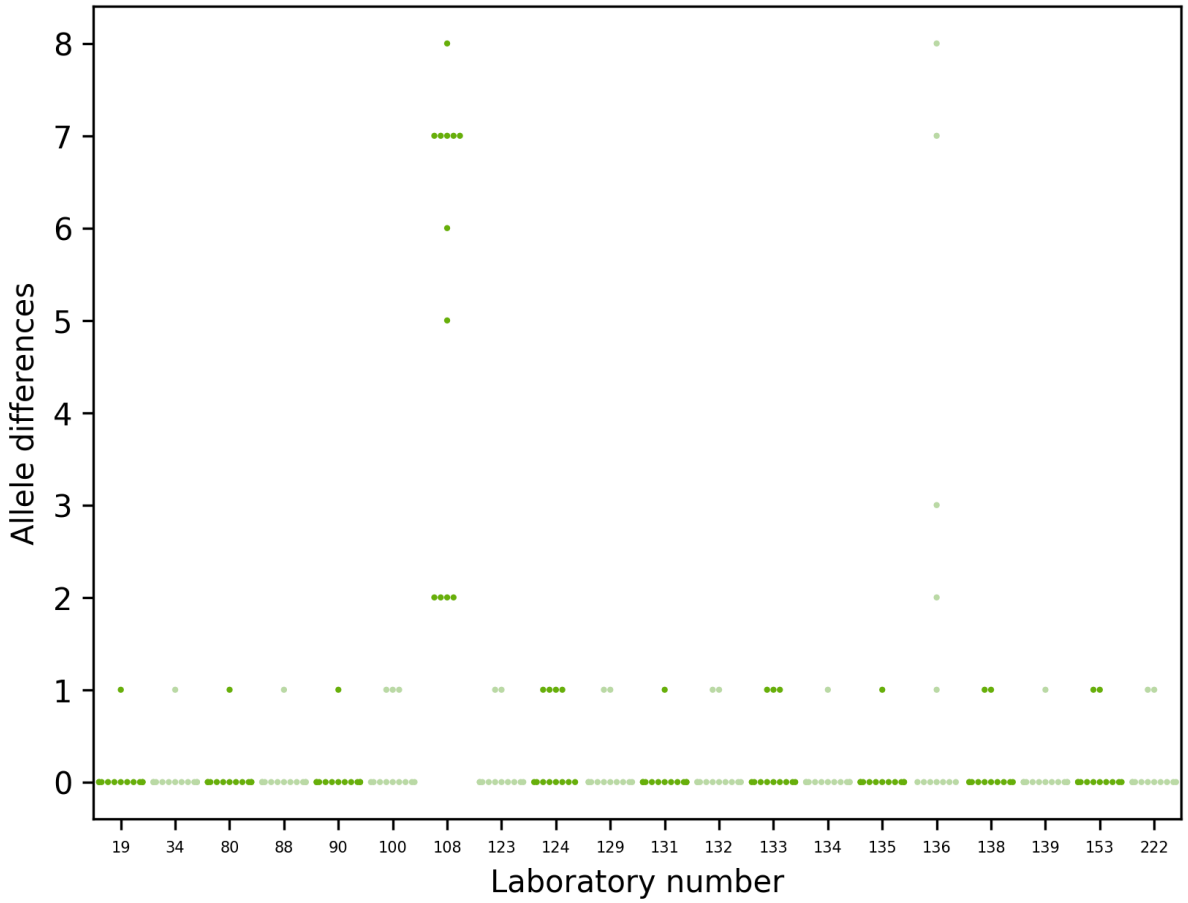


Minimum spanning tree (MST) in log scale of core genome multilocus sequence typing (cgMLST) [10] based on submitted raw reads (FASTQ files).
Each of the strain1–7 test strains have a different colour. EQA-provided sequences strains1–12 from the EQA provider are in grey, the provided sequences (strains14–20) in white. The provided modified sequences with poor quality (strain13 and strain19) were not included in the analysis. Strains3, 7, and 20 (sequence) were technical triplicates.
Results from laboratories 108 and 222 were run in CE (using Ion Torrent setup for allele calling).

The allele differences in Figure 9 do not exactly match those illustrated in the individual reports and consequently those in Figure 10, 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. Joint analysis accordingly contains fewer loci.

For each laboratory, cgMLST was performed on the submitted raw reads (FASTQ files), applying Applied Maths allele calling with the Enterobase scheme [10]. A hierarchical single linkage clustering was performed on the submitted data for each laboratory along with the EQA provider's reference strains. Figure 10 shows the allele differences between each submitted sequence and the corresponding reference.

Figure 10. 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.

For 196 of 240 results (81%), no allele difference was identified. For 30 results (13%), a difference of one allele from the reference strain was calculated, and for six results (3%) a difference of two-three alleles was observed. Ten reported results showed a difference between five and eight, mostly reported by laboratory 108.

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. Using different Q score parameters (Phred) was the most reported parameter, followed by confirmation of genus, and coverage with acceptance thresholds ranging from 20–60X were the most widely used QC parameters. 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 < 5% contamination with other genus	Min. x 50	No	4.640.000 bp - 5.560.000 bp	Min. 95% core percent and max.15 loci with multiple consensus
34	KRAKEN	70	>30	>4.8	>95%
80	KRAKEN	No	No	No	>= 90% good targets
88	Kraken (version 2.0.7, minikraken2_v2_8GB database) with reads and contigs. We also used confindr (version 0.0.4) with reads.	15x as a min. coverage. strain0013 we lowered the threshold to obtain an assembly.	Filtered lower quality sequences using a sliding window of 5bp with an average Q of at least 20. Reads less than 55bp after filtering were removed.	Assembly size lower than 80% or higher than 150% of the provided estimated genome size	§ Main analysis: cgMLST using only loci called (by chewBBACA) in all samples.
90	rMLST (pubmlst.org), Mash Screen (Ridom)	30x	No	4.9 - 5.9 Mb	>95%
100	KmerFinder 3.1 Center for Genomic Epidemiology	60x	FastQC, threshold to 30	SeqSphere assembler, genome size 5.0-5.5	SeqSphere cgMLST sheme, 95 % good targets threshold
108	Species Escherichia coli Genome size and coverage	20X	20X	4 000-6 000 Mbp	7
123	Contamination Check (Mash Screen) in SeqSphere	>50	No	5,0-5,5 Megabases	>98%
124	length, GC% and in silico PCR e coli det	>100 (acceptable >30 in BioNumerics)	Q30 >60 (threshold used in BioNumerics)	3.9 MB - 6.5 Mb	> 80% (alleles called available in BioNumerics)
129	No	30	No	No	90%
131	KmerFinder-3.2	>50x	>30	4,7-5,7 Mb	>95%
132	Bifrost pipeline (Based on Kraken) at SSI	No	No	5 200 000+/- 200 000	Match % was evaluated
133	PubMLST Species ID & Predicted pathotype in Bionumerics (+ Kraken)	>30	> 30	5Mb-5.8Mb	corePercent >= 96
134	Mash screen	coverage assembled > = 30	No	ref genome size +/- 10 %	No less than 10% of missing targets
135	Kraken2/Bracken built in our in-house assembly pipeline	>30	>30	between 4.6 and 5.8 Mb	>90% of alleles
136	KmerFinder on CGE website	Coverage threshold = 30X	No	5M pb and 5.7M pb	No
138	kraken2	40	30	4.909Mb -5.500	No
139	Kraken2 version 2.1.1	>45	Low-quality reads with phred scores below 15 and a length of less than 50 bp are discarded.	4.5 - 6.0 Mbp	No
153	KmerFinder, SpeciesFinder	>50x coverage. Or acceptable quality >30x coverage.	For Illumina Qscore >30	4-6MB.	No
222	No mismatches in the alignment with the 7 housekeeping genes of MLST panel (Warwick)	Min. 50x average depth of coverage across the genome	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	**95%**	**90%**	**55%**	**85%**	**80%**

See Annex 9 for additional information. § 88: We repeated the same procedure as before, using an extended set of loci present in 95% of the samples. Additionally, samples with less than 95% of these loci called were excluded (sample0013 and sample0019). By maximising the shared genome, this dynamic approach allowed increasing the resolution power and confidence on the clusters detected previously. In both cases the number of loci fall within the range of previous works (Llarena et al. 2018).

For each laboratory, the submitted raw reads (FASTQ files) were evaluated by the EQA provider's in-house quality control pipeline [15]. For the full QC evaluation of all strains, see Annex 10.

According to the QC parameters, sequencing quality was uniformly good. Three laboratories (34, 80 and 133) received warnings as *Pseudomonas tolaasii / Shigella flexneri was* identified as additional species, the contamination part of Bifrost is based on Kraken [15]. Four laboratories (88, 132, 134 and 153) have warnings as the average coverage is below the threshold of 50 but within the range of 26-44, the Bifrost pipeline is strict. Additional four laboratories (123, 133, 138 and 135) receive a waring of the '% unclassified' is above 20% threshold.

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. coverage (Mbp)	Length [1-25] x min. coverage (kbp)	No. of contigs at 25 x min. coverage	Contigs at [1,25] x coverage	Average coverage	No. of reads (x 1000)	Average read length	Average insert size	N50 (kbp)	QC status (Bifrost)
19	Ec	81.1-93.7	0.3-3.1	5.4-14.2	5.0-5.4	26.1-123.1	341.0-709.0	36.0-148.0	58.0-91.0	2308.0-3443.0	141.0-144.0	234.0-258.0	18.0-35.0	
34	Ec, Pt	72.7-88.7	5.0-11.1	5.7-13.9	5.0-5.5	0.0	173.0-451.0	0.0-0.0	143.0-319.0	6500.0-17763.0	151.0	112.0-193.0	73.0-147.0	Warning
80	Ec, Pt	72.5-92.6	1.9-7.7	4.6-17.8	5.1-5.6	0.0-127.1	88.0-304.0	0.0-8.0	67.0-105.0	2895.0-4402.0	151.0	273.0-336.0	69.0-160.0	Warning
88	Ec	77.8-95.9	0.4-2.4	2.8-17.9	5.0-5.6	0.0-58.3	183.0-396.0	0.0-56.0	43.0-119.0	928.0-2761.0	230.0-240.0	318.0-373.0	39.0-118.0	Warning
90	Ec	84.7-95.9	0.1-2.5	2.9-12.0	5.0-5.6	0.6-56.7	216.0-601.0	2.0-85.0	64.0-153.0	2475.0-6128.0	115.0-146.0	142.0-336.0	25.0-62.0	
100	Ec	85.7-98.1	0.0-2.4	1.7-10.9	5.0-5.6	0.0	92.0-312.0	0.0-0.0	375.0-718.0	13991.0-25157.0	145.0-148.0	275.0-316.0	89.0-181.0	
108#	Ec	89.6-97.9	0.4-1.6	1.4-7.1	4.9-5.3	0.0-5.1	811.0-3490.0	0.0-20.0	76.0-114.0	1471.0-2000.0	238.0-305.0	0.0	2.0-12.0	
123	Ec	62.7-97.2	0.2-2.0	2.2-34.3	5.1-5.6	0.0-12.2	90.0-351.0	0.0-14.0	78.0-105.0	1704.0-2455.0	239.0-260.0	275.0-327.0	53.0-140.0	Warning
124	Ec	90.1-98.1	0.3-2.5	1.0-5.1	5.1-5.6	0.0	71.0-246.0	0.0-0.0	258.0-288.0	6000.0-6000.0	251.0-251.0	400.0-416.0	91.0-199.0	
129	Ec	78.9-96.6	0.2-2.1	2.9-17.5	5.0-5.6	2.0-34.0	125.0-546.0	3.0-39.0	74.0-166.0	2690.0-6347.0	144.0-149.0	296.0-387.0	24.0-106.0	
131	Ec	84.2-96.8	0.1-2.6	2.7-11.5	5.1-5.6	0.0	89.0-321.0	0.0-0.0	132.0-160.0	4758.0-5982.0	148.0-149.0	263.0-342.0	72.0-168.0	
132	Ec	80.1-95.7	0.2-1.8	3.4-16.9	2.9-5.6	0.0-2257.1	89.0-352.0	0.0-66.0	26.0-201.0	927.0-7833.0	138.0-148.0	224.0-492.0	40.0-166.0	Warning
133	Ec, Sf	70.4-97.6	0.2-5.1	1.9-25.4	5.1-5.6	0.0	79.0-262.0	0.0-0.0	101.0-187.0	1965.0-3820.0	279.0-289.0	337.0-402.0	86.0-181.0	Warning

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. coverage (Mbp)	Length [1-25] x min. coverage (kbp)	No. of contigs at 25 x min. coverage	Contigs at [1,25] x coverage	Average coverage	No. of reads (x 1000)	Average read length	Average insert size	N50 (kbp)	QC status (Bifrost)
134	Ec	79.9-98.0	0.1-2.3	1.7-16.4	5.1-5.6	0.0-3.4	93.0-294.0	0.0-3.0	38.0-95.0	1362.0-3459.0	144.0-148.0	294.0-325.0	72.0-180.0	Warning
135	Ec	71.4-96.3	0.1-2.1	3.1-25.1	5.1-5.6	0.0	106.0-331.0	0.0-0.0	130.0-223.0	5092.0-7873.0	149.0	300.0-319.0	72.0-160.0	Warning
136	Ec	86.3-97.1	0.2-3.2	2.3-8.7	5.1-5.6	0.0	74.0-254.0	0.0-0.0	219.0-1421.0	8638.0-51514.0	146.0-149.0	326.0-492.0	90.0-169.0	
138	Ec	73.0-98.4	0.1-2.3	1.4-23.6	5.1-5.6	0.0	91.0-344.0	0.0-0.0	540.0-906.0	19386.0-31809.0	148.0-150.0	236.0-347.0	86.0-166.0	Warning
139	Ec	88.0-96.1	0.2-3.2	3.2-6.9	5.0-5.6	0.0-13.9	216.0-480.0	0.0-13.0	132.0-356.0	4853.0-13157.0	143.0-143.0	356.0-429.0	34.0-81.0	Warning
153	Ec	63.6-97.6	0.1-2.2	1.9-33.7	0.3-5.5	0.0-5124.7	99.0-280.0	0.0-229.0	32.0-50.0	1161.0-1742.0	148.0-149.0	310.0-356.0	59.0-157.0	Warning
222	Ec	90.4-97.5	0.2-2.7	2.0-5.8	5.1-5.6	0.0	81.0-281.0	0.0-0.0	219.0-586.0	7798.0-20849.0	151.0	337.0-375.0	84.0-166.0	

**: indicative QC range.*
Ec: E. coli, Sf: Shigella flexneri, Pt: Pseudomonas tolaasii (listed if >5%).
Warning: Some issues 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, the actions taken if errors were detected, the usefulness of the QC evaluation of the participant sequenced data, the usefulness of including low quality data, and suggestions for improvements (15/26 participants). 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 appreciated. However, one laboratory lacked a guideline with standardised QC criteria, and some commented on the format and details of the comprehensive STEC EQA. Two suggestions: to possibly include the detection of bacterial species from a clinical specimen in the coming EQAs, and the EQA schemes for different pathogens to be more evenly distributed around the year will be discussed with ECDC.

Table 11. Results of evaluation of the EQA scheme

Questions	Response (Yes)	Comments /actions
1) Used for accreditation/licensing purposes?	13/15 (87%)	One reported applying for accreditation last year.
2) Satisfied with the format/comments?	15/15 (100%)	One reported that the available PCR kit they applied did no obtain the expected results.
		One reported that for molecular analysis there were some confusements and remained unclear.
		One reported that it was clear and useful.
3) Differed any of your analytical test results?	7/15 (47%)	One reported that they will apply a new PCR kit for subtyping the STEC.
		One reported that they made an error in stx subtyping and use this result to fine tune our pipeline.
		One reported that one result was a mistake.
		One reported that they decided to use a new PCR-based test in order to improve the detection of the key virulence gene markers.
4) Usefulness of the manipulated sequences?	12/14 (86%)	One reported that it was useful with challenging strains.
5) Usefulness of the QC status of your submitted sequences?	13/14 (93%)	One reported that it is a useful comparison taking into consideration the lack of standardized QC criteria.
6) Improvements/remarks		Less labour-intensive, please. It is not the only EQA we have to do.
		We did like pool-format.
		The questions were really too many. Too many details on the analysis performed are requested, which would only be justified if an accurate discussion about this part would be made in the final report.
		The more details we receive in the evaluation report regarding what was expected from us (in fact, technical guidance), the better to figure out what to focus our attention on.
		Consider including direct detection from clinical specimens.

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 participants were satisfied with the format of the individual report and the additional feedback from the EQA provider. Only one participant reported that the molecular part was unclear. As the evaluation is based 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. In addition, the inclusion of the modified sequences in the cluster analysis and the QC feedback of the uploaded sequences was well received by most participants. The suggestions are listed in section 6.

4.1 Serotyping

Twenty-five (96%) laboratories participated in the serotyping part of the EQA-11, and nine of those participants (36%) provided phenotypic serotyping results and 16 (64%) provided molecular serotyping results (one by PCR and 15 by WGS). In this EQA, 19 participated in both O group and H typing, which is an increase from EQA-10 (16 participants), and 84% (16/19) reported the correct serotype for all 12 test strains also an improvement from EQA-10 were 12 (75%) correctly assigned all 10 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 EQA-11 we have seen an increase to 25/26 participants.

The performance of O grouping was similar to the performance in EQA-10 but differed from EQA-9. Seventeen (68%) of the 25 participants were able to correctly O-type all 12 test strains which is comparable to 71% in EQA-10 and only 50% in EQA-9. However, unlike EQA-10, not all the incorrect O group results were reported by laboratories using phenotypic methods. Laboratories 88 and 130 used WGS-based and PCR-based methods and did not determine several O groups (Annex 3). Twelve of the 46 incorrect results were reported as an incorrect type, while the rest were reported as non-typable/rough or not done. Note that the majority (83%, 38/46) of the incorrect results were reported by four laboratories.

The included O group O187, highlights the importance of correctly identify non-O157 and a new emerging strain causing human infections [17]. Seven participating laboratories reported it as either O74, O111 or non-typable. All seven laboratories except one (PCR-based) used phenotypic method for the O grouping. Note all laboratories using WGS-based methods correctly identified O187. The EQA provider has no knowledge of any known cross-reaction between O187 and O74 and O111 as well as the other mis-typed O-groups.

Some of the more common O groups, generated high performances except for O157 and O104. (O157: 88%, O145: 96%, O26:96%, O104:76% and O91: 81%). The average score was lower (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 to the average lower O-group performance in EQA-11 is due to four laboratories (one of the four have not participated in previous EQAs), using phenotypic serotyping, accounting for 83% of the incorrect results.

4.1.2 H type

Unlike the previous EQAs (EQA-10 94%, EQA-9 94%, and EQA-8 92%) the average performance for correctly H-typing the 12 tests strains was lower (84%). However, there was an increase in H-typing participation (19 laboratories) compared to EQA-10 (16 participants). The general performance for correctly reporting the H type, of all 12 test strains, was higher (84%) than the O grouping (68%). This might be explained by fewer participating laboratories and that the majority (16/19) used WGS-based methods. Only one participant using WGS-based method reported an incorrect H type (H- in strain5). By PCR-based method one laboratory incorrectly reported H42 in strain2 and one laboratory incorrectly reported two NT and H- in strain4 and strain5, respectively. The latter laboratory used phenotypic H-typing. The EQA provider has no knowledge of cross-reaction between H42 and H28.

4.1.2 OH serotyping

Complete O:H serotyping was performed by 19 (73%) participants with an average score of 95%, and for each strain the score ranged from 84% (17/19) for strain5 (O91:H14) to 100% (19/19) for strain3 (O157:H-/H7), strain7 (O157:H-/H7), and Strain10 (O145:H-/H28). 84% (16/19) of the participants who performed the O:H serotyping reporting the correct serotype on all 12 strains (Figure 2, Annex 3).

The average percentage O:H serotyping in this EQA was, as last year, higher (95%) compared to 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 O187:H28, O80:H2, and O91:H14 proved more difficult to identify particular if participants used phenotypic methods.

In addition to O grouping, H typing is crucial for outbreak detection, epidemiological surveillance, taxonomic differentiation of *E. coli* and detection of pathogenic serotypes. As such, it remains a main challenge to enable more NPHRLs to perform complete and reliable O:H serotyping, particular H typing. However, with the use of WGS, this might be more feasible for some countries in 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 the same as EQA-10, as such the highest rate was as follows; *stx* genes (96%), *eae* (92%), *aggR* (81%), subtyping of *stx* genes (85%), and the lowest rate was the detection of the *esta* (65%), which was not included in the previous EQAs

4.2.1 Detection of *aggR* and *esta*

The performance of detection the EAEC *aggR* genes was high with 95% of the participants correctly detecting *aggR* (21/22). This is comparable to EQA-10 where 94% correctly identified *aggR*. Likewise, the performance for *esta* was also high (89%) where only two laboratories reported a false negative result in. Both laboratories used other methods than WGS.

4.2.2 Detection of *eae*

Genotyping of *eae* had a high participation rate (92%) and performance; 21 (88%) laboratories obtained a 100% score, giving an average score of 97%. The average correct score has been fairly unchanged through the EQAs. (EQA-4 to EQA-10, 96%-99%).

4.2.3 Detection of *stx1* and *stx2*

Both the participation (96%) and performance rates were high for genotyping of *stx1* (99%) and *stx2* genes (96%), 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*

Strain4 harbours both *stx2a* and stx2c and it can be difficult to correctly identify both subtypes by using WGS-based methods. This implication, for the reliability of methods of *stx* subtype analysis, might result in laboratories need to utilise a PCR targeting both *stx2a* and *stx2c*. The difficulty of subtyping both *Stx2a* and *Stx2c*, was also reflected in the reported results, as only 32% (7/22) of participants correctly identified both subtypes. Therefore, the EQA provider has decided to disregard any subtyping results reported by the participants of Strain4.

Comparable to EQA-10, the average score of laboratories that correctly performed the *stx* subtypning were; 99% for *stx1*, 94% for *stx2*, and 93% combined *stx1* and *stx2*. Though not as high, as last year's EQA-10 (*stx1* 100% and *stx2* 97%), 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 three laboratories. 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, the true errors ('not done' results excluded) were nine incorrect *stx2* subtyping results. All errors were reported by four laboratories for three different strains. The average score for the *stx2* subtyping was 94% with the majority of the mis-subtyping *stx2* and reported results could be attributed to two laboratories. The EQA provider included the *stx2g* variant this year which was correctly identified by 91% (20/22) of the participants. The incorrect results were reported by two laboratories 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, and Stx2o, 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 of 26 laboratories (77%) performed cluster analysis all used WGS-derived data, no laboratories submitted PFGE-derived data.

4.3.1 WGS-derived data

Only one laboratory reported the use of external assistance for sequencing, and the majority (16/20) reported using an Illumina platform. All reported using commercial kits for preparing the library.

The EQA provider QC evaluation of the raw reads submitted by the participants showed good quality data, however 11/20 did received warnings by the Bifrost QC pipeline. The contamination assessment part of Bifrost is based on Kraken [11] which identified that three laboratories had contaminations by either *Pseudomonas tolaasii or Shigella flexneri.* Additional four laboratories had over 20% unclassified reads. The average coverage is an important QC parameter, but the threshold of 50 is a little strict; some argue that 30-40 would be enough, depending on the analysis [18].

As previous years the main QC parameters reported used by the participants in EQA-11 were a threshold of coverage and the checking of genus/species confirmation. The percentage of participant using assessment of the genome size have increased from 71% (EQA-9) to 85% and 95% use confirmation of genus as a QC parameter.

The performance was very high, with 16 (80%) laboratories correctly identifying the cluster of closely related strains, which is a little lower than last year (93%). Of the 20 laboratories, 16 (80%) reported using an allele-based method as the main analysis and four (20%) reported using SNP analysis. Three of the laboratories that used SNP-based analysis did not identify the pre-determined cluster. The distances reported using SNP-based analyses (and identifying the correct cluster) were 0–4 inside the cluster and the number of allele differences using cgMLST were 0–3 inside the cluster.

When assessing the reported allele difference or SNP distances, only the cgMLST approach showed comparable results, showing a clear separation of the cluster and non-cluster strains. One exception was the results from one laboratory using allelic-based analysis cgMLST, as strain15 and strain17 showed high number of ADs in the participants' analysis. However, the EQA provider could not repeat the large number of ADs when using only assembly-based calls in BioNumerics as reported by the participant. The result showed the cluster with very small differences. For the laboratories able to identify the correct cluster, a high level of similarity was seen for the reported cgMLST results based on Enterobase – most had three allele differences or below three within the cluster. Only the laboratory not identifying the correct cluster reported up to six allelic differences within the triplicated strains (sequences) an increase compared to other laboratories using the same scheme and approach.

SNP analyses can provide valid cluster detection at a national level and can be used for communication about cluster definitions, but the four laboratories using SNP as the main analysis reported very different results. One did not use strain20 but strain4 (a non-cluster strain) as the reference for the SNP analysis, and selected to include all ST11 in the reported cluster. Furthermore, neither SNP (NASP) nor BioNumerics analysis using wgMLST (of the submitted sequences of the cluster strain7, which laboratory 138 discarded as the distance was 50 SNPs) clarified the high number of SNPs observed by the participant. Both laboratories (132 and 138) have only recently started using WGS-derived data, and EQAs are a good way to test the status of this progress. Both laboratories provided good quality data, with only a few warnings, but 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 8, there is only a clear separation between the cluster strains and the remaining strains for laboratory 153, which did identify the cluster. The EQA provider highlights that a definitive cut-off in STEC WGS analysis has not been formally established, as this is of course difficult to do.

The EQA provider's analysis of the submitted raw data showed that when using a standardised cgMLST analysis, it is not unlikely to observe a random variation of one allele, even with high coverage (Figure 10). As has been the case in previous years, one participant (108) deviated consistently. This laboratory provided Ion Torrent data for which the EQA provider's analysis is not optimised, making correct assembly difficult. Therefore, the observed AD may be method artefacts. However, the use of Ion Torrent data can complicate the communication and investigation of multi-country outbreaks when only using the allelic method.

The majority of participants (80%, 16/20) correctly identified the poor quality of strain13, (a non-cluster sequence with reduced coverage and removal of genes) and thereafter excluded the sequence from the cluster analysis. An additional four participants accepted the quality for outbreak investigation. Comparable to the above, 85% of participants (17/20) correctly observed the contamination of the strain19 sequence (a non-cluster sequence contaminated with approx. 14% *E. albertii*). Three participants, however, reported the sequence to have acceptable quality (QC status A). One described the contamination. It is conceivable that it is more straightforward for laboratories to identify this very low-quality genome such as strain13 than to identify contamination with a different species. However, 14% contamination – in this case *E. albertii* – should be identified.

Comparable to EQA-10, most laboratories used more time to assess the modified genomes. In general, the participants described in detail what they observed and not just as previously suggested the re-run of the strain. It seems that the participants accepted the challenge, as advised by the contractor, and used the time to try to analyse the more questionable data and suggest if it was a cluster strain or not.

5 Conclusions

Twenty-six laboratories participated in the EQA-11 scheme; 25 (96%) performed the serotyping part, 25 (96%) the virulence profile determination part, and 20 (77%) cluster identification. Participation in serotyping and cluster analysis saw an increase from EQA-10 (21/26 81% and 16/26 62%, respectively). Unlike EQA-9 and EQA-10, this EQA included molecular typing-based cluster analysis using only WGS-derived data, as no participants reported PFGE data this year. This adjustment of STEC 'finger-printing' likely reflects the transitioning from PFGE to WGS among Member States.

The O:H serotyping was performed by 73% of the participants (19/26), with an average score of 95%. As in previous EQAs, participation in the O grouping was higher than in H typing. As in previous EQAs, not all laboratories demonstrated the capacity to determine all O groups and H types. In general, the more common European serotypes generated the highest scores compared to the less common serotypes, such as O187:H28, O80:H2, and O91:H14, which proved more difficult to identify, particularly if participants used phenotypic methods.

The EQA provider this year included two other DEC pathotypes – EAEC (*aggR* gene) and ETEC (*esta* gene) – testing the participating laboratories' abilities to detect STEC hybrid strains. The performance of the detection of the *aggR* genes was high (21/22, 95%), which is comparable to EQA-10 in which 94% correctly identified *aggR*. Likewise, the performance for *esta* was also high (89%), with only two laboratories reporting a false negative result. Both laboratories used other methods than WGS.

Detection of the *eae* gene had high participation rates, and average scores through the EQAs has always been above 96% (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%).

Similarly to previous EQAs, the participation and average scores for *stx1* and *stx2* gene detection were high, with an average score of 99% for *stx1* and 96% 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 [7]. The high participation rate of 85% (22/26) is unchanged from EQA-10 which is still encouraging. The average score of laboratories that correctly performed the *stx* subtypning were *stx1*, 99% for *stx1*, 94% for *stx2*, and 93% combined *stx1* and *stx2*.

Incorporating of the molecular typing-based cluster analysis in this EQA is up to date with the development of surveillance methods used by NPHRLs in Europe. Twenty laboratories performed the cluster analysis which is four more than EQA-10, and all 20 used WGS-derived data. Notably, no laboratory used PFGE for cluster analysis anymore.

Performance was very high, with 16 (80%) laboratories correctly identifying the cluster of closely related strains and two of four laboratories not identifying the correct cluster has just recently started using WGS data, the results are encouraging. Further, of the 20 laboratories, 16 (80%) reported using an allele-based method as the main analysis and four (20%) 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 and can be used for communication about cluster definitions.

As provided in EQA-10, the strain sequence data were made accessible by the EQA provider, and the participants were asked to include these in the cluster analysis and report characteristics and quality issues. Contamination with a different species can be more difficult to identify than low quality sequences. However, most of the participants did identity the contamination (quality issues) in strain19.

The current EQA scheme for typing STEC is the 11th 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. 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 employ each method as an individual test irrespective of results obtained in the screening and detection or any other test. Therefore, when a participant signs up for a test and subsequently participates, all strains must be tested using this test, e.g. 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 such as 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 evaluation of the provided genome sequences was a success. Almost all participants performed the analysis and identified the modifications introduced by the EQA provider. For the following EQA rounds, the EQA provider will continue and expand this part of the EQA in order to challenge the participants in their assessment of poor-quality genomes as well as contaminated ones. This will emphasise importance of assessing the genomes despite a low-level contamination or other quality issues – but of course concluded with the utmost caution.

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

The EQA provider emphasises the difficulty correctly identify both *stx2a* and *stx2c* in some STEC strains. Therefore, they recommend that laboratories might find it useful to utilise PCR as an additional method if WGS does not suffice.

Based on the feedback-survey, we conclude that the assessment of the QC of the participants submitted sequences is being appreciated, however one laboratory lacked a guideline with standardised QC criteria and some commented on the format and details of the comprehensive STEC EQA. Two suggestions; to possibly include detection of bacterial species from a clinical specimen in the coming EQAs and the EQA schemes for different pathogens to more evenly distributed around the year will be discussed with ECDC.

References

1. 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
2. European Parliament and European Council. Decision No 1082/2013/EU of the European Parliament and of the Council of 22 October 2013 on serious cross-border threats to health and repealing Decision No 2119/98/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
3. European Centre for Disease Prevention and Control (ECDC). Chlamydia infection. In: ECDC. Annual epidemiological report for 2021. Stockholm: ECDC; 2022. Available at: https://www.ecdc.europa.eu/sites/default/files/documents/AER%20STEC%20-%202021.pdf
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-Ord_o~nez A, Bover-Cid S, Chemaly M, Davies R, De Cesare A, Herman L, Hilbert F, Lindqvist R et al. Scientific Opinion on the 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. 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
10. Warwick Medical School. EnteroBase [Internet]. Coventry: University of Warwick; 2018 [cited 21 August 2018]. Available at: http://enterobase.warwick.ac.uk
11. 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.
12. 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
13. 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.
14. 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
15. Statens Serum Institut (SSI). Bifrost_QC [Internet; software package]. Copenhagen: Statens Serum Institut; 2019. Available at: https://github.com/ssi-dk/bifrost
16. 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.
17. 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	National Reference Center for Escherichia coli including VTEC	Institute for Medical Microbiology and Hygiene
Belgium	National Reference Laboratory STEC	Universitair Ziekenhuis Brussels
Bulgaria	NRL for Enteric Diseases	National Center of Infectious and Parasitic Diseases
Czechia	NRL for E. coli and Shigella	National Institute of Public Health
Denmark	International Escherichia and Klebsiella Centre	Statens Serum Institut
Estonia	Laboratory of Communicable Diseases	Health Board
Finland	Expert Microbiology Unit	Finnish Institute for Health and Welfare
France	Laboratoire de Microbiologie - Centre de Référence Escherichia coli	Centre Hospitalo-Universitaire Robert-Debré
Germany	NRC Salmonella	Robert Koch Institute
Greece	National Reference Centre for Salmonella, Shigella, VTEC	Department of Public Health Policy, School of Public Helath
Hungary	Reference Laboratories, Department of Bacteriology	National Public Health Center
Iceland	Dept. of Clinical Microbiology	Landspítali University Hospital
Ireland	NRL-VTEC	Public Health Laboratory
Italy	Microbiological Food Safety and Foodborne Disease Unit	Istituto Superiore di Sanità
Latvia	National Microbiology Reference laboratory	Infectology Centre of Latvia
Lithuania	National Public Health Surveillance laboratory	National public health survellance labortaory
Luxembourg	Epidemiology and Microbial Genomics (EPIGEM)	Laboratoire National de Sante
Montenegro	Laboratory for sanitary microbiology	Institute for Public Health of Montenegro
The Netherlands	Centre for Infectious Disease Research, Diagnostics and Laboratory Surveillance	RIVM
Norway	National Reference Laboratory for Enteropathogenic Bacteria	Norwegian Institute of Public Health
Poland	Department of Bacteriology and Biocontamination Control	National Institute of Public Health – National Institute of Hygiene
Portugal	LNR Infeções Gastrintestinais	Instituto Nacional de Saúde Dr. Ricardo Jorge
Romania	Molecular Epidemiology Laboratory	Cantacuzino National Medico-Military Institute of Research and Development
Slovenia	Department for Public Health Microbiology	National Laboratory of Health, Environment and Food
Spain	Unidad de Enterobacterias	Instituto de Salud Carlos III
Sweden	Mikrobiologi	Folkhäslomyndigheten

Annex 2. Participation overview EQA-10/-11

	2019-2020 (EQA-10)					2021-2022 (EQA-11)			
				Cluster					Cluster
Laboratory number	Participation (min. 1 part)	Serotyping	Virulence	PFGE	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
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	x	
128	x	x				x	x	x	
129	x	x	x			x	x	x	x
130	x		x			x	x	x	
131	x	x	x			x	x	x	x
132	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
137*	x	x	x		x				
138	x	x	x			x	x	x	x
139	x	x	x		x	x	x	x	x
145	x		x			x	x	x	
153	x	x	x			x	x	x	x
180*	x		x						
222	x	x	x		x	x	x	x	x
230						x	x	x	
240						x	x		
Number of participants	26	21	25	2	15	26	25	25	20

** = Laboratory did not participate in EQA-11*
#= Laboratory did not participate in EQA-10

Annex 3. Serotyping result scores

O group

	Strain number												
Laboratory number	1	2	3	4	5	6	7	8	9	10	11	12	
EQA	O26	O187	O157	O177	O91	O80	O157	O157	O128	O145	O146	O104	Method
19	26	187	157	177	91	80	157	157	128	145	146	104	A
34	26	187	157	177	91	80	157	157	128	145	146	104	C
80	26	187	157	177	91	80	157	157	128	145	146	104	C
88	NT	187	157	177	NT	80	157	NT	NT	145	NT	NT	C
100	26	187	157	177	91	80	157	157	128	145	146	104	A
108	26	187	157	177	91	80	157	157	128	145	146	104	C
123	26	187	157	177	91	80	157	157	128	145	146	104	C
124	26	187	157	177	91	80	157	157	128	145	146	104	C
127	26	74*	157	177	91	128*	157	157	128	145	146	104	A
128	26	187	157	177	91	80	157	157	128	145	146	104	A
129	26	187	157	177	91	80	157	157	128	145	146	104	C
130	26	ND	157	ND	ND	ND	157	157	ND	ND	ND	ND	B
131	26	NT	157	NT	91	80	157	157	128	145	146	104	A
132	26	187	157	177	91	80	157	157	128	145	146	104	C
133	26	187	157	177	91	80	157	157	128	145	146	104	C
134	26	187	157	177	91	80	157	157	128	145	146	104	C
135	26	187	157	177	91	80	157	157	128	145	146	104	C
136	26	187	157	177	91	80	157	157	128	145	146	104	C
138	26	NT	NT	145*	45*	NT	NT	NT	NT	145	NT	NT	A
139	26	187	157	177	91	80	157	157	128	145	146	104	C
145	26	ND	157	177	91	ND	157	157	ND	145	146	ND	A
153	26	187	157	177	91	80	157	157	128	145	146	104	C
222	26	187	157	177	91	80	157	157	128	145	146	104	C
230	26	NT	157	NT	158*	NT	157	26*	128	145	26*	NT	A
240	26	111*	157	145*	104*	ND	157	157	103*	145	ND	157*	A

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

H type

	Strain number												
Laboratory number	1	2	3	4	5	6	7	8	9	10	11	12	
EQA	H11	H28	H7/H-	H25/H-	H14	H2	H7/H-	H7/H-	H2/H-	H28/H-	H21	H4	Method
19	11	28	H-	H-	14	2	H-	H-	H-	H-	21	4	A
34	11	28	7	25	14	2	7	7	2	28	21	4	C
80	11	28	7	25	14	2	7	7	2	28	21	4	C
88	11	28	7	25	14	2	7	7	2	28	21	4	C
100	11	28	7	25	14	2	7	7	2	28	21	4	C
108	11	28	7	25	14	2	7	7	2	28	21	4	C
123	11	28	7	25	14	2	7	7	2	28	21	4	C
124	11	28	7	25	14	2	7	7	2	28	21	4	C
127	11	42*	7	25	14	2	7	7	2	28	21	4	B
129	11	28	7	25	14	2	7	7	2	28	21	4	C
131	11	28	H-	NT	H-*	2	H-	H-	H-	H-	21	4	A
132	11	28	7	25	14	2	7	7	2	H-	21	4	C
133	11	28	7	25	14	2	7	7	2	28	21	4	C
134	11	28	7	25	14	2	7	7	2	28	21	4	C
135	11	28	7	25	14	2	7	7	2	28	21	4	C
136	11	28	7	25	14	2	7	7	2	28	21	4	C
139	11	28	7	25	14	2	7	7	2	28	21	4	C
153	11	28	7	25	H-*	2	7	7	2	H-	21	4	C
222	11	28	7	25	14	2	7	7	2	28	21	4	C

n=19 participants
Purple shading: incorrect results
A: phenotypic serotyping, B: PCR-based serotyping, C: WGS-based serotyping
Some H- results was accepted as correct results (Strain3, Strain4, Strain7-10), 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												
Laboratory number	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	
EQA	**-**	**-**	**-**	**-**	**-**	**-**	**-**	**-**	**-**	**-**	**-**	**+**	**Method**
19	-	-	-	-	-	-	-	-	-	-	-	+	B
34	-	-	-	-	-	-	-	-	-	-	-	+	B
80	-	-	-	-	-	-	-	-	-	-	-	+	B
88	-	-	-	-	-	-	-	-	-	-	-	+	A
90	-	-	-	-	-	-	-	-	-	-	-	+	A
100	-	-	-	-	-	-	-	-	-	-	-	+	A
108	-	-	-	-	-	-	-	-	-	-	-	-*	B
123	-	-	-	-	-	-	-	-	-	-	-	+	B
124	-	-	-	-	-	-	-	-	-	-	-	+	B
127	-	-	-	-	-	-	-	-	-	-	-	+	A
128	-	-	-	-	-	-	-	-	-	-	-	+	A
129	-	-	-	-	-	-	-	-	-	-	-	+	B
130	-	-	-	-	-	-	-	-	-	-	-	+	A
131	-	-	-	-	-	-	-	-	-	-	-	+	A
132	-	-	-	-	-	-	-	-	-	-	-	+	B
133	-	-	-	-	-	-	-	-	-	-	-	+	B
134	-	-	-	-	-	-	-	-	-	-	-	+	B
135	-	-	-	-	-	-	-	-	-	-	-	+	B
136	-	-	-	-	-	-	-	-	-	-	-	+	A
139	-	-	-	-	-	-	-	-	-	-	-	+	B
153	-	-	-	-	-	-	-	-	-	-	-	+	B
222	-	-	-	-	-	-	-	-	-	-	-	+	B

n=22 participants
Purple shading: incorrect results
A: Other than WGS, B: WGS-based

Detection of *eae*

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

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

Detection of *esta*

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

n=19 participants
Purple shading: incorrect results.
A: Other than WGS, B: WGS-based

Detection of *stx1*

	Strain number												
Laboratory number	1	2	3	4	5	6	7	8	9	10	11	12	
EQA	+	-	+	-	+	-	+	+	-	-	+	-	Method
19	+	-	+	-	+	-	+	+	-	-	+	-	B
34	+	-	+	-	+	-	+	+	-	-	+	-	B
80	+	-	+	-	+	-	+	+	-	-	+	-	B
88	+	-	+	-	+	-	+	+	-	-	+	-	A
90	+	-	+	-	+	-	+	+	-	-	+	-	A
100	+	-	+	-	+	-	+	+	-	-	+	-	A
108	+	-	+	-	+	-	+	+	-	-	+	-	B
123	+	-	+	-	+	-	+	+	-	-	+	-	B
124	+	-	+	-	+	-	+	+	-	-	+	-	B
127	+	-	+	-	+	-	+	+	-	-	+	-	A
128	+	-	+	-	+	-	+	+	-	-	+	+*	A
129	+	-	+	-	+	-	+	+	-	-	+	-	B
130	+	-	+	-	+	-	+	+	-	-	+	-	A
131	+	-	+	-	+	-	+	+	-	-	+	-	A
132	+	-	+	-	+	-	+	+	-	-	+	-	B
133	+	-	+	-	+	-	+	+	-	-	+	-	B
134	+	-	+	-	+	-	+	+	-	-	+	-	B
135	+	-	+	-	+	-	+	+	-	-	+	-	B
136	+	-	+	-	+	-	+	+	-	-	+	-	A
138	+	-	+	-	+	-	+	+	-	-	+	-	A
139	+	-	+	-	+	-	+	+	-	-	+	-	B
145	+	-	+	-	+	-	+	+	-	-	+	-	A
153	+	-	+	-	+	-	+	+	-	-	+	-	B
222	+	-	+	-	+	-	+	+	-	-	+	-	B
230	+	-	+	-	+	-	+	+	+*	-	+	-	-

n=25 participants
Purple shading: incorrect results.
A: Other than WGS, B: WGS-based

Detection of *stx2*

	Strain number												
Laboratory number	1	2	3	4	5	6	7	8	9	10	11	12	
EQA	+	+	-	+	+	+	-	+	+	+	+	-	Method
19	+	+	-	+	+	+	-	+	+	+	+	-	B
34	+	+	-	+	+	+	-	+	+	+	+	-	B
80	+	+	-	+	+	+	-	+	+	+	+	-	B
88	+	+	-	+	+	+	-	+	+	+	+	-	A
90	+	+	-	+	+	+	-	+	+	+	+	-	A
100	+	+	-	+	+	+	-	+	+	+	+	-	A
108	+	+	-	+	+	+	-	+	+	+	+	-	B
123	+	+	-	+	+	+	-	+	+	+	+	-	B
124	+	+	-	+	+	+	-	+	+	+	+	-	B
127	+	+	-	+	+	+	-	+	+	+	+	-	A
128	+	+	-	+	+	+	-	+	-*	+	+	+*	A
129	+	+	-	+	+	+	-	+	+	+	+	-	B
130	+	+	-	+	+	+	-	+	-*	+	+	-	A
131	+	+	-	+	+	+	-	+	-*	+	+	-	A
132	+	+	-	+	+	+	-	+	+	+	+	-	B
133	+	+	-	+	+	+	-	+	+	+	+	-	B
134	+	+	-	-*	+	+	-	+	+	+	+	-	B
135	+	+	-	+	+	+	-	+	+	+	+	-	B
136	+	+	-	+	+	+	-	+	+	+	+	-	A
138	+	+	-	+	+	+	-	+	+	+	+	-	A
139	+	+	-	+	+	+	-	+	+	+	+	-	B
145	+	+	-	+	+	+	-	+	+	+	+	-	A
153	+	+	-	+	+	+	-	+	+	+	+	-	B
222	+	+	-	+	+	+	-	+	+	+	+	-	B
230	-*	-*	-	-*	-*	+	-	-*	-*	+	-*	-	-

n=25 participants
Purple shading: incorrect results.
A: Other than WGS, B: WGS-based

stx subtyping

stx1

	Strain number												
Laboratory number	1	2	3	4	5	6	7	8	9	10	11	12	
EQA	*stx1a*	-	*stx1a*	-	*stx1a*	-	*stx1a*	*stx1a*	-	-	*stx1c*	-	Method
19	*1a*	-	*1a*	-	*1a*	-	*1a*	*1a*	-	-	*1c*	-	B
34	*1a*	-	*1a*	-	*1a*	-	-*	*1a*	-	-	*1c*	-	B
80	*1a*	-	*1a*	-	*1a*	-	*1a*	*1a*	-	-	*1c*	-	B
88	*1a*	-	*1a*	-	*1a*	-	*1a*	*1a*	-	-	*1c*	-	A
90	*1a*	-	*1a*	-	*1a*	-	*1a*	*1a*	-	-	*1c*	-	A
100	*1a*	-	*1a*	-	*1a*	-	*1a*	*1a*	-	-	*1c*	-	A
108	*1a*	-	*1a*	-	*1a*	-	*1a*	*1a*	-	-	*1c*	-	B
123	*1a*	-	*1a*	-	*1a*	-	*1a*	*1a*	-	-	*1c*	-	B
124	*1a*	-	*1a*	-	*1a*	-	*1a*	*1a*	-	-	*1c*	-	B
127	*1a*	-	*1a*	-	*1a*	-	*1a*	*1a*	-	-	*1c*	-	A
129	*1a*	-	*1a*	-	*1a*	-	*1a*	*1a*	-	-	*1c*	-	B
131	*1a*	-	*1a*	-	*1a*	-	*1a*	*1a; 1c**	-	-	*1c*	-	A
132	*1a*	-	*1a*	-	*ND*	-	*1a*	*1a*	-	-	*1c*	-	B
133	*1a*	-	*1a*	-	*1a*	-	*1a*	*1a*	-	-	*1c*	-	B
134	*1a*	-	*1a*	-	*1a*	-	*1a*	*1a*	-	-	*1c*	-	B
135	*1a*	-	*1a*	-	*1a*	-	*1a*	*1a*	-	-	*1c*	-	B
136	*1a*	-	*1a*	-	*1a*	-	*1a*	*1a*	-	-	*1c*	-	A
138	*1a*	-	*1a*	-	*1a*	-	*1a*	*1a*	-	-	*1c*	-	A
139	*1a*	-	*1a*	-	*1a*	-	*1a*	*1a*	-	-	*1c*	-	B
145	*1a*	-	*1a*	-	*1a*	-	*1a*	*1a*	-	-	*1c*	-	A
153	*1a*	-	*1a*	-	*1a*	-	*1a*	*1a*	-	-	*1c*	-	B
222	*1a*	-	*1a*	-	*1a*	-	*1a*	*1a*	-	-	*1c*	-	B

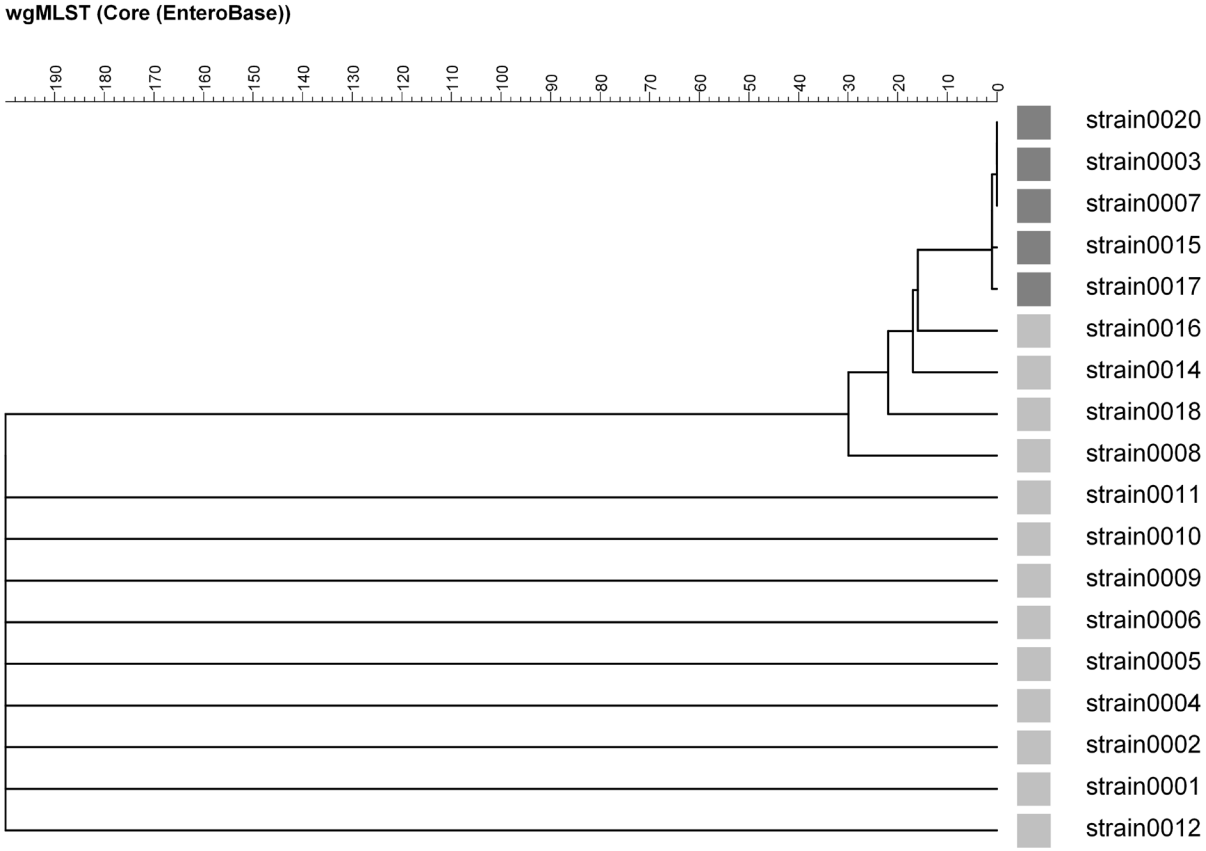
n=22 participants.
Purple shading: incorrect results.
A: Other than WGS, B: WGS-based.
ND = not done.

stx2

	Strain number												
Laboratory number	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	
EQA	*stx2a*	*stx2g*	-	*stx2a; stx2c*	*stx2b*	*stx2d*	-	*stx2c*	*stx2f*	*stx2a*	*stx2b*	-	**Method**
19	*2a*	*2g*	-	*2a; 2c*	*2b*	*2d*	-	*2c*	*2f*	*2a*	*2b*	-	B
34	*2a*	*2g*	-	*2a; 2c*	*2b*	*2d*	-	*2c*	*2f*	*2a*	*2b*	-	B
80	*2a*	*2g*	-	*2a; 2c*	*2b*	*2d*	-	*2c*	*2f*	*2a*	*2b*	-	B
88	*2a*	*2g*	-	*2a; 2c; 2d*	*2b*	*2d*	-	*2c*	*2f*	*2a*	*2b*	-	A
90	*2a*	*2g*	-	*2a; 2c*	*2b*	*2d*	-	*2c*	*2f*	*2a*	*2b*	-	A
100	*2a*	*2g*	-	*2c*	*2b*	*2d*	-	*2c*	*2f*	*2a*	*2b*	-	A
108	*2a*	*2g*	-	*2a*	*2b*	*2d*	-	*2c*	*2f*	*2a*	*2b*	-	B
123	*2a*	*2g*	-	*2c; 2d*	*2b*	*2d*	-	*2c*	*2f*	*2a*	*2b*	-	B
124	*2a*	*2g*	-	*2a; 2c*	*2b*	*2d*	-	*2c*	*2f*	*2a*	*2b*	-	B
127	*2a*	*2g*	-	*2c*	*2b*	*2d*	-	*2c*	*2f*	*2a*	*2b*	-	A
129	*2a*	*2g*	-	*2a; 2d*	*2b*	*2d*	-	*2c*	*2f*	*2a*	*2b*	-	B
131	*2a; 2c**	*2a; 2b**	-	*2a; 2c; 2d*	*2a**	*2a; 2c; 2d**	-	*2a; 2c; 2d**	*-**	*2a; 2c**	*2a**	-	A
132	*2a*	*2g*	-	*2a*	*2b*	*2d*	-	*2c*	*2f*	*2a*	*2b*	-	B
133	*2a*	*2g*	-	*2a; 2c*	*2b*	*2d*	-	*2c*	*2f*	*2a*	*2b*	-	B
134	*2a*	*2g*	-	*-*	*2b*	*2d*	-	*2c*	*2f*	*2a*	*2b*	-	B
135	*2a*	*2g*	-	*2d*	*2b*	*2d*	-	*2c*	*2f*	*2a*	*2b*	-	B
136	*2a*	*2g*	-	*2c*	*2b*	*2d*	-	*2c*	*2f*	*2a*	*2b*	-	A
138	*2a*	*2g*	-	*2a; 2c*	*2b*	*2d*	-	*2c*	*2f*	*2a*	*2b*	-	A
139	*2a*	*2g*	-	*ND*	*2b*	*2d*	-	*2c*	*2f*	*2a*	*2b*	-	B
145	*2a*	*ND**	-	*2a*	*2a**	*ND**	-	*ND**	*ND**	*ND**	*ND**	-	A
153	*2a*	*2g*	-	*2a; 2d*	*2b*	*2d*	-	*2c*	*2f*	*2a*	*2b*	-	B
222	*2a*	*2g*	-	*2c*	*2b*	*2d*	-	*2c*	*2f*	*2a*	*2b*	-	B

n=22 participants.
Purple shading: incorrect results.
Grey shading: disregarded strain4 see results section.
A: Other than WGS, B: WGS-based.
ND = not done.

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-11 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.
Strain3, Strain7 and Strain20 are technical triplicates.

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	Nextera	MiSeq
80	In own laboratory	Commercial kits	Kapa HyperPlus (Kapa Biosystems)	NextSeq
88	In own laboratory	Commercial kits	Illumina Nextera XT	MiSeq
90	In own laboratory	Commercial kits	Nextera XT DNA Library Prep Illumina	MiSeq
100	In own laboratory	Commercial kits	Illumina DNA Prep	NextSeq
108	In own laboratory	Commercial kits	IonXpress Plus fragment library kit	Ion GeneStudio S5 Prime system (Thermo Fisher Scientific, USA)
123	In own laboratory	Commercial kits	Nextera XT Library Prep Kit (Illumina)*	MiSeq
124	Externally	Commercial kits	KAPA HyperPlus Kit and Pippin prep size selection	NovaSeq 6000
129	In own laboratory	Commercial kits	Illumina Nextera XT*	MiSeq
131	In own laboratory	Commercial kits	Illumina® DNA Prep, (M) Tagmentation (96 Samples)	NextSeq
132	In own laboratory	Commercial kits	Illumina Nextera XT	MiSeq
133	In own laboratory	Commercial kits	Illumina DNA prep	MiSeq
134	In own laboratory	Commercial kits	DNA prep Illumina*	MiniSeq
135	In own laboratory	Commercial kits	Illumina DNA prep kit	NextSeq
136	In own laboratory	Commercial kits	Nextera DNA Flex Sample Preparation kit	NextSeq
138	In own laboratory	Commercial kits	Illumina DNAprep	Illumina NovaSeq 6000 with SP 300 reagents
139	In own laboratory	Commercial kits	Nextera XT kit (Illumina, San Diego, CA, USA)	NextSeq
153	In own laboratory	Commercial kits	Illumina DNA Prep	MiSeq
222	In own laboratory	Commercial kits	NEBNext® Fast DNA Fragmentation & Library Prep Set for Ion Torrent, New England Biolabs**	Ion GeneStudio S5 System

**: adjusted volume of reagents.*
***: decreased shearing time*

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

Laboratory	Reported cluster	Corresponding to EQA provider strains	Correct
Provider		Strain3, Strain7, Strain15, Strain17, Strain20	Yes
19	8151, 8831, 0015, 0017, 0020	Strain3, Strain7, Strain15, Strain17, Strain20	Yes
34	8252, 8260, 0015, 0017, 0020	Strain3, Strain7, Strain15, Strain17, Strain20	Yes
80	8246, 8767, 0015, 0017, 0020	Strain3, Strain7, Strain15, Strain17, Strain20	Yes
88	8293, 8161, 0015, 0017, 0020	Strain3, Strain7, Strain15, Strain17, Strain20	Yes
90	8762, 8146, 0015, 0017, 0020	Strain3, Strain7, Strain15, Strain17, Strain20	Yes
100	8097, 8842, 0015, 0017, 0020	Strain3, Strain7, Strain15, Strain17, Strain20	Yes
108	0015, 0020	Strain15, Strain20	No
123	8088, 8139, 0015, 0017, 0020	Strain3, Strain7, Strain15, Strain17, Strain20	Yes
124	8691, 8388, 0015, 0017, 0020	Strain3, Strain7, Strain15, Strain17, Strain20	Yes
129	8809, 8477, 0015, 0017, 0020	Strain3, Strain7, Strain15, Strain17, Strain20	Yes
131	8820, 8274, 0015, 0017, 0020	Strain3, Strain7, Strain15, Strain17, Strain20	Yes
132	8654, 8538, 8196, 0014, 0015, 0016, 0017, 0018, 0020	Strain3, Strain7, Strain8, Strain14, Strain15, Strain16, Strain17, Strain18, Strain20	No
133	8973, 8712, 0020	Strain3, Strain7, Strain20	No
134	8350, 8672, 0015, 0017, 0020	Strain3, Strain7, Strain15, Strain17, Strain20	Yes
135	8466, 8120, 0015, 0017, 0020	Strain3, Strain7, Strain15, Strain17, Strain20	Yes
136	8510, 8955, 0015, 0017, 0020	Strain3, Strain7, Strain15, Strain17, Strain20	Yes
138	8204, 8244, 0014, 0015, 0016, 0017, 0018, 0020	Strain3, Strain8, Strain14, Strain15, Strain16, Strain17, Strain18, Strain20	No
139	8447, 8756, 0015, 0017, 0020	Strain3, Strain7, Strain15, Strain17, Strain20	Yes
153	8715, 8921, 0015, 0017, 0020	Strain3, Strain7, Strain15, Strain17, Strain20	Yes
222	8358, 8526, 0015, 0017, 0020	Strain3, Strain7, Strain15, Strain17, Strain20	Yes

Strain3, Strain7 and Strain20 are technical triplicates.

Annex 8. Reported SNP distance and allelic differences

SNP distances

			Laboratory No.					
Strain ID	ST	Provider	19*	100*	108	132	138	153
Strain1	21	ND	ND	71546	80845	8463	41	9085
Strain2	200	ND	ND	82858	78538	8487	13	9102
Strain3#‡	11	0	0	0	44	0	0	0
Strain4¤	342	ND	ND	89844	79399	8767	47	9325
Strain5	33	ND	ND	58225	79963	8552	11	9164
Strain6	301	ND	ND	32462	77257	8571	43	9036
Strain7#‡	11	0	0	0	42	0	50	1
Strain8	11	90	79	33	175	15	5	33
Strain9	20	ND	ND	71159	80066	8500	18	9153
Strain10	32	ND	ND	24901	63579	7031	53	7323
Strain11	442	ND	ND	71076	80341	8364	16	8996
Strain12	678	ND	ND	83443	78096	8491	27	9164
Strain13	-	-	ND	ND	-	ND	-	-
Strain14	11	53	53	39	78	16	3	22
Strain15‡	11	5	4	2	4	0	2	1
Strain16	11	41	37	18	44	14	3	19
Strain17‡	11	2	2	1	12	0	0	0
Strain18	11	127	68	40	139	14	5	25
Strain19	-	-	ND	ND	151	-	-	-
Strain20#¤‡	11	0	0	0	0	0	0	0

Allelic differences

			Laboratory No.																
Strain ID	ST	EQA	19	34	80	88	88*	90	100	123	124	129	131	133	134	135	136	139	222
Strain1	21	2354	2305	2333	2337	1512	2613	2332	11184	2334	2340	1426	2316	200	2333	2333	4148	2368	2229
Strain2	200	2345	2311	2328	2333	1509	2617	2329	6215	2330	2340	1422	2315	200	2328	2329	2356	2353	2229
Strain3#‡	11	0	0	1	0	0	0	0	0	1	0	0	0	5	0	0	0	1	0
Strain4	342	2356	2307	2318	2322	1517	2613	2318	8258	2320	2340	1427	2303	200	2318	2318	4437	2369	2232
Strain5	33	2339	2301	2320	2326	1509	2605	2320	9451	2321	2330	1412	2305	200	2320	2320	4069	2352	2226
Strain6	301	2334	2298	2313	2317	1504	2593	2314	9921	2316	2330	1415	2300	200	2313	2314	4546	2346	2220
Strain7#‡	11	0	0	0	0	0	0	0	0	1	0	0	0	6	0	2	0	1	0
Strain8	11	34	28	33	33	21	39	33	48	34	30	18	33	42	33	33	38	37	32
Strain9	20	2324	2295	2311	2316	1512	2611	2309	4536	2312	2320	1420	2296	200	2311	2311	4061	2348	2220
Strain10	32	2272	2244	2243	2247	1460	2520	2244	2247	2245	2260	1374	2228	200	2243	2244	2285	2285	2159
Strain11	442	2344	2313	2329	2336	1513	2622	2331	7761	2332	2340	1420	2317	200	2330	2330	3921	2365	2230
Strain12	678	2345	2315	2327	2331	1511	2614	2327	9382	2328	2340	1428	2312	200	2327	2327	4027	2357	2229
Strain13	-	-	-	-	-	100	ND	-	-	37	10	-	-	-	-	-	-	-	-
Strain14	11	23	17	25	25	19	27	25	38	26	20	14	25	73	25	25	27	27	24
Strain15‡	11	1	0	2	2	2	3	2	1	3	1	2	3	17	2	3	3	3	3
Strain16	11	17	13	18	18	14	23	18	18	19	20	9	17	39	18	19	20	20	19
Strain17‡	11	2	1	1	1	1	2	2	1	2	1	1	1	98	1	1	2	2	1
Strain18	11	27	20	27	28	20	32	28	41	29	30	14	28	36	27	28	31	31	29
Strain19	-	-	-	34	-	27	ND	34	-	-	-	14	33	84	-	-	-	-	-
Strain20#¤‡	11	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

ST: sequence type.
‡: closely related strains (in grey).
#: technical triplicates strains.
¤: strain used as cluster representative by participant.

ND: Not done.

Annex 9. Reported QC parameters

Lab no.	1		2		3		4		5	
	Parameter	Threshold	Parameter	Threshold	Parameter	Threshold	Parameter	Threshold	Parameter	Threshold
19	N50 value	Available from QC analysis but no threshold	Number of contigs	Available from QC analysis but no threshold	Number of unidentified bases (N) or ambiguous sites	Available from QC analysis but no threshold				
80	Contamination	KRAKEN								
88	True coverage (module part of the INNUca pipeline)	This module calculates an improved estimation of the true bacterial chromosome coverage via read mapping against reference gene sequences distributed throughout the genome (17 genes in the case of E. coli).	Maximum number of contigs	assembly max. 100 contigs per 1.5Mb of expected genome size	contamination	Kraken (version 2.0.7, minikraken2_v2_8GB database) with reads and contigs. confindr (version 0.0.4) with reads				
90	Number of reads	>1 000 000	Average read length	>140 or > 230, according to the chemistry used	Total no of contigs >=200	<500	N50	>30 000	Contamination check	<5% other species
100	N50	40.000	contig count	500	SAV-NextSeq run parameters	clusters passing filter, no. of generated feads and Q30 score were all to Illumina recommendations	contamination check with KmerFinder	most reads classified as E. coli	read length	corresponds to expected length of sequencing platform and kit
123	assembly length	>5 000	N50	>50 000						
124	GC%	E.coli GC% +- 51%	N50	Threshold set in the quality control window of BioNumerics > 52100	non-ACGT bases	Scatterplot (length vs non-ACGT)	Nr BAFPerfect	Scatterplot (length vs BAFPerfect)		
131	N50	>10 000 bp								
133	Average Quality	>= 30	N50:	>=70 000	Ncontigs:	<500	NonAGCT	<2 500	Contamination	species ID PubMLST
134	countig count	1 000								
135	number of contigs	<= 650	GC%	49.5 - 51.0%	N50	>=30000	contamination (CheckM)	<4%	completeness (CheckM)	>96%

Lab no.	1		2		3		4		5	
	Parameter	**Threshold**	**Parameter**	**Threshold**	**Parameter**	**Threshold**	**Parameter**	**Threshold**	**Parameter**	**Threshold**
136	N50	>10k pb	Read number	<500	Low quality bases	The proportion of bases with quality lower than 10 should not be greater than 3% of total bases				
138	N50	> 72 000	Number of contigs	< 600	Number of contigs with length >1 000bp	<200	Read fraction majority species	> 95%	GC content, %	50.3 < x 50.9
139	N50 value	EnteroBase QC procedure: N50 >20kb	Number of contigs	<500	Proportion of scaffolding placeholders	EnteroBase procedure: <3%				
153	Contamination analysed by KmerFinder	analysed individually if necessary	N50	>30 000 bp	Number of contigs	<500				
222	Assembly quality (N50)	N50 >30 000 was used as a threshold for acceptable quality of the assembled contigs.								

Annex 10. Calculated qualitative/quantitative parameters

Quality Assessment made by the SSI in-house quality control pipeline https://github.com/ssi-dk/bifrost [18]

		Laboratory 19											
Parameters	Ranges*	8125	8151	8237	8301	8321	8379	8517	8616	8734	8777	8831	8960
Detected species	{Ea}, {Ec}, {Pt} or {Sf}	Ec	Ec	Ec	Ec	Ec	Ec	Ec	Ec	Ec	Ec	Ec	Ec
% Species 1		92.0	93.7	91.9	91.2	88.0	81.1	91.3	90.2	93.5	90.0	92.8	93.7
% Species 2		0.9	0.3	0.9	1.2	0.7	2.8	0.7	0.7	0.4	3.1	0.3	0.5
Unclassified reads (%)	{<100}	6.2	5.7	6.7	6.7	10.4	14.2	6.6	7.4	5.6	5.9	6.5	5.4
Length at >25 x min. coverage (Mbp)	{>4.5 ^ <5.8}	5.4	5.1	5.3	5.0	5.4	5.3	5.0	5.0	5.1	5.2	5.1	5.1
Length [1-25] x min. coverage (kbp)	{<250}	69.0	67.7	123.1	26.1	40.6	59.0	68.5	56.4	51.4	38.1	70.8	49.5
No. of contigs at 25 x min. coverage	{>0}	614	450	709	366	475	527	605	405	484	341	439	394
No. of contigs [1-25] x min. coverage	{<1000}	89	82	148	36	59	67	92	73	63	47	99	60
Average coverage	{>50}	66	79	58	78	86	76	77	71	78	83	79	91
No. of reads (x 1 000)		2 619	2 964	2 308	2 783	3 443	2 974	2 895	2 556	2 894	3 089	3 009	3 416
Average read length		143	143	142	144	142	144	142	143	144	144	141	143
Average insert size		256	248	238	251	242	245	243	256	253	258	240	234
N50 (kbp)		23	26	18	27	28	27	21	25	27	35	28	33
QC status (Bifrost)		OK	OK	OK	OK	OK	OK	OK	OK	OK	OK	OK	OK

		Laboratory 34											
Parameter	Ranges*	8061	8213	8252	8260	8383	8441	8489	8562	8570	8599	8739	8886
Detected species	{Ea}, {Ec}, {Pt} or {Sf}	Ec, Pt	Ec, Pt	Ec, Pt	Ec, Pt	Ec, Pt	Ec, Pt	Ec, Pt	Ec, Pt	Ec, Pt	Ec, Pt	Ec	Ec, Pt
% Species 1		74.2	84.6	88.2	88.2	79.5	84.5	83.3	87.7	76.5	84.5	88.7	72.7
% Species 2		11.1	5.8	5.4	5.1	8.3	5.4	7.3	5.6	10.1	7.3	5.0	8.4
Unclassified reads (%)	{<100}	13.9	6.7	6.0	6.2	9.9	7.4	8.0	6.0	11.8	7.0	5.7	13.5
Length at >25 x min. coverage (Mbp)	{>4.5 ^ <5.8}	5.4	5.3	5.2	5.2	5.2	5.1	5.5	5.2	5.0	5.4	5.2	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}	277	173	217	220	387	191	449	234	177	451	326	347
No. of contigs [1-25] x min. coverage	{<1 000}	0	0	0	0	0	0	0	0	0	0	0	0
Average coverage	{>50}	291	143	255	261	251	152	276	250	319	236	314	253
No. of reads (x 1 000)		17 340	6 500	11 736	12 076	13 166	6 786	1 780	11 440	17 763	12 162	14 103	13 188

Average read length		151	151	151	151	151	151	151	151	151	151	151	151
Average insert size		112	193	164	161	128	159	137	163	112	155	167	143
N50 (kbp)		94	131	145	145	94	147	84	143	145	97	140	73
QC status (Bifrost)		warning	warning	warning	warning	warning	warning	warning	warning	warning	warning	warning	warning

All 12 strains display warnings since the '% Species 1' + '% unclassified' is below 95%, 11 strains have Pseudomonas tolaasii listed as 2. species.

		Laboratory 80											
Parameter	**Ranges***	**8093**	**8131**	**8180**	**8246**	**8366**	**8514**	**8585**	**8628**	**8767**	**8816**	**8858**	**8978**
Detected species	{Ea}, {Ec}, {Pt} or {Sf}	Ec, Pt	Ec	Ec, Pt	Ec	Ec	Ec	Ec, Pt	Ec, Pt	Ec, Pt	Ec, Pt	Ec, Pt	Ec, Pt
% Species 1		82.0	84.5	80.7	92.6	84.9	87.6	79.7	79.9	83.6	84.3	85.3	72.5
% Species 2		5.5	4.7	6.6	1.9	4.8	4.7	7.3	7.7	6.4	5.3	5.9	5.3
Unclassified reads (%)	{<100}	11.5	9.3	10.5	4.6	8.2	6.9	9.5	10.5	9.3	9.0	7.9	17.8
Length at >25 x min. coverage (Mbp)	{>4.5 ^ <5.8}	5.6	5.6	5.1	5.2	5.2	5.2	5.3	5.1	5.2	5.5	5.2	5.4
Length [1-25] x min. coverage (kbp)	{<250}	0.5	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	127.1
No. of contigs at 25 x min. coverage	{>0}	173	304	108	175	272	219	126	88	165	297	170	247
No. of contigs [1-25] x min. coverage	{<1000}	1	0	0	0	0	0	0	0	0	0	0	8
Average coverage	{>50}	81	95	87	105	92	87	67	88	104	82	82	77
No. of reads (x 1 000)		3 541	4 131	3 611	4 036	3 767	3 504	2 895	3 734	4 402	3 580	3 416	3 322
Average read length		151	151	151	151	151	151	151	151	151	151	151	151
Average insert size		324	309	329	273	326	323	336	326	295	330	328	323
N50 (kbp)		117	76	141	146	91	160	135	159	123	95	118	69
QC status (Bifrost)		warning	OK	warning	OK	OK	OK	warning	warning	warning	warning	warning	warning

Four strains passed the QC, eight strains have warnings since '% Species 1' + '% unclassified' is below 95%, contamination with Pseudomonas tolaasii.

		Laboratory 88											
Parameter	**Ranges***	**8161**	**8162**	**8198**	**8266**	**8293**	**8687**	**8754**	**8826**	**8873**	**8880**	**8899**	**8981**
Detected species	{Ea}, {Ec}, {Pt} or {Sf}	Ec	Ec	Ec	Ec	Ec	Ec	Ec	Ec	Ec	Ec	Ec	Ec
% Species 1		95.9	91.6	91.4	93.6	94.6	89.3	90.5	94.7	93.6	77.8	93.1	95.2
% Species 2		0.4	0.5	2.4	0.8	0.6	0.5	1.4	0.6	0.5	2.1	0.8	0.4
Unclassified reads (%)	{<100}	2.8	6.7	4.4	4.2	4.1	8.9	5.7	3.9	4.1	17.9	4.9	3.6
Length at >25 x min. coverage (Mbp)	{>4.5 ^ <5.8}	5.3	5.5	5.2	5.6	5.2	5.5	5.1	5.2	5.2	5.5	5.0	5.2
Length [1-25] x min. coverage (kbp)	{<250}	0.5	19.5	19.8	16.0	6.0	12.7	10.8	16.3	26.4	14.7	58.3	0.0
No. of contigs at 25 x min. coverage	{>0}	183	396	189	347	242	247	201	212	343	279	266	247
No. of contigs [1-25] x min. coverage	{<1000}	1	21	17	14	6	16	12	11	20	18	56	0

Average coverage	{>50}	103	93	80	81	107	74	88	84	73	86	43	119
No. of reads (x 1 000)		2 438	2 278	1 866	2 020	2 489	1 819	1 994	1 925	1 676	2 091	928	2 761
Average read length		230	235	237	234	235	236	236	238	239	237	240	234
Average insert size		318	356	364	353	356	357	354	373	373	370	362	348
N50 (kbp)		118	45	66	59	68	61	77	80	54	55	39	111
QC status (Bifrost)		OK	OK	OK	OK	OK	OK	OK	OK	OK	OK	warning	OK

Eleven strains passed the QC, 8899 have an 'Average coverage' below 50.

		Laboratory 90											
Parameter	**Ranges***	**8002**	**8134**	**8146**	**8190**	**8473**	**8483**	**8613**	**8730**	**8762**	**8799**	**8849**	**8891**
Detected species	{Ea}, {Ec}, {Pt} or {Sf}	Ec	Ec	Ec	Ec	Ec	Ec	Ec	Ec	Ec	Ec	Ec	Ec
% Species 1		86.8	89.6	94.1	93.8	93.3	84.7	93.1	87.0	93.9	94.7	94.0	95.9
% Species 2		0.5	1.5	0.1	1.0	0.3	2.5	0.2	1.7	0.1	0.4	0.4	0.3
Unclassified reads (%)	{<100}	12.0	8.3	5.7	4.6	6.2	11.2	6.4	10.6	5.9	3.7	4.7	2.9
Length at >25 x min. coverage (Mbp)	{>4.5 ^ <5.8}	5.5	5.0	5.2	5.4	5.2	5.4	5.2	5.2	5.2	5.2	5.1	5.6
Length [1-25] x min. coverage (kbp)	{<250}	0.9	2.1	18.6	56.7	1.1	12.6	2.4	2.0	0.6	4.2	2.4	14.2
No. of contigs at 25 x min. coverage	{>0}	387	374	427	601	384	497	468	463	427	348	216	374
No. of contigs [1-25] x min. coverage	{<1000}	3	6	22	85	2	21	6	5	2	7	4	13
Average coverage	{>50}	130	110	98	64	88	65	96	130	153	101	94	96
No. of reads (x 1 000)		5 790	4 263	3 804	2 475	3 456	2 567	3 785	6 003	6 128	3 738	3 372	3 726
Average read length		127	133	136	144	135	139	134	115	133	144	144	146
Average insert size		182	219	261	301	287	287	244	142	212	318	336	317
N50 (kbp)		58	36	37	25	42	32	38	36	39	59	62	56
QC status (Bifrost)		OK	OK	OK	OK	OK	OK	OK	OK	OK	OK	OK	OK

		Laboratory 100											
Parameter	**Ranges***	**8017**	**8066**	**8097**	**8179**	**8287**	**8430**	**8512**	**8532**	**8577**	**8698**	**8842**	**8941**
Detected species	{Ea}, {Ec}, {Pt} or {Sf}	Ec	Ec	Ec	Ec	Ec	Ec	Ec	Ec	Ec	Ec	Ec	Ec
% Species 1		96.4	85.7	98.1	95.9	95.2	93.5	96.5	97.6	96.1	98.0	97.9	96.9
% Species 2		0.3	2.4	0.1	0.8	1.3	0.3	0.4	0.1	0.4	0.2	0.0	0.6
Unclassified reads (%)	{<100}	2.4	10.9	1.8	3.1	2.9	5.8	2.2	2.1	2.8	1.7	2.0	2.1
Length at >25 x min. coverage (Mbp)	{>4.5 ^ <5.8}	5.4	5.5	5.2	5.5	5.0	5.6	5.2	5.2	5.1	5.2	5.2	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}	269	236	169	312	92	165	296	249	113	170	166	125
No. of contigs [1-25] x min. coverage	{<1000}	0	0	0	0	0	0	0	0	0	0	0	0

Average coverage	{>50}	375	627	605	571	688	619	615	503	718	591	597	594
No. of reads (x 1 000)		13 991	23 590	22 061	21 732	23 754	23 679	22 323	18 335	25 157	21 286	21 860	21 455
Average read length		147	148	146	148	148	148	147	146	147	147	145	146
Average insert size		316	295	275	292	296	304	279	275	296	295	276	293
N50 (kbp)		92	89	181	101	166	136	97	166	168	147	148	147
QC status (Bifrost)		OK	OK	OK	OK	OK	OK	OK	OK	OK	OK	OK	OK

		Laboratory 108											
Parameter	**Ranges***	**8054**	**8137**	**8267**	**8340**	**8390**	**8424**	**8426**	**8502**	**8648**	**8685**	**8741**	**8904**
Detected species	{Ea}, {Ec}, {Pt} or {Sf}	Ec	Ec	Ec	Ec	Ec	Ec	Ec	Ec	Ec	Ec	Ec	Ec
% Species 1		94.5	96.8	89.6	97.9	97.8	95.9	95.7	97.8	95.5	97.2	95.3	93.4
% Species 2		0.8	1.0	1.6	0.4	0.5	0.8	1.6	0.6	1.4	0.6	0.9	0.4
Unclassified reads (%)	{<100}	4.0	1.8	7.1	1.4	1.5	2.2	1.8	1.4	2.5	2.0	2.3	5.5
Length at >25 x min. coverage (Mbp)	{>4.5 ^ <5.8}	4.9	5.2	5.3	5.2	5.0	5.3	5.0	5.2	4.9	5.0	4.9	5.3
Length [1-25] x min. coverage (kbp)	{<250}	1.7	2.7	1.1	0.0	2.0	5.1	1.5	0.0	2.5	2.9	2.7	3.2
No. of contigs at 25 x min. coverage	{>0}	2513	3061	2398	1031	2838	2948	2865	811	2619	2687	3490	3278
No. of contigs [1-25] x min. coverage	{<1000}	8	13	5	0	9	20	7	0	11	13	12	14
Average coverage	{>50}	98	103	80	114	107	76	89	108	92	84	92	87
No. of reads (x 1 000)		1 778	1 960	1 895	2 000	2 000	1 471	1 628	2 000	1 586	1 514	1 654	1 716
Average read length		289	298	238	305	287	293	293	289	299	297	299	290
Average insert size		0	0	0	0	0	0	0	0	0	0	0	0
N50 (kbp)		3	3	4	9	3	3	3	12	3	3	2	3
QC status (Bifrost)		OK	OK	OK	OK	OK	OK	OK	OK	OK	OK	OK	OK

		Laboratory 123											
Parameter	**Ranges***	**8088**	**8109**	**8139**	**8153**	**8200**	**8339**	**8401**	**8569**	**8635**	**8647**	**8753**	**8957**
Detected species	{Ea}, {Ec}, {Pt} or {Sf}	Ec	Ec	Ec	Ec	Ec	Ec	Ec	Ec	Ec	Ec	Ec	Ec
% Species 1		96.8	97.2	97.0	93.4	95.0	93.8	93.0	89.6	96.2	62.7	94.0	93.6
% Species 2		0.3	0.3	0.2	0.4	0.5	0.7	1.1	0.5	0.5	1.5	2.0	0.7
Unclassified reads (%)	{<100}	2.6	2.2	2.4	5.5	3.3	4.1	4.3	9.1	2.6	34.3	2.7	4.7
Length at >25 x min. coverage (Mbp)	{>4.5 ^ <5.8}	5.2	5.2	5.3	5.5	5.6	5.2	5.1	5.6	5.2	5.5	5.3	5.1
Length [1-25] x min. coverage (kbp)	{<250}	0.0	0.7	0.4	1.7	0.5	6.7	0.7	1.1	1.7	12.2	0.8	0.0
No. of contigs at 25 x min. coverage	{>0}	228	222	197	351	334	314	129	171	251	288	140	90
No. of contigs [1-25] x min. coverage	{<1000}	0	2	1	2	1	5	1	1	3	14	2	0

Average coverage	{>50}	103	99	85	84	80	93	96	105	97	78	103	90
No. of reads (x 1 000)		2 285	2 216	1 830	1 879	1 875	2 003	2 013	2 455	2 005	1 704	2 169	1 815
Average read length		243	239	251	253	246	251	251	246	260	255	256	259
Average insert size		285	275	302	309	292	303	307	294	325	312	314	327
N50 (kbp)		106	91	96	60	68	62	124	111	98	53	116	140
QC status (Bifrost)		OK	OK	OK	OK	OK	OK	OK	OK	OK	Warning	OK	OK

Eleven strains passed the QC, 8647 have '% unclassified' value above threshold (20%).

		Laboratory 124											
Parameter	**Ranges***	**8018**	**8144**	**8147**	**8188**	**8388**	**8414**	**8453**	**8586**	**8627**	**8691**	**8786**	**8905**
Detected species	{Ea}, {Ec}, {Pt} or {Sf}	Ec	Ec	Ec	Ec	Ec	Ec	Ec	Ec	Ec	Ec	Ec	Ec
% Species 1		94.2	98.0	90.1	95.2	98.0	96.4	95.8	98.0	96.2	98.1	95.7	97.2
% Species 2		0.3	0.3	2.5	2.1	0.4	0.5	0.7	0.3	0.7	0.4	1.0	1.0
Unclassified reads (%)	{<100}	4.5	1.0	5.1	1.4	1.2	1.8	1.4	1.1	1.4	1.1	2.1	1.1
Length at >25 x min. coverage (Mbp)	{>4.5 ^ <5.8}	5.6	5.2	5.5	5.3	5.3	5.1	5.2	5.2	5.6	5.3	5.1	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}	123	189	209	100	147	80	234	141	237	149	71	246
No. of contigs [1-25] x min. coverage	{<1000}	0	0	0	0	0	0	0	0	0	0	0	0
Average coverage	{>50}	260	276	263	274	276	288	276	278	258	275	286	264
No. of reads (x 1 000)		6 000	6 000	6 000	6 000	6 000	6 000	6 000	6 000	6 000	6 000	6 000	6 000
Average read length		251	251	251	251	251	251	251	251	251	251	251	251
Average insert size		411	407	408	405	409	416	408	414	413	400	414	412
N50 (kbp)		166	199	91	147	172	175	97	148	107	172	192	104
QC status (Bifrost)		OK	OK	OK	OK	OK	OK	OK	OK	OK	OK	OK	OK

		Laboratory 129											
Parameter	**Ranges***	**8216**	**8218**	**8223**	**8380**	**8456**	**8477**	**8513**	**8546**	**8717**	**8791**	**8809**	**8979**
Detected species	{Ea}, {Ec}, {Pt} or {Sf}	Ec	Ec	Ec	Ec	Ec	Ec	Ec	Ec	Ec	Ec	Ec	Ec
% Species 1		93.6	93.0	94.0	95.8	93.9	96.6	89.5	93.8	78.9	96.4	95.7	94.3
% Species 2		0.5	1.4	0.7	0.4	0.4	0.2	0.4	0.8	2.1	0.2	0.2	0.3
Unclassified reads (%)	{<100}	4.9	4.5	4.7	2.9	5.0	2.9	9.5	4.1	17.5	3.0	3.8	5.0
Length at >25 x min. coverage (Mbp)	{>4.5 ^ <5.8}	5.0	5.2	5.1	5.6	5.5	5.2	5.5	5.1	5.4	5.2	5.2	5.2
Length [1-25] x min. coverage (kbp)	{<250}	34.0	7.4	3.1	2.0	15.3	14.6	4.0	5.7	5.1	13.7	9.3	10.6
No. of contigs at 25 x min. coverage	{>0}	546	185	125	368	385	237	213	217	426	265	182	204
No. of contigs [1-25] x min. coverage	{<1000}	39	7	3	3	13	14	4	6	6	10	7	6

Average coverage	{>50}	142	126	129	158	103	99	139	121	166	74	91	78
No. of reads (x 1 000)		5 121	4 615	4 546	6 127	3 924	3 554	5 443	4 248	6 347	2 690	3 371	2 897
Average read length		147	147	147	147	148	149	146	148	147	147	145	144
Average insert size		309	311	318	297	328	355	302	312	296	369	387	352
N50 (kbp)		24	85	101	72	56	63	83	66	38	94	106	105
QC status (Bifrost)		OK	OK	OK	OK	OK	OK	OK	OK	OK	OK	OK	OK

		Laboratory 131											
Parameter	Ranges*	8145	8274	8371	8415	8508	8549	8563	8731	8794	8820	8902	8910
Detected species	{Ea}, {Ec}, {Pt} or {Sf}	Ec	Ec	Ec	Ec	Ec	Ec	Ec	Ec	Ec	Ec	Ec	Ec
% Species 1		93.9	96.7	95.2	91.3	96.8	93.9	95.4	93.9	84.2	96.6	96.3	94.2
% Species 2		0.4	0.1	0.8	0.3	0.3	1.0	0.4	2.2	2.6	0.1	0.2	1.0
Unclassified reads (%)	{<100}	4.7	3.0	3.7	7.9	2.7	4.4	3.4	3.2	11.5	3.1	3.2	3.3
Length at >25 x min. coverage (Mbp)	{>4.5 ^ <5.8}	5.1	5.2	5.5	5.6	5.2	5.1	5.6	5.3	5.5	5.2	5.2	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}	133	162	312	189	182	89	321	150	244	162	231	297
No. of contigs [1-25] x min. coverage	{<1000}	0	0	0	0	0	0	0	0	0	0	0	0
Average coverage	{>50}	152	156	144	141	142	154	155	142	136	132	160	159
No. of reads (x 1 000)		5 311	5 625	5 470	5 391	5 125	5 348	5 982	5 128	5 127	4 758	5 731	5 732
Average read length		148	149	149	148	148	149	148	148	149	149	149	148
Average insert size		275	296	309	268	263	321	281	268	321	342	296	277
N50 (kbp)		168	148	101	136	146	159	84	131	72	146	166	97
QC status (Bifrost)		OK	OK	OK	OK	OK	OK	OK	OK	OK	OK	OK	OK

		Laboratory 132											
Parameter	Ranges*	8073	8126	8185	8196	8538	8601	8614	8654	8737	8789	8986	8999
Detected species	{Ea}, {Ec}, {Pt} or {Sf}	Ec	Ec	Ec	Ec	Ec	Ec	Ec	Ec	Ec	Ec	Ec	Ec
% Species 1		94.4	94.1	92.3	95.7	95.4	94.1	92.4	94.6	94.2	80.1	95.1	93.9
% Species 2		0.5	0.4	0.4	0.3	0.2	1.7	0.7	0.2	0.4	1.8	0.2	0.6
Unclassified reads (%)	{<100}	4.2	4.6	6.7	3.8	4.2	3.4	6.2	5.1	4.2	16.9	4.5	5.0
Length at >25 x min. coverage (Mbp)	{>4.5 ^ <5.8}	5.1	5.6	5.5	5.1	5.2	2.9	5.1	5.2	5.2	5.5	5.2	5.5
Length [1-25] x min. coverage (kbp)	{<250}	2.2	2.5	8.6	78.5	6.4	2257.1	0.0	0.0	0.0	7.8	0.0	3.9
No. of contigs at 25 x min. coverage	{>0}	119	324	197	271	196	103	89	157	266	300	279	352
No. of contigs [1-25] x min. coverage	{<1000}	1	3	7	43	4	66	0	0	0	9	0	3

Average coverage	{>50}	102	118	83	41	54	26	95	99	128	121	201	102
No. of reads (x 1 000)		3 592	4 826	3 210	1 477	2 038	927	3 554	3 869	4 786	4 666	7 833	3 929
Average read length		147	140	147	146	142	148	138	138	144	145	138	146
Average insert size		350	273	350	451	379	492	280	272	288	303	224	344
N50 (kbp)		127	73	90	40	106	89	158	160	88	54	166	70
QC status (Bifrost)		OK	OK	OK	warning	OK	warning	OK	OK	OK	OK	OK	OK

Ten strains passed the QC, 8196 and 8601 displays warning as the 'Average coverage' is below 50.

		Laboratory 133											
Parameter	**Ranges***	**8102**	**8117**	**8490**	**8695**	**8712**	**8769**	**8830**	**8834**	**8835**	**8966**	**8973**	**8992**
Detected species	{Ea}, {Ec}, {Pt} or {Sf}	Ec	Ec	Ec	Ec	Ec	Ec	Ec	Ec	Ec	Ec	Ec	Ec, Sf
% Species 1		90.9	94.1	93.1	70.4	96.9	94.8	95.1	95.6	96.2	94.8	97.6	91.5
% Species 2		0.4	0.4	0.6	2.8	0.2	0.8	0.9	0.6	0.4	0.5	0.2	5.1
Unclassified reads (%)	{<100}	8.0	4.9	4.4	25.4	2.5	3.3	2.2	2.4	3.0	4.1	1.9	2.2
Length at >25 x min. coverage (Mbp)	{>4.5 ^ <5.8}	5.6	5.2	5.1	5.5	5.3	5.1	5.2	5.6	5.2	5.5	5.3	5.3
Length [1-25] x min. coverage (kbp)	{<250}	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
No. of contigs at 25 x min. coverage	{>0}	140	181	90	215	146	79	244	241	161	262	153	98
No. of contigs [1-25] x min. coverage	{<1000}	0	0	0	0	0	0	0	0	0	0	0	0
Average coverage	{>50}	187	124	159	131	101	132	138	163	122	161	179	119
No. of reads (x 1 000)		3 820	2 360	2 988	2 632	1 965	2 498	2 655	3 322	2 347	3 282	3 472	2 285
Average read length		286	289	282	288	279	281	284	286	285	286	285	283
Average insert size		371	396	350	402	337	348	364	384	367	379	368	360
N50 (kbp)		146	180	175	86	181	168	97	107	147	104	148	146
QC status (Bifrost)		OK	OK	OK	warning	OK	OK	OK	OK	OK	OK	OK	warning

Ten strains passed the QC, 8695 have '% unclassified above 20%, 8992 have '% Species 1' + '% unclassified' below 95% (contamination with Shigella flexneri).

		Laboratory 134										
Parameter	**Ranges***	**8050**	**8072**	**8087**	**8201**	**8236**	**8350**	**8421**	**8667**	**8672**	**8763**	**8857**	**8953**
Detected species	{Ea}, {Ec}, {Pt} or {Sf}	Ec	Ec	Ec	Ec	Ec	Ec	Ec	Ec	Ec	Ec	Ec	Ec
% Species 1		92.7	97.9	95.1	95.2	95.0	96.8	96.6	79.9	98.0	97.8	95.9	94.9
% Species 2		0.2	0.3	0.4	0.9	0.7	0.1	0.3	2.3	0.1	0.1	0.4	2.2
Unclassified reads (%)	{<100}	6.7	1.7	3.6	3.3	4.0	3.0	2.3	16.4	1.8	1.8	2.5	2.3
Length at >25 x min. coverage (Mbp)	{>4.5 ^ <5.8}	5.6	5.2	5.1	5.1	5.5	5.2	5.6	5.5	5.2	5.2	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	3.4	0.0
No. of contigs at 25 x min. coverage	{>0}	162	160	104	93	294	162	294	223	158	215	275	130

No. of contigs [1-25] x min. coverage	{<1000}	0	0	0	0	0	0	0	0	0	0	3	0
Average coverage	{>50}	70	95	61	69	78	69	76	92	91	79	38	53
No. of reads (x 1 000)		2 654	3 379	2 124	2 396	2 975	2 556	2 931	3 459	3 269	2 822	1 362	1 894
Average read length		148	148	148	148	148	144	148	148	148	148	148	148
Average insert size		313	316	324	318	325	294	315	314	319	315	316	312
N50 (kbp)		143	148	175	158	96	146	88	72	148	180	88	131
QC status (Bifrost)		OK	OK	OK	OK	OK	OK	OK	OK	OK	OK	warning	OK

Eleven strains passed the QC, one sequence has an 'Average coverage' below 50.

		Laboratory 135											
Parameter	**Ranges***	**8032**	**8046**	**8059**	**8104**	**8119**	**8120**	**8136**	**8219**	**8241**	**8466**	**8696**	**8854**
Detected species	{Ea}, {Ec}, {Pt} or {Sf}	Ec	Ec	Ec	Ec	Ec	Ec	Ec	Ec	Ec	Ec	Ec	Ec
% Species 1		89.3	96.3	93.5	94.6	71.4	95.8	92.8	96.0	88.0	96.3	94.3	94.3
% Species 2		0.3	0.2	0.5	0.5	2.1	0.2	1.1	0.2	0.1	0.2	0.3	1.4
Unclassified reads (%)	{<100}	9.7	3.1	5.3	3.7	25.1	3.7	5.1	3.3	11.4	3.2	4.2	3.5
Length at >25 x min. coverage (Mbp)	{>4.5 ^ <5.8}	5.5	5.2	5.1	5.6	5.5	5.2	5.1	5.2	5.6	5.2	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}	314	177	106	331	259	165	115	219	182	172	283	138
No. of contigs [1-25] x min. coverage	{<1000}	0	0	0	0	0	0	0	0	0	0	0	0
Average coverage	{>50}	130	200	223	166	179	171	192	197	164	197	199	200
No. of reads (x 1 000)		5 092	7 290	7 873	6 494	6 901	6 254	6 841	7 182	6 612	7 221	7 263	7 315
Average read length		149	149	149	149	149	149	149	149	149	149	149	149
Average insert size		304	300	302	301	319	300	319	311	316	312	310	314
N50 (kbp)		92	131	140	78	72	145	148	160	117	123	88	131
QC status (Bifrost)		OK	OK	OK	OK	warning	OK	OK	OK	OK	OK	OK	OK

Eleven strains passed the QC, one sequence has a '% unclassified' above 20%

		Laboratory 136											
Parameter	**Ranges***	**8195**	**8288**	**8299**	**8400**	**8486**	**8510**	**8531**	**8584**	**8609**	**8783**	**8888**	**8955**
Detected species	{Ea}, {Ec}, {Pt} or {Sf}	Ec	Ec	Ec	Ec	Ec	Ec	Ec	Ec	Ec	Ec	Ec	Ec
% Species 1		92.8	86.3	93.7	95.6	92.5	95.6	95.8	93.7	94.6	97.1	94.5	96.4
% Species 2		1.0	3.2	1.1	0.4	0.3	0.5	0.3	2.7	0.8	0.2	0.6	0.3
Unclassified reads (%)	{<100}	5.0	8.7	4.3	3.3	6.5	3.4	2.6	2.8	2.9	2.3	3.9	2.4
Length at >25 x min. coverage (Mbp)	{>4.5 ^ <5.8}	5.5	5.5	5.1	5.2	5.6	5.2	5.6	5.3	5.2	5.2	5.1	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}	250	200	74	179	134	136	243	105	254	157	112	149
No. of contigs [1-25] x min. coverage	{<1000}	0	0	0	0	0	0	0	0	0	0	0	0
Average coverage	{>50}	219	330	1300	381	415	444	576	1281	1004	1421	768	316
No. of reads (x 1 000)		8 638	12 561	45 948	14 034	15 903	16 378	22 404	46 443	36 826	51 514	27 214	11 615
Average read length		146	148	147	146	148	146	148	148	147	148	149	147
Average insert size		492	460	428	462	391	489	393	386	326	404	429	396
N50 (kbp)		91	90	169	160	146	149	95	135	97	147	168	159
QC status (Bifrost)		OK	OK	OK	OK	OK	OK	OK	OK	OK	OK	OK	OK

		Laboratory 138											
Parameter	**Ranges***	**8204**	**8240**	**8244**	**8248**	**8367**	**8445**	**8446**	**8496**	**8553**	**8580**	**8845**	**8938**
Detected species	{Ea}, {Ec}, {Pt} or {Sf}	Ec	Ec	Ec	Ec	Ec	Ec	Ec	Ec	Ec	Ec	Ec	Ec
% Species 1		98.0	95.7	98.0	96.5	95.5	98.4	95.8	73.0	96.7	93.8	98.1	93.2
% Species 2		0.1	0.9	0.3	1.2	0.4	0.1	0.5	2.3	0.3	0.4	0.1	0.2
Unclassified reads (%)	{<100}	1.8	2.8	1.6	1.9	3.1	1.4	2.8	23.6	2.1	5.4	1.5	6.3
Length at >25 x min. coverage (Mbp)	{>4.5 ^ <5.8}	5.2	5.1	5.2	5.3	5.2	5.2	5.1	5.5	5.6	5.5	5.2	5.6
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}	159	91	170	126	272	171	151	229	344	293	225	158
No. of contigs [1-25] x min. coverage	{<1000}	0	0	0	0	0	0	0	0	0	0	0	0
Average coverage	{>50}	543	563	670	678	635	738	906	540	721	627	737	761
No. of reads (x 1 000)		19 424	19 386	23 846	24 256	22 775	26 475	31 809	22 220	28 052	24 945	26 368	30 438
Average read length		150	150	150	150	150	150	149	150	148	150	149	149
Average insert size		342	324	311	317	319	275	236	347	252	322	293	315
N50 (kbp)		148	160	145	135	97	159	148	86	92	101	166	139
QC status (Bifrost)		OK	OK	OK	OK	OK	OK	OK	warning	OK	OK	OK	OK

Eleven strains passed the QC, 8496 have a '% unclassified' above 20%

		Laboratory 139											
Parameter	**Ranges***	**8005**	**8015**	**8020**	**8040**	**8166**	**8318**	**8404**	**8410**	**8447**	**8576**	**8620**	**8756**
Detected species	{Ea}, {Ec}, {Pt} or {Sf}	Ec	Ec	Ec	Ec	Ec	Ec	Ec	Ec	Ec	Ec	Ec	Ec
% Species 1		91.8	93.6	96.1	94.3	94.2	94.7	95.9	93.2	96.1	94.2	88.0	96.1
% Species 2		0.7	1.3	0.4	0.8	0.5	1.5	0.3	2.2	0.3	0.7	3.2	0.2
Unclassified reads (%)	{<100}	6.6	4.0	3.2	3.8	4.3	3.4	3.4	3.7	3.3	3.5	6.9	3.4
Length at >25 x min. coverage (Mbp)	{>4.5 ^ <5.8}	5.5	5.0	5.2	5.6	5.1	5.4	5.1	5.2	5.2	5.2	5.4	5.2
Length [1-25] x min. coverage (kbp)	{<250}	0.6	3.3	2.0	9.1	0.6	11.0	13.9	3.5	0.0	0.0	8.8	3.6

No. of contigs at 25 x min. coverage	{>0}	294	223	345	446	216	480	357	243	227	379	365	360
No. of contigs [1-25] x min. coverage	{<1000}	1	4	2	10	1	13	13	2	0	0	10	4
Average coverage	{>50}	273	179	196	147	280	154	132	202	356	275	140	188
No. of reads (x 1 000)		10 686	6 347	7 174	5 834	10 043	5 968	4 853	7 485	13 157	10 170	5 416	6 900
Average read length		143	143	143	143	143	143	143	143	143	143	143	143
Average insert size		429	387	360	357	386	370	370	397	409	356	374	371
N50 (kbp)		47	52	40	38	72	34	43	49	81	53	41	38
QC status (Bifrost)		OK	OK	OK	OK	OK	OK	OK	OK	OK	OK	Warning	OK

Eleven strains passed the QC, one sequence displays a warning since '% species 1' + '% unclassified' is below 95%.

		Laboratory 153											
Parameter	**Ranges***	**8086**	**8211**	**8359**	**8360**	**8411**	**8419**	**8715**	**8718**	**8846**	**8921**	**8950**	**8995**
Detected species	{Ea}, {Ec}, {Pt} or {Sf}	Ec	Ec	Ec	Ec	Ec	Ec	Ec	Ec	Ec	Ec	Ec	Ec
% Species 1		95.8	88.4	94.0	95.0	95.1	90.6	97.0	63.6	97.4	97.6	93.9	95.8
% Species 2		0.3	0.3	2.2	0.5	0.3	0.1	0.3	1.7	0.2	0.2	0.5	0.3
Unclassified reads (%)	{<100}	3.1	10.8	2.8	3.4	3.7	8.9	2.5	33.7	2.0	1.9	5.0	3.6
Length at >25 x min. coverage (Mbp)	{>4.5 ^ <5.8}	5.2	0.3	5.3	5.1	5.0	5.5	5.2	0.4	5.2	5.2	5.1	5.2
Length [1-25] x min. coverage (kbp)	{<250}	373.4	5124.7	0.0	0.0	187.7	13.4	88.6	5057.9	0.8	0.0	2.2	24.6
No. of contigs at 25 x min. coverage	{>0}	280	160	136	109	255	169	167	111	229	169	99	168
No. of contigs [1-25] x min. coverage	{<1000}	62	229	0	0	32	8	8	169	1	0	1	2
Average coverage	{>50}	32	35	42	50	40	37	32	38	40	44	44	44
No. of reads (x 1 000)		1 231	1 318	1 497	1 742	1 430	1 421	1 161	1 396	1 450	1 573	1 527	1567
Average read length		149	149	148	149	148	149	149	149	148	149	149	148
Average insert size		322	356	310	322	311	333	322	332	317	323	322	312
N50 (kbp)		72	61	131	148	87	136	145	59	133	157	147	148
QC status (Bifrost)		warning	warning	warning	OK	warning	warning	warning	warning	warning	warning	warning	Warning

Eleven strains have warnings, since 'Average coverage' is below 50, one passed the QC.

		Laboratory 222											
Parameter	**Ranges***	**8076**	**8094**	**8358**	**8392**	**8420**	**8464**	**8526**	**8692**	**8813**	**8856**	**8889**	**8997**
Detected species	{Ea}, {Ec}, {Pt} or {Sf}	Ec	Ec	Ec	Ec	Ec	Ec	Ec	Ec	Ec	Ec	Ec	Ec
% Species 1		95.4	95.3	97.0	95.2	95.7	90.4	97.5	97.0	94.0	95.8	91.7	95.8
% Species 2		0.4	0.4	0.2	0.6	1.2	2.7	0.2	0.4	0.2	1.2	1.5	0.5
Unclassified reads (%)	{<100}	3.8	3.4	2.5	3.0	2.3	5.4	2.0	2.3	5.2	2.6	5.8	2.3
Length at >25 x min. coverage (Mbp)	{>4.5 ^ <5.8}	5.2	5.1	5.3	5.6	5.3	5.4	5.3	5.2	5.6	5.5	5.1	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}	213	97	155	268	111	214	163	160	148	281	81	262
No. of contigs [1-25] x min. coverage	{<1000}	0	0	0	0	0	0	0	0	0	0	0	0
Average coverage	{>50}	219	514	586	488	504	439	567	530	495	499	427	508
No. of reads (x 1000)		7798	17755	20849	18600	18006	16096	20191	18801	18698	18717	14693	18016
Average read length		151	151	151	151	151	151	151	151	151	151	151	151
Average insert size		349	352	375	364	354	346	345	359	349	348	362	337
N50 (kbp)		166	160	159	84	135	88	145	147	143	101	166	97
QC status (Bifrost)		OK	OK	OK	OK	OK	OK	OK	OK	OK	OK	OK	OK

*Quality assessment made by the EQA provider in-house quality control pipeline. *: indicative QC ranges; Ec: E. coli, Ss: Shigella sonnei, Pt: Pseudomonas tolaasii (listed if >5%).*

Annex 11. Accessing provided sequences

Lab ID	Sero/ Stx sub	ST	Cluster	QC Status	Description Strain13
EQA provider				**C**	**A non-cluster sequence with reduced coverage and removal of genes**
19				C	The quality of Strain 0013 is not accepted. The average read coverage is too low resulting in a poorly assembled genome (low N50, many contigs and many N's) and a low core % in the cgMLST analysis. The strain has to be resequenced.
34				C	Coverage only 9fold, only 37% of the targets were found.
80				C	36.8% good targets cgMLST, low coverage.
88	7777:H7	11	No	B	This sample had low estimated coverage (9.4x), below the threshold we use by default (15x). We could still assemble a genome covering ~90% of the expected size (5MB) by lowering the default cutoff.
90				C	QC parameters used: Number of reads (>1000000) – 332736 Average read length (>140 bases) – passed Average genome coverage (>30x) - 10x Size of assembled genome (4.9-5.9 Mb) – passed Total no of contigs >=200 (<500) – 2015N50 (>30000) - 4 500cgMLST good targets (>95%) - 80% Contamination check (<5% other species) - no evidence for contamination found/passed Note: QC failed due to low no. of reads, low genome coverage, high no. of contigs, low N50, low % of cgMLST good targets.
100				C	Percentage of good targets too low (29.7), avg. coverage to low (9x), contig count too high (4.797), N50 too low (1.186), genome size too low (4.1 Mb).
108				C	This strain did not pass our QC cut off, which is 15X coverage for in-house SNP pipeline which is reference-based.
123	O157:H7	11	No	B	Percentage of good targets E. coli cgMLST: only 84 % (our threshold: 98%) species match: E. coli (no evidence for contamination) GC-content: 50.8 Genome size: 5.0Av. Coverage: only 8 (our threshold: 50N50: only 6425 (our threshold: 50000).
124		11	No	B	Strain 13 has a genome size of 5.0 Mb, a high number of N bases, a low coverage (average read coverage 9) and a low N50, suggesting sequencing problems. The E.coli plugin tool in BioNumerics could not determine the O-type. No stx genes were detected. Normally, this would have been performed via PCR before performing NGS. The MLST is ST11. These data fail the QC metrics to be reliable but, given the PCR free sequencing alleles that are called, could be trusted (31% of the alleles were called). The strain has 10 AD with the reference strain 20. With all this information, we can presume that strain 13 is not part of the cluster.
129				C	Percentage of good targets very low (35.8%), contig count very high (5153), average coverage very low (9).
131				C	cgMLST Perc. Good Targets too low (according to the cgMLST Perc Warning Targets too high) Avg. Coverage (Assembled) too low Approximated Genome Size lower than the expected N50 too short Number of contigs is too high.
132		11	No	B	Number of contigs are very high and the genome size is on the lower side and we decided to exclude this strain from the SNP analysis.
133				C	Poor Average Quality: >= 30 =TRUE Average Coverage: >=30 =FALSE, 9N50: >=70000 =FALSE, 6915 Ncontigs:<500 =FALSE, 1538 Non AGCT:<2500 =FALSE, 4295 Length: 5Mb-5.8Mb =TRUE Core Percent: >96% =FALSE, 31% Therefore, this strain is not acceptable for analysis it has 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 The QC failed 5/7 criteria not acceptable for analysis, low core %.
134				C	QC result = very bad. Coverage = 9 good targets found = 36.8% only contigs count = 4645 approx genome size = 4.2 Mb (instead of 5.2 on average).
135				C	too many contigs, N50 too low, total length too low, GC% too high, coverage too low, contamination too high, completeness too low, % matching alleles too low. Sequence depth is insufficient and culture may have been contaminated
136				C	Coverage and N50 values were low (Coverage = 14 and N50=4528), Contig number and low quality bases were high.
138				C	Insufficient number of readsQC – fail.

139				C	The Assembly failed in the EnteroBase quality control: coverage: 14 Number of bases (length): 5030838, criteria OK Number of contigs: 1947, > threshold of 800N50 value: 4528, < threshold of 20kbProportion of N's: 486454, > threshold of 3% Correct Species Assignment in Kraken Conclusion: Depth of coverage too low.
153				C	Very low coverage of the sequence.
222				C	The coverage was about 9x which is not acceptable for the analysis.

Lab ID	Sero	ST	Cluster	QC Status	Description Strain19
EQA provider	**O157:H7**	**11**	**No**	**B/C**	**A non-cluster sequence contaminated with approx. 14 *E. albertii***
19				C	The quality of Strain 0019 is not accepted. The N50 value is low, the genome is too large, and the assembled genome has many contigs. Moreover, many multiple consensus calls in the cgMLST analysis are seen, indicating a contamination. Kraken analysis confirmed a contamination with E. Albertii. The strain has to be restreaked and resequenced.
34	O157:H7	9999	No	B	Probably mixed culture, intimin type gamma1 and omicron were found.
80				C	86.7% good targets. KRAKEN: mix of E. coli and Escherichia albertii. During an outbreak situation we would have spread the isolate once more and picked a NSF O157:H7 colony and sequenced this. The sequence showed serotype O157:H7, stx1a stx2c.
88	O157:H7	11	No	B	QC detected a possible contamination with ~15% of E. albertii. After filtering contigs we obtained an assembly with the expected size.
90	O157:H7	11	No	B	QC parameters used: Number of reads (>1000000) – passed Average read length (>140 bases) – 137 Average genome coverage (>30x) – passed Size of assembled genome (4.9-5.9 Mb) - 6.8 Total no of contigs >=200 (<500) - 684N50 (>30000) – passed cgMLST good targets (>95%) - 79.9% Contamination check (<5% other species) - 13% contamination with Escherichia albertii Note: A new assembly was generated after excluding E. albertii reads by BWA mapping the raw reads against strain 0020 assembly. After this, an improvement of the failed parameters was observed (i.e., average read length - 141 bases, size of assembled genome - 5.3 Mb, total no. of contigs - 365, N50 – 70 068, cgMLST good targets - 98.9%), which changed the QC status to 'passed'.
100				C	Perc. of good targets too low (88,3%), genome size too high (6,6 MB), contig count too high (2202), N50 is a bit too low (18 284).
108		11	No	A	Genome size too big=8.143 Mbp, none of the options below meet our QC parameter. I only choose option 1 because you have to choose something to get further in the form. We would try to isolate EHEC from E. albertini if we had the colony in the laboratory.
123				C	Percentage of good targets E. coli cgMLST: 39.0 (our threshold: 98%) species match: (E. coli) potential contamination by second species above 10% detected E. albertii GC-content: 50.0 Genome size: 9.3 (expected is ~ 5.0–5.5 Megabases for E. coli) Av. Coverage: 68N50: only 13500 (our threshold: 50000).
124				C	Strain 19 has genome size of 9.6 Mb, high number of N bases and a high number of multiple alleles, suggesting a mix of two species: E.coli (E.coli det+) + E.albertii (E.albertii det+) (which was confirmed after running Kmer Finder). The E.coli plugin tool in BioNumerics could determine the O-type (O157) but not the ST. Both germs are from the same genus which makes the analysis very complex. The sample cannot be used for cgMLST analysis as both germs are analysed with the same scheme.
129	-	11	No	A	Good quality (percentage of good targets 98.3%, contig count 989, average coverage 103).
131	-	11	No	A	All values are in the expected range.
132	-	-	-	C	Number of contigs are very high and the genome size is too large and we decided to exclude this strain from the SNP analysis.
133	O157:H7	11	No	B	Below average, Average Quality: >= 30 =TRUE, Average Coverage: >=30 =TRUE, N50: >=70,000 =FALSE, very low, 18,690 Ncontigs:<500 =FALSE very 1532, suggests culture may be mixed NonAGCT:<2500 =FALSE, high 4133, suggests culture may be mixedLength: 5Mb-5.8Mb =FALSE, too long suggest two different speciesCorePercent: >96% =TRUE, The Quality, the coverage &

					Core percent all TRUE, strain is acceptable for Outbreak, will be able to carry out analysis using cgMLST(enterobase) E. albertii is the other strain, and was found using species ID on PubMLST.
134	-	-	-	C	Coverage = 94 contig count = 2202 (above the cut-off) good targets found = 86.7% (below the cut-off of 90%), approx genome size = 6.8 Mb (above the cut-off, genome size ref +/- 10%).In the MLST scheme, two different alleles were identified for mdh locus (allele 68 and allele 15). Two populations are mixed.
135	-	-	-	C	Total length too high, number of contigs too high, contamination too high, % assigned alleles too low. This seems to be a contaminated culture. NGS data suggest a mixture of 60% Escherichia albertii and 40% Escherichia coli.
136	-	-	-	C	Genome length higher than threshold (6.8 pb) and number of contig higher than 500 (619).
138	-	-	-	C	Sufficient number of readsContamination with E. albertiTyping possible, but not recommendedQC – Fail.
139	-	-	-	C	The Assembly failed in the EnteroBase quality control:Number of bases (length) : 6777198, > threshold of 6,4 MbpNumber of contigs : 619, criteria OK but highN50 value : 45191, criteria OKProportion of N's : 148025, criteria OKCorrect Species Assignment in KrakenConclusion: probable contamination with another E. coli sequence.
153	-	-	-	C	Potential contamination with different E. coli strain.
222	-	-	-	C	1799/2360 loci of cgMLST scheme correctly mapped, not satisfying our quality threshold set at 80% for reliability of cluster analysis. Even if the coverage was good (136x) and the 7 genes of conventional MLST (Warwick scheme) were all found 100% in length, the sequence is not accepted for cluster investigation because of the low number of loci of cgMLST scheme. Moreover, N50 calculated on the assembled contigs was <30000, used as a threshold for acceptable quality of the assembly.

-: 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 'Go to').

1. STEC EQA-11 2022-2023

Dear Participant

Welcome to the eleventh External Quality Assessment (EQA-11) scheme for typing of STEC in 2022-2023.

NOTE: new virulence gene esta (STa)

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

Please note that most of the fields are required to be filled in before the submission can be completed.

Any comments can be written at the end of the form.

You are always welcome to contact us at ecoli.eqa@ssi.dk.

Please start by filling in your country, your Laboratory name and your LAB_ID.

Available options in this submission form include:

- Click "Options" and "Pause" to save your results and finish at a later time (using the same link)
- Click "Options" and "Print" to print your answers. This can be done at any time, but before pressing "Submit results"
- Click "Previous" to go back to the questions you have already answered

Note: After pressing "Submit results" you will not be able to review your results.

2. Country

(State one answer only)

- ❑ Austria
- ❑ Belgium
- ❑ Bulgaria
- ❑ Czechia
- ❑ Denmark
- ❑ Estonia
- ❑ Finland
- ❑ France
- ❑ Germany
- ❑ Greece
- ❑ Hungary
- ❑ Iceland
- ❑ Ireland
- ❑ Italy
- ❑ Latvia
- ❑ Lithuania
- ❑ Luxembourg
- ❑ Montenegro

- ❑ Norway
- ❑ Poland
- ❑ Portugal
- ❑ Romania
- ❑ Slovenia
- ❑ Spain
- ❑ Sweden
- ❑ The Netherlands
- ❑ Türkiye

3. Institute name

4. Laboratory name

5. Laboratory ID

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

6. E-mail

7. 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 ___
Strain 11 ___
Strain 12 ___

8. Serotyping and virulence gene determination of STEC

9. 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

10. Submitting results - Serotyping

(State one answer only)

- ❑ Both O group and H type – Go to 11
- ❑ Only O Group – Go to 11
- ❑ Only H type – Go to 13
- ❑ Did not participate in serotyping – Go to 15

11. Results for serotyping (O Group)

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

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 ___

12. Please specify the method used:

Phenotypic or molecular (PCR-based, WGS-based)

(State only one answer per question)

Method

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

13. Results for serotyping (H Type)

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

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 ___

14. Please specify the method used:

Phenotypic or molecular (PCR-based, WGS-based)

(State only one answer per question)

Method

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

15. 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

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

(State only one answer per question)

- ❑ WGS
- ❑ Other

17. Results for virulence gene determination

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

	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 strains 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 strains 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*	*stx2b; stx2c*	*stx2b stx2d*	*stx2c stx2d*	*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 0020 (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
- ❑ SeqPhere – 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 Math (wgMLST) – Go to 43
- ❑ Applied Math (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 20 (as 0020 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 20 (as 0020 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 20 (as 0020 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 20 (as 0020 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 20 (as 0020 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 20 (as 0020 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 20 (as 0020 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 20 (as 0020 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 20 (as 0020 downloaded sequence)

Please use 9999 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 20 (as 0020 downloaded sequence)

Please use 9999 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 20 (as 0020 downloaded sequence)

Please use 9999 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 20 (as 0020 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 20 (as 0020 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 20 (as 0020 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 20 (as 0020 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 20 (as 0020 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 20 (as 0020 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 20 (as 0020 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 20 (as 0020 downloaded sequence)

Please use 9999 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 20 (as 0020 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 an additional 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

– 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 0020, downloaded sequences as reference) Otherwise indicate Multi-locus Sequence Type (e.g. ST8) and strain 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
- ❑ SeqPhere – 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 202

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

203. Please select scheme used for the allele analysis

(State one answer only)

- ❑ Applied Math (wgMLST) – Go to 205
- ❑ Applied Math (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 0020 (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 0020, downloaded sequences as reference) Otherwise indicate Multi-locus Sequence Type (e.g. ST8) and strain 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
- ❑ SeqPhere – 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 Math (wgMLST) – Go to 224
- ❑ Applied Math (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 0020 (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 five selected criteria, which were the most frequently reported by in previous EQAs.

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

For each criteria please also report the threshold or procedure used to 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. Thank you for your participation

Thank you for filling out the Submission for the STEC EQA-11.

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

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

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

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

European Centre for Disease Prevention and Control (ECDC)

Gustav III:s Boulevard 40
16973 Solna, Sweden

Tel. +46 858601000
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