Thirteenth external quality assessment scheme for typing of Shiga toxin-producing *Escherichia coli* in EU/EEA and EU enlargement countries, 2024

ECDC ASSESSMENT

Thirteenth external quality assessment scheme for typing of Shiga toxin-producing *Escherichia coli* in EU/EEA and EU enlargement countries, 2024

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 Kasper Rømer Villumsen, Vera Irene Erickson, Susanne Schjørring, Anne Sophie Majgaard Uldall, Flemming Scheutz, and Eva Møller Nielsen of the Foodborne Infections Unit at Statens Serum Institut, Copenhagen, Denmark.

Suggested citation: European Centre for Disease Prevention and Control. Twelfth external quality assessment scheme for typing of Shiga toxin-producing *Escherichia coli*. Stockholm: ECDC; 2025.

Stockholm, July 2025

ISBN 978-92-9498-813-3
DOI 10.2900/6760020
Catalogue number TQ-01-25-043-EN-N

Contents

Abbreviations

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

Executive summary

This report presents the results of the 13th round of the external quality assessment (EQA-13) 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-13 contained serotyping, detection of virulence genes, and molecular typing-based cluster analysis.

Human STEC infection is a zoonotic disease. For 2023, 10 217 confirmed human cases of STEC illness in the EU were reported by 27 Member States. This placed STEC as the third most commonly reported gastrointestinal food-borne illness, responsible for a reported 66 foodborne outbreaks in the EU during 2023. Twenty-six countries reported at least one confirmed STEC case and only Cyprus reported zero cases. In 2023, the EU notification rate was 3.1 per 100 000 population. This marked an increase of 30.0% in the annual notification rate reported compared with the previous year (2022; 2.4 cases per 100 000 population). The rise can partly be attributed to new laboratory diagnostic testing methods used in several countries. In 2023, information on serogroup was available for just 31.9% of the total reported cases in the EU, which was a decrease compared to previous year. The six most frequently reported serogroups were O157, O26, O146, O103, O145 (4.4%), and O63. Together, these six made up over 60% of cases, for which serogroup data were available [1].

Since 2007, ECDC has been responsible for the EU-wide surveillance of STEC, including facilitating the detection and investigation of food-borne outbreaks. Surveillance data, including basic typing parameters and molecular typing data for the isolated pathogen, are reported by Member States to The European Surveillance System (TESSy, since 2025 called EpiPulse Cases). The surveillance system relies on the capacity of NPHRLs in FWD-Net providing data to produce comparable typing results. To ensure that the EQA is linked to the development of surveillance methods used by NPHRLs, a molecular typing-based cluster analysis using whole genome sequencing (WGS)-derived data has been included since EQA-8.

The objectives of the EQAs are to assess the quality and comparability of typing data reported by NPHRLs participating in FWD-Net. The EQA test strains 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. In addition, eight strains (sequences) were included for the molecular typing-based cluster analysis. Twenty-two laboratories registered and all of them completed the exercise.

In total, serotyping was performed by 21 laboratories (95%), 22 laboratories determining the virulence profile, and 19 (86%) engaging in cluster identification using WGS data analysed by different approaches. The full O:H serotyping was performed by 86% (18/21) of participating laboratories. In O:H serotyping, participants achieved a high average score of 98%. Despite the high overall score, however, not all laboratories demonstrated the capacity to determine all included O groups and H types, and the participation in H typing was lower (18/22) compared to the O grouping (21/22). In the reported O-grouping results 71% (15/21) used WGS-based methods, which is higher than EQA-12 (68%), EQA-11 (60%), EQA-10 (52%), EQA-9 (50%), and EQA-8 (26%), indicating a shift towards using WGS based methods for serotyping.

The performance in detecting the virulence genes was also high: 97% for *stx1* and 95% for s*tx2*, and 98% for the *eae* gene. The average score of laboratories that correctly performed the *stx* subtyping were 100% for *stx1*, 97% for *stx2*, and 97% for stx1 and stx2 combined. These results were similar to previous EQAs. In EQA-13, two other diarrhoeagenic *E. coli* (DEC) pathotypes were included, ETEC strain1 (*esta* gene) and EAEC Strain10 (*aggR* gene) testing the participating laboratories in their abilities to detect STEC hybrid strains. The detection performance of the *aggR* gene was lower (20/22, 91%) than in EQA-12 (23/24, 98%) and EQA-11 (95%). Similar to *aggR*, the performance for *esta* was also lower (17/20, 85%) than EQA-12 (98%) and EQA-11 (89%). Fifteen laboratories utilised a WGS-based method to identify the genes.

Of the 22 laboratories participating in EQA-13, 19 (86%) performed molecular typing-based cluster analysis using WGS data analysed by different approaches. Since EQA-11, all participating laboratories have chosen the WGS based method and again none chose PFGE, a decrease from EQA-10 (2 laboratories) and EQA-9 (8 laboratories). The purpose of the cluster analysis part of the EQA was to assess the NPHRL's ability to identify a cluster of genetically closely related strains, i.e. to correctly categorise the cluster test strains regardless of the method used. The focus is on the result, not a specific procedure.

All participants (100%) correctly identified the cluster of closely related ST335 strains defined by pre-categorisation from the EQA provider among the 12 test strains and eight test strains (genomic sequences).

In this EQA, participants were free to choose their preferred analytical method for the WGS-based cluster identification. An allele-based method was most frequently used; 89% (17/19) used core genome Multi Locus

Sequence Typing (cgMLST) compared to 11% (2/19) 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, which can make the comparison and communication of the results between laboratories difficult. However, in EQA-13, all laboratories that completed the cluster analysis correctly identified the pre-determined cluster, regardless of the method used.

As part of the clustering analysis, the participants assessed additional genomes, some of which were modified by the EQA provider to provide a realistic view of various quality issues, and to challenge quality control efforts. Notably, 63% (12/19) of the participants reported quality issues with the modified sequence containing 9.3% contamination with *E. albertii*. In contrast, 95% (18/19%) correctly identified the poor quality of strain20, a non-cluster sequence with reduced coverage and removal of genes. Assessing both contamination with a different species and poor quality is crucial before conducting WGS analysis.

A feedback survey was sent to assess the STEC EQA-13 scheme. The questionnaire contained both questions related to accreditation and information on the individual report; 59% (13/22) responded. Overall, the survey revealed an appreciation for QC assessment but highlighted the need for a simplified process for reporting results. Streamlining the reporting form, especially for virulence gene determination, was suggested. All the respondents appreciated the format, and some listed recommendation for improvements.

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 extends such cooperation and supports the implementation of quality assurance schemes [2].

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

ECDC has outsourced the organisation of EQA schemes for EU/EEA countries in the disease networks. EQAs aim to identify areas for improvement in laboratory diagnostic capacities relevant for epidemiological surveillance of communicable diseases as set forth in Decision No 1082/2013/EU [3] and ensure the reliability and comparability of results generated by laboratories across all EU/EEA countries. When operational, the network of EU reference laboratories coordinated by ECDC will be responsible for coordinating the network of national reference laboratories' activities, including EQAs, according to the Regulation 2022/2371/EU.

The main purposes of EQA schemes are to:

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

Since 2012, the unit of Foodborne Infections at SSI, Denmark, has been the EQA provider for the three EQA schemes covering typing of *Salmonella enterica* ssp. *enterica*, Shiga toxin/verocytotoxin-producing *Escherichia coli* (STEC/VTEC) and *L. monocytogenes*. In 2021, SSI was granted the new round of tenders (2022–2025) for *Listeria* and STEC. The STEC EQA covers serotyping, virulence profile determination, and molecular typing-based cluster analysis. This report presents the results of STEC EQA-13.

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 uremic syndrome (HUS), which is clinically defined as a combination of haemolytic anaemia, thrombocytopenia and acute renal failure.

In 2023, 10 217 confirmed cases of STEC infection were reported in the EU by 27 Member States. The overall EU notification rate was 3.1. cases per 100 000 population, which exceeded the pre-pandemic level and represented a 30% increase compared to the notification rate in 2022, continuing a trend of increasing case reporting over the past 5 years [1]. Information on serogroup was available for 3 259 cases (31.9%) In 2023, the six most frequently reported serogroups were O157, O26, O146, O103, O145, and O63. These serogroups together accounted for over 60% of the total number of confirmed STEC cases with known serogroups in 2023. A total of 505 HUS cases, as well as 15 casualties, were reported across 20 EU Member States. In cases for which serogroup information was available, the most common serogroups associated with HUS cases were O26 (37.8%), O157 (19.6%) O145 (7.6%) and O80 (5.3%) [1].

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 EpiPulse Cases. In addition to the basic characterisation of the pathogens isolated from human infections, there is public health value in using more discriminatory typing techniques for pathogen characterisation in the surveillance of food-borne infections. Since 2012, ECDC has enhanced EU surveillance by incorporating molecular typing data through isolate-based reporting. Three selected FWD pathogens were included: *Salmonella enterica* ssp. *enterica*, *L. monocytogenes*, and STEC. The overall aims of integrating molecular typing into EU level surveillance are to:

- foster the rapid detection of dispersed international clusters/outbreaks;
- facilitate the detection and investigation of transmission chains and relatedness of isolates across Member States and contribution to global investigations;
- detect the emergence of new evolving pathogenic isolates;
- support investigations to trace the source of an outbreak and identify new risk factors; and
- aid the study of a particular pathogen's characteristics and behaviour in a community of hosts.

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

1.3 STEC characterisation

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

The recommended method for *stx* subtyping is a specific polymerase chain reaction (PCR) [4]. STEC serotype O157:H7 may be divided into two groups: non-sorbitol fermenters (NSF) and a highly virulent sorbitol fermenting (SF) variant of O157. STEC EQA-13 included O:H serotyping, detection of virulence genes (*eae, stx1* and *stx2*, including subtyping of *stx* genes), the *aggR* gene specific for enteroaggregative *E. coli* (EAEC), the *esta* gene specific for enterotoxigenic *E. coli* (ETEC), and molecular typing-based cluster analysis.

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

1.4 Objectives of EQA-13 on STEC

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

As a result, and part of the recommendations in EQA-10, the EQA provider does not include *aaiC* gene in EQA-13. 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-13 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-13 were to assess the ability to assign the correct virulence profile; the presence/absence of *stx1*, *stx2*, *eae, esta,* and *aggR* genes and subtyping of *stx* genes (*stx1a, stx1c, stx1d, stx2a, stx2b, stx2c, stx2d, and stx2g*).

1.4.3 Molecular typing-based cluster analysis

The objective of the molecular typing-based cluster analysis of STEC EQA-13 was to assess the ability of the participants to correctly identify the cluster of closely related strains. Laboratories could perform analysis using PFGE and/or derived data from WGS. The cluster analysis should be conducted on the 12 test strains and eight additional test strains (provided genomic sequences). Some of the provided sequences were modified to have quality control (QC) issues.

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

2 Study design

2.1 Organisation

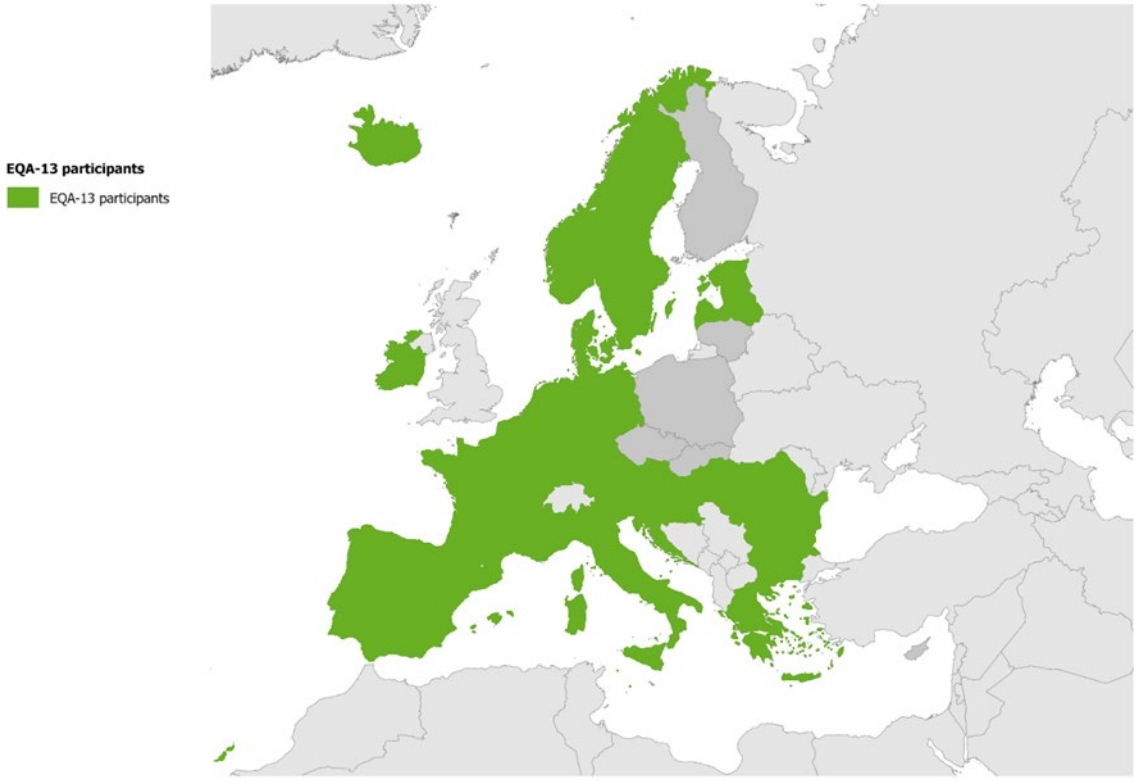
STEC EQA-13 was funded by ECDC and arranged by SSI following ISO/IEC 17043:2010 [10]. EQA-13 included serotyping, virulence gene determination, and a molecular typing-based cluster analysis, and was carried out between May and December 2024.

Invitations were emailed by the EQA provider to ECDC's contact points in the FWD-Net (31 countries) by 12 April 2024, with a deadline to respond by 29 April 2024. In addition, invitations were sent to the EU candidate countries.

Twenty-two NPHRLs in EU/EEA and EU candidate countries accepted the invitation to participate, and all submitted their results (Figure 1, Annex 1, Table 12). EQA test strains were sent to participants between 21 May and 4 June 2024. In Annex 2, participation details in EQA-12 and EQA-13 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 2024 (Annex 12).

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

Figure 1. Countries participating in the 13th round of the external quality assessment (EQA-13) scheme for typing of Shiga toxin-producing *Escherichia coli* (STEC)



EQA-13 participating countries are shown in green. Administrative boundaries: © EuroGeographics © UN-FAO © Turkstat. The boundaries and names shown on this map do not imply official endorsement or acceptance by the European Union.

2.2 Selection of test strains/genomes

Seventeen test strains were selected to fulfil the following criteria:

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

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

Similarly to EQA-12, we included two hybrid *E. coli* pathotype test strains; Shiga toxin-producing and enterotoxigenic *E. coli* (STEC/ETEC) and Shiga toxin-producing and enteroaggregative *E. coli* (STEC/EAEC). As was seen with the emergence of Shiga Toxin producing enteroaggregative *E. coli* (Stx-EAEC), hybrid strains can possess a major challenge for the public health, due to the needs to now implement diagnostic procedures that will identify the most virulent clones. The selected hybrid strains comprised of O187:H28 (STEC/ETEC) and O159:H4 (STEC/EAEC).

Based on the WGS-derived data, the selected cluster of closely related strains consisted of four STEC ST335 strains (including the technical duplicate set strain7/strain11/strain18). Characteristics of all the STEC test strains are listed in Table 1 and Annexes 3-9. The EQA provider found at most zero allele differences or 1 SNPs between any two strains in the cluster (Annex 8). The EQA provider's cluster analysis of WGS-derived data was based on an allele-based (cgMLST [11]) and SNP analysis (NASP [12]). The 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. However, this year none of the participating laboratories used PFGE for cluster identification. An additional eight strains (sequences) for cluster analysis were selected to include strains with different varying relatedness of sequence types (ST335) and other STs. A set of duplicates were included in the test strains (strain7, strain11 and strain18). Two of the sequences were modified by the EQA provider; one sequence with reduced coverage, and one sequence contaminated with 9.3% *E. albertii* (Table 5). The characteristics of all the strains and sequences are listed as 'EQA provider' in Annexes 4–10.

Table 1. Characteristics of test strains and sequences

Method	Serotyping		Virulence profile		Cluster analysis			
No. strains/sequences	12 strains		12 strains		12 strains / 8 sequences			
Annex	3		4		5, 7-8			
Strain ID						ST	QC-status	Cluster
Strain1	Strains for Serotyping	O187:H28	Strains for virulence profile	*stx2g, esta*	Strains/sequences for cluster analysis	200	-	
Strain2		O157:H-/H7		*stx2c, eae*		11	-	
Strain3		O171:H-/H2		*stx2d*		332	-	
Strain4		O146:H21		*stx1c, stx2b*		442	-	
Strain5		O55:H7		*eae*		335	-	
Strain6		O27:H30		*stx2b*		735	-	
Strain7#‡		O55:H7		*stx1a, eae*		335	-	Yes
Strain8		O26:H-/H11		*stx1a, stx2a, eae*		21	-	
Strain9		O145:H-/H28		*stx1a, stx2a, eae*		32	-	
Strain10		O159:H4		*stx2a, aggR*		678	-	
Strain11#‡		O55:H7		*stx1a, eae*		335	-	Yes
Strain12		O91:H14		*stx1a, stx2b*		33	-	
Strain13‡ - sequence	-	O55:H7		*stx1a*		335	A	Yes
Strain14 - sequence	-	O157:H7		*stx2a*		11	A	
Strain15^ - sequence	-	O55:H7		-		335	B/C	
Strain16 - sequence	-	O55:H7		-		335	A	
Strain17 - sequence	-	O55:H7		*stx2a*		335	A	
Strain18#‡ - sequence	-	O55:H7		*stx1a*		335	A	Yes
Strain19 - sequence	-	O55:H7		-		335	A	
Strain20^ - sequence	-	O55:H7		-		335	C	

‡: closely related strains; #: technical duplicates strains; ST: sequence type; ^modified sequences: strain15, a nonCluster sequence contaminated with app. 9.3% E. albertii and strain20, a nonCluster sequence with low coverage; A: Acceptable quality, B: Quality only acceptable for outbreak situations (less good quality) and C: Not acceptable quality – strain not analysed.

2.3 Distribution of strains and sequences

The 12 test strains were blinded and shipped from 21 May 2024 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. Twelve participants received the strains within two days, and 10 within three to five days after shipment, respectively. No participants reported damage to the shipment or errors in the unique strain IDs.

In June 2024, instructions for the submission of results procedure were emailed to the participants. This included the links to the online site for downloading the additional sequences, viewing the empty submission form and uploading the produced FASTQ files.

2.4 Testing

The serotyping part comprised 12 STEC test strains and the purpose was to assess the participants' ability to obtain the correct serotype. The participants could perform conventional serological methods according to suggested protocol [13] or molecular-based serotyping (PCR or WGS). The results of serotyping were submitted in the online form.

The same set of the above 12 STEC test strains were also used to generate the virulence profile. The analyses were designed to assess the participants' ability to obtain the correct virulence profile. The participants could choose to perform detection of *the aggR* (EAEC associated gene), *esta* (ETEC associated gene) *eae* and *stx1* and *stx2,* as well as subtyping of subtyping of *stx* genes (*stx1a, stx1c, stx1d, stx2a, stx2b, stx2c, stx2d, and stx2g*) according to suggested protocol [14, 15]. The results were submitted in the online form.

For the molecular typing-based cluster analysis the participants could choose to use either WGS-derived data or PFGE-derived data. In EQA-13, all the participants 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 was 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 (strain18) 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 (strain15 and strain20) (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-13 BioNumerics (BN) database. The EQA provider contacted two participants in order to ensure they submitted their result, and one additional laboratories were contacted as some of the sequences were uploaded with incomplete and truncated data due to upload

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

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

Molecular typing-based cluster analysis was evaluated according to correct or incorrect identification of the expected cluster of closely related strains based on a pre-defined categorisation by the organiser. The EQA provider's WGS-derived cluster analysis was based on allele-based cgMLST [11] and SNP analysis (NASP) [12]. The cluster categorisation is based on WGS data and the correct cluster delineation might be difficult to obtain by the use of less discriminatory methods, e.g. PFGE. The ST335 cluster comprised four strains or sequences: strain7, strain11, strain13, and strain18, where three were replicates from the same strain (strain7, strain11 and strain18). To simulate real-world data integrity issues, the sequence of non-cluster strain15 was modified by the EQA provider, contaminating the sequence with 9.3% *E. albertii ,* while the sequence of non-cluster strain20 was modified to reduce the overall sequencing coverage. The EQA provider determined that there were, at most, zero allele differences or 1 SNP between any two strains within the cluster.

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

Individual evaluation reports and certificates of attendance were distributed to participants in December 2024. 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 22 participants who signed up, all 22 (100%) completed and submitted their results. Eighty-six percent of the participants (19/22) completed all three parts of EQA-13 (serotyping, virulence determination, and cluster analysis). In total, 21 (95%) of the participants performed serotyping, 22 (100%) participated in the detection of one or more of the virulence genes and 19 (86%) 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	21	22	19
% of participants	95*	100*	86*

[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 (22) of participating laboratories.*

O grouping results were provided by 21 participants (95%) and H typing results were provided by 18 (86%). The majority of participating laboratories used molecular-based serotyping (71%, 15/21 for O group, and 94%, 17/18 for H group), and a minor fraction performed phenotypic serotyping (29%, 6/21 for O group, 6%, 1/18 for H group). None of the participants reported using PCR methods (Annex 3). All participants (100%, 22/22) performed the detection of virulence genes *stx1*, *stx2, eae,* and the detection of the enteroaggregative gene, *aggR.* Slightly fewer participants reported the heat stable enterotoxin gene, *esta* (91%, 20/22). In addition, *stx1* and *stx2* subtyping detection were reported by 86% (19/22) (Annex 4). The majority of the participants performed the cluster analyses (86%, 19/22), all using WGS-derived data (Table 3).

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

	Serotyping n=21		Virulence profile determination n=22					Cluster analysis n=19
	O group	H type	*aggR*	*eae*	*esta*	*stx1* and *stx2*	*stx* subtyping	WGS
Number of participants	21[#]	18Δ	22	22	20	22	19	19
Percentage of participants^	100%	86%	100%	100%	91%	100%	86%	100%
Percentage of participants *	95%	82%	100%	100%	91%	100%	86%	86%

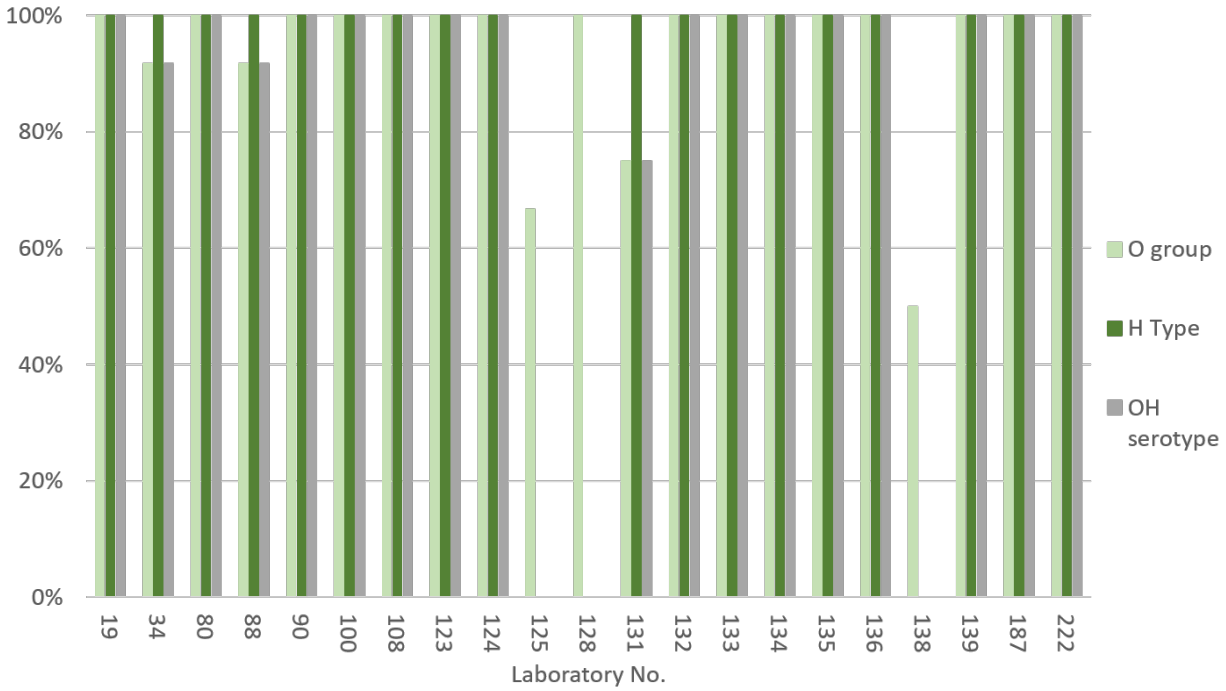
^: percentage of participants in respective part of EQA
**: percentage of total number of participating laboratories (22)*
[#]: phenotypic (n=6)/PCR-based (n=0)/WGS-based (n=15)
Δ: phenotypic (n=1)/PCR-based (n=0)/WGS-based (n=17)

3.2 Serotyping

The majority of participating laboratories took part in O typing (96%, 21/22). Upon reviewing the combined results from the O typing, Strain10 proved to be challenging across participants, with just one laboratory (5%, 1/21) being able to correctly assign the expected O type (O159). Given the systematic difficulties with this strain, it was decided by the EQA provider that all participants that reported the strain as "not typeable", were registered as having correctly typed this strain. With this correction, the average success rate for O typing was 94% (Figure 2).

Eighteen (82%, 18/22) laboratories performed H typing. Of the 21 laboratories participating in O grouping, 86% (18/21) also reported H type. The overall performance for H typing was excellent, and superior to that of the O grouping, with all participating laboratories (100%; 18/18) correctly H typing all 12 test strains (Figure 2).

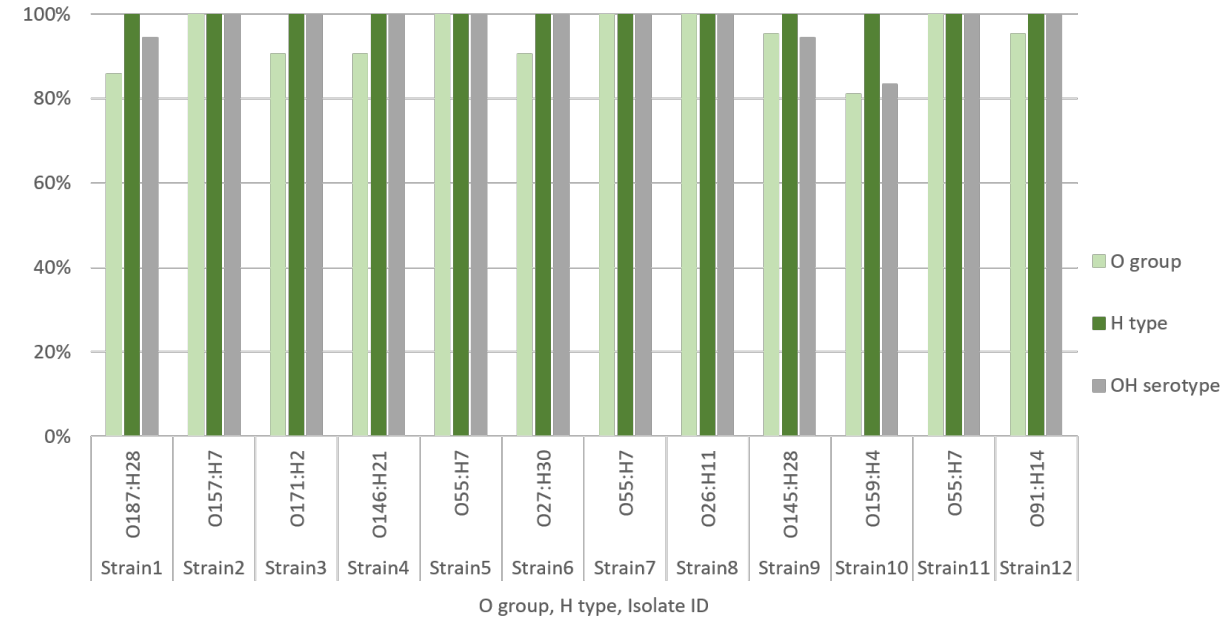
Figure 2. Participant percentage scores for O grouping and H typing



Arbitrary numbers represent participating laboratories.
Bars represent the percentage of correctly assigning O groups (light green), n= 21 participants, H types (dark green), n=18 participants, Combined O:H serotypes (grey), n=18 participants.

Complete O:H serotyping was performed by 18 of the 21 (86%) participants with a high average overall score of 98%. Scores for each individual strain ranged from 83% to 100% correct O:H typing by participants (Figure 3). Correct O:H serotypes of all 12 strains were reported by 15 of the 18 laboratories participating (83%, after correcting for Strain10 O typing difficulties).

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



Bars represent the percentage of laboratories correctly assigning O groups (light green): n=21 participants.
H types (dark green): n=18 participants. Combined O:H serotypes (grey): n=18 participants.
Average scores: O group, 94%; H type, 100% and combined O:H serotype, 98%, for laboratories reporting in each category

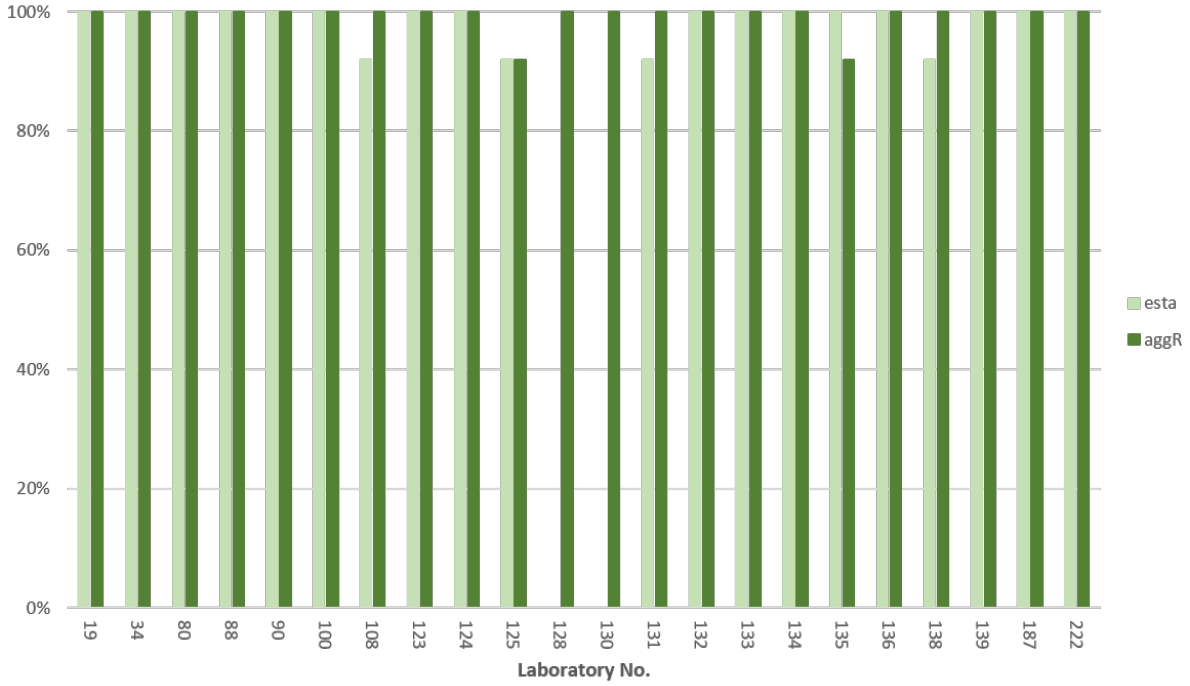
3.3 Virulence profile determination

Most, if not all 22 participants submitted results for the following virulence genes; *aggR* (22 participants), *eae* (22 participants), *esta* (20 participants), *stx1* (22 participants), *stx2* (22 participants), and subtyping of *stx1* (19 participants), and *stx2* (19 participants).

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

Two of the 12 strains included in EQA-13 harboured virulence genes associated with pathotypes other than STEC; Strain1 harbouring the ETEC associated *esta* gene and Strain10 harbouring the EAEC defining gene *aggR*. All laboratories, except for two (125 and 135), correctly identified *aggR* in Strain10, corresponding to correct responses from 91% of the participants. The average lab-specific success rate across all 12 strains was 99% (Figure 4, Annex 4). The ETEC associated *esta* was correctly identified by all but three participants (laboratories 108, 125 and 138) (85%, 17/20), and was erroneously reported for Strain10 by laboratory 131. Thus, 80% of all participants (16/20) reported the expected results, and the average lab-specific success rate across all 12 strains was 98% (Figure 4, Annex 4).

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

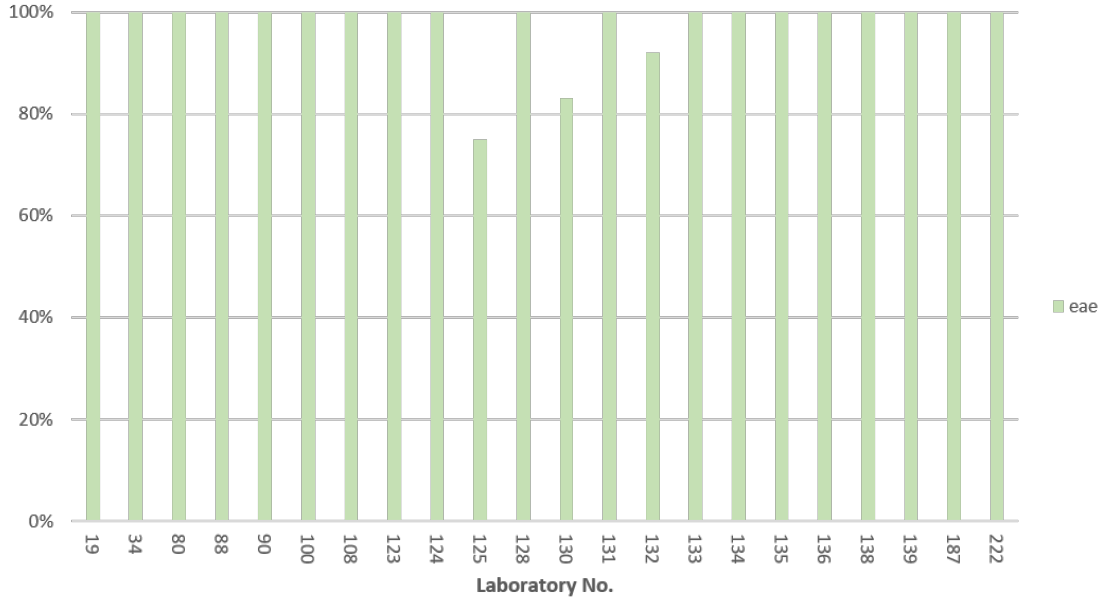


Arbitrary numbers represent participating laboratories.
Bars represent percentage of correct genotyping of esta (light green) n=20 participants and aggR (dark green): n=22 participants. Average scores: esta, 98%; aggR, 99%.

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

Detection of virulence genes *eae*, *stx1* and *stx2* was performed by all 22 (100%) participating laboratories with a high overall performance (Figures 5–6). In EQA-13, six of the 12 included strains were positive for *eae*. These were successfully identified by all but three of the participants (86%, 19/22) (laboratories 125, 130 and 132). For laboratory 125, three false negatives were reported, for laboratory 130 a false negative and a false positive was reported, and for laboratory 132 a false negative was reported. The predominance of false negatives indicates difficulties with detection of *eae* as the main issue. Across all strains, the average lab-specific success rate for *eae* detection was 98% (Figure 5).

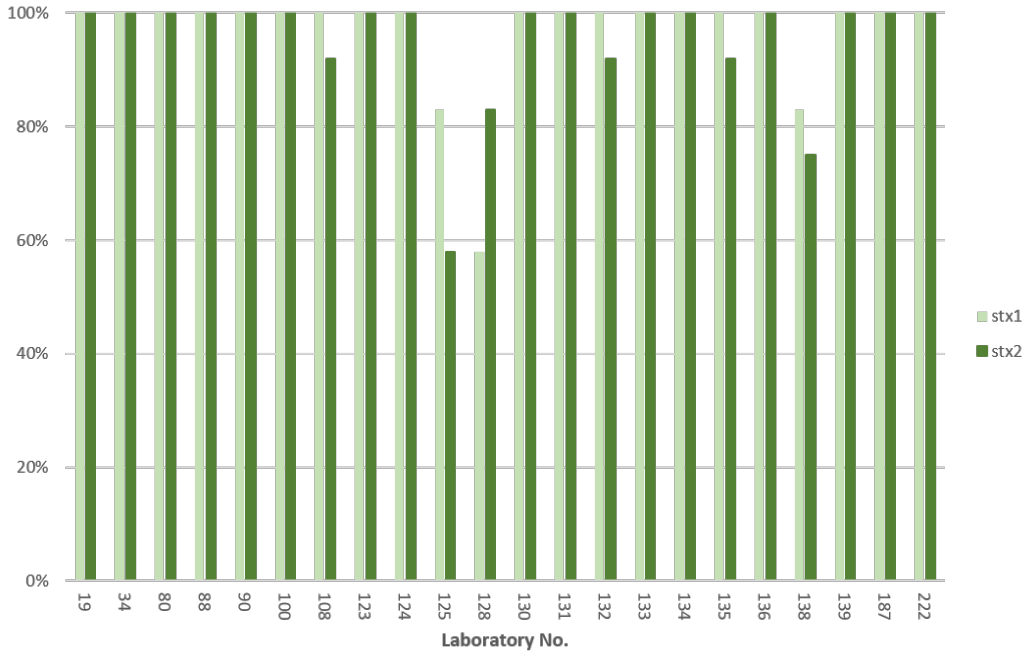
Figure 5. Participant percentage scores for genotyping of *eae*



Arbitrary numbers represent participating laboratories.
Bars represent percentage of correct genotyping of eae (light green): n=22 participants. Average score: eae, 98%.

Eleven of the 12 strains were positive for stx1 or stx2, alone or in combination. The performance for the detection of both *stx1* and *stx2* was high, and 16 laboratories reported 100% accuracy for both *stx1* and *stx2* (Figure 6). There were a total of nine errors in stx1 detection reported by three participants; two from laboratory 125 (false negatives), five from laboratory 128 (all false positives), and two from laboratory 138 (false negatives). For stx2, a total of 13 errors were observed from six participants; one from laboratory 108 (false negative), five from laboratory 125 (all false negatives), two from laboratory 128 (false positives), one from laboratory 132 (false negative), one from laboratory 135 (false positive) and three from laboratory 138 (false negatives). Only Strain6 and Strain9 had multiple (two) laboratories reporting erroneous stx2 results, with the remaining nine errors spread across the remaining strains. Therefore, reported errors were primarily grouped by participant, rather than strain (Annex 4, Table 19-20).

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



Arbitrary numbers represent participating laboratories.
Bars represent percentage of correct genotyping of stx1 *(light green) and* stx2 *(dark green): n=22 participants.*
Average scores: stx1, 97%; stx2, 95%.

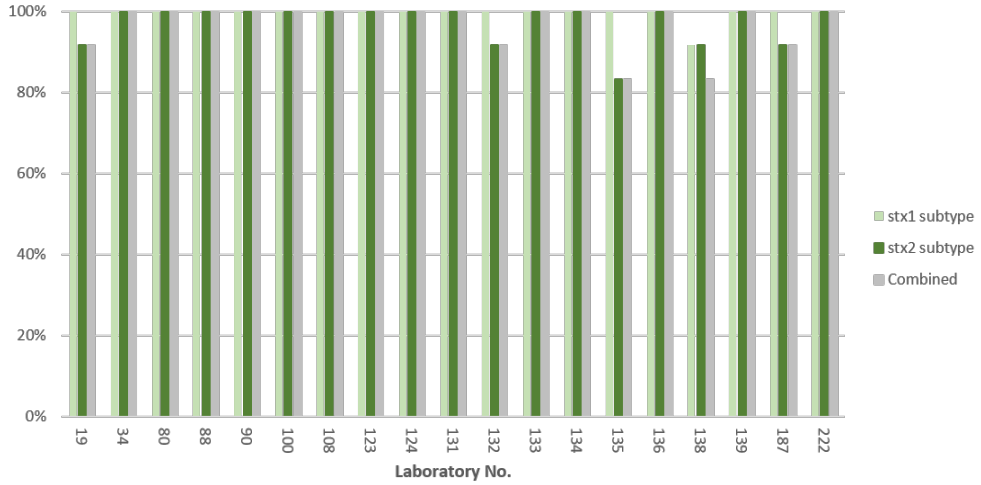
3.3.3 Subtyping of *stx1* and *stx2*

Nineteen of the 22 participants performed subtyping of *stx1* and *stx2* genes (86%). Across all 12 included strains, 18 participants correctly subtyped *stx1* for 100% of the strains, while the remaining participant (laboratory 138) had a 92% success rate, incorrectly reporting an *stx1a* from a negative strain (Figure 7; Annex 4, Table 21). The average success rate rounded up to 100% across participants. For *stx2*, 14 participants (74%, 14/19) correctly subtyped 100% of the strains, four participants (laboratories 19, 132, 138 and 187) correctly subtyped 92% of the strains, and one (laboratory 135) had an 83% success rate (Figure 7; Annex 4, Table 22). The average success rate among participants for *stx2* subtyping was 97%.

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.

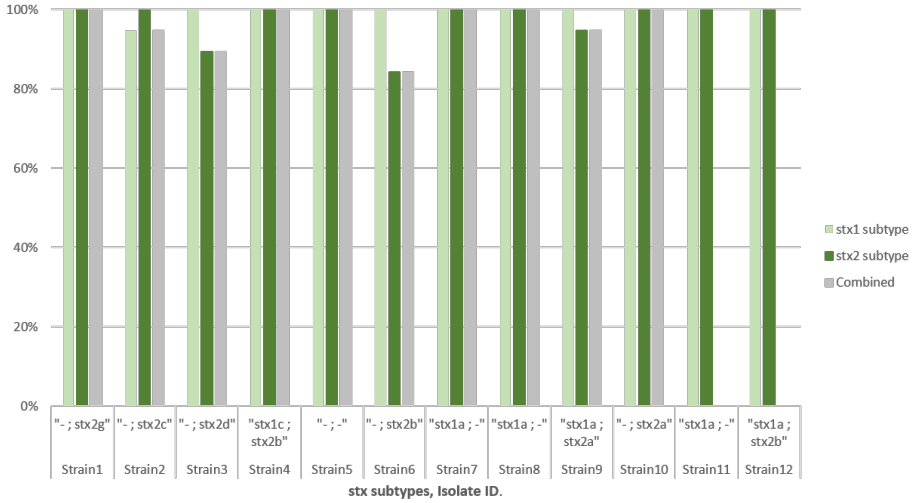
Seventy-four percent of participants (14/19) correctly subtyped both *stx1* and *stx2* for all (100%) of the 12 strains. Three participants (laboratories 19, 132 and 187) correctly subtyped both for 92% of the strains, and two participants (laboratories 135 and 138) were successful for 83% of the strains. The overall average success rate for the combined subtyping was 97%.

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



Arbitrary numbers represent participating laboratories.
Bars represent percentage of correct subtyping of stx1 (light green), stx2 (dark green), combined stx1 and stx2 (grey), n=19 participants.

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



Bars represent percentage of laboratories correctly subtyping stx1 (light green), stx2 (dark green) and combined stx1 and stx2 (grey), n=19. Average scores: stx1, 100%; stx2, 97% and combined stx1 and stx2, 97%.

The incorrect results of the *stx2* subtyping are shown in Table 4, which is divided into two categories: false negatives (1/19), incorrect reported *stx2* subtype 6/19.

Table 4. Incorrect *stx2* subtype results

		Incorrect subtype results		
Strain ID	**EQA provider**	**False negative**	**Incorrect**	**Total true errors**
Strain3	*stx2d*	-	*stx2c; stx2d* (2)	2
Strain6	*stx2b*	1	*stx2b; stx2d (1), stx2a (1)*	3
Strain9	*stx2a*	-	*stx2b* (1)	1
Total		**1**		**6**

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 four *stx1a* producing *E. coli* ST335, based on WGS-derived data (Table 1). The EQA provider's cluster analysis of WGS-derived data was based on an allele-based (cgMLST [11]) and SNP analysis (NASP [12]).

The correct cluster based on WGS-derived data contained four ST3335 strains: strain7, strain11, strain13 and strain18 (strain7/strain11/strain18) were triplicates). As previously mentioned, the strain15 sequence was contaminated with approximately 9,3% *E. albertii* sequence by the EQA provider, while strain20 was modified for low coverage. The EQA provider found at most 0 allele differences or 1 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

Nineteen participants (86%, 19/22) performed cluster analysis using WGS-derived data. All participating laboratories reported using in-house sequencing. The participants reported using different sequencing platforms: NextSeq (10), MiSeq (3), Ion Torrent platforms (2), NovaSeq (2) and MiniSeq (2) (Annex 6). All laboratories reported using commercial kits for library preparation. The predominant kits used were Illumina DNA Prep and Nextera kits (32% each, 6/19) (Annex 6).

3.4.1.2 Assessment of the QC- status of the provided sequences

Participants were instructed to describe their QC observations and considerations when assigning QC status, as well as during the following cluster analysis for the additional test strains (provided genome sequences) strain13-20. A three-tier QC-status was used; 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 (strain15 and strain20): one with sequence contamination and one with low coverage (Table 5).

The manipulations of the two strains, and participant assessments of them were as follows:

Strain15: A non-cluster sequence contaminated with app. 9.3% *E. albertii* sequence data. This contamination was expected by the EQA provider to prompt either a B or a C QC-status. Participant evaluations varied, with 32% (6/19) reporting a QC-status of C, 32% (6/19) reporting QC-status B and finally 37% (7/19) reporting QC-status A. The majority of participating laboratories thus deemed the sequence quality acceptable for downstream analyses.

Strain20: A non-cluster sequence modified for reduced coverage, expected by the EQA provider to result in QC-status C classification. While one laboratory (Laboratory 132, 5%, 1/19) reported QC-status A, deeming the sequence acceptable for further analyses, the remaining participants assigned QC-status C (95%, 18/19), rejecting strain20 from further analyses.

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

Genome	Characteristics	Provider	A	B	C
Strain15	A non-cluster sequence (ST335) contaminated with app. 9.3% *E. albertii*	B/C	7	6	6
Strain20	A non-cluster sequence (ST335) with low coverage.	C	1	0	18

A: Acceptable quality
B: Quality only acceptable for outbreak situations (less good quality)
C: Not acceptable quality
Raw data available in Annex 11

The remaining six provided sequences were reported to be of acceptable quality by all participants, with QC-status A reported in all participant reports, except for three B statuses each (16%, 3/19) reported for strains 14 and 17.

3.4.1.3 Cluster analysis

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

Performance in the cluster analysis with WGS-derived data was excellent (100%), with all participants reporting the clustering pattern of closely related strains, as defined by pre-categorisation from the EQA provider, among the 12 test strains and eight sequences (Table 6). While approximately 1/3 of participants chose to include Strain15 in their analysis, this did not affect their ability to correctly report the expected cluster.

Laboratories were instructed to report the data analysis used for cluster identification and use strain18 (sequence) as a representative in the cluster for reporting SNP distance or allelic differences. Laboratories could report results up to three analyses (one main and up to two additional, but the detected cluster had to be based on results from the main analysis.

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

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

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

Two of the 19 participating laboratories (Laboratories 108 and 132, 11%, 2/19) reported using SNP as the main cluster analysis method, while Laboratory 132 also reported using SNP as additional cluster analysis approach. Laboratory 19 reported using SNP as an additional analysis approach (5%, 1/19). The rest of the participants utilised allele-based approaches, with Laboratory 124 using different allele-based approaches as both main and additional analysis (Table 6). The three laboratories reporting SNP-based cluster analyses used different pipelines, read mappers

and variant callers. As 100% of participants correctly identified the expected cluster, no differences in reported clusters could be attributed to the choice of clustering analysis. Notably, Laboratory 108, using an in-house pipeline, reported a very high range of within-cluster SNP differences. Furthermore, it appears the maximum reported within-cluster distance exceeds the lower boundary of reported distances to non-cluster strains (Table 7, Figure 9). Whether the clustering analysis was supported by alternative methods, is not evident from the reported information.

The two analyses reported by Laboratory 132 appear identical, based on the submitted information, but resulted in different SNP distances within, as well as outside of the reported clusters. From the reported results, laboratory 132 reports including fewer strains in the secondary analyses. Based on this, this could be a strategy of sequential refinement of the cluster analysis. This, however, is speculative, and the underlying reason for these discrepancies are unknown.

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 [12]	Rb	Strain18	BWA	GATK	Yes	0–1	78-178
19*	NASP	Rb	Strain18	BWA	GATK	Yes	0-1	81-186
108	In-house pipeline	Rb	Strain18	CLC assembly cell	CLC assembly cell	Yes	1-1632	457-80787
132	CFSAN SNP Pipeline v2.2.1	Rb	Strain18	Bowtie2	SAMtools	Yes	0-0	5-168
132*	CFSAN SNP Pipeline v2.2.1	Rb	Strain18	Bowtie2	SAMtools	Yes	0-1	86-189

**: additional SNP-based analysis (See Annex 8, Table 25).* *Rb: Reference-based*

Seventeen participants reported using allele-based analyses as the main method for cluster detection (Table 8), with one of them submitting additional allele-based analyses (Laboratory 124), and one submitting additional SNP-based clustering analyses (Laboratory 19). Twelve of the 17 laboratories (71%) reported using only assembly-based allele calling methods, while five (29%, 5/17) reported using assembly- and mapping-based allele calling methods. One of the participants (Laboratory 124) reported only assembly-based as the main analysis method, and an assembly- and mapping-based analysis as additional analysis.

Table 8. Results of allele-based cluster analysis

Lab No.	Allele-based analysis							
	Approach	Allelic calling method	Assembler	Scheme	No. of loci	Identified Pre-defined Cluster	Difference within cluster	Difference outside cluster
Provider	BioNumerics	A&M	SPAdes	Applied Mathss (cgMLST/Enterobase)	2513	Yes	0–0	26-2326
19	BioNumerics	A&M	SPAdes	Applied Maths (cgMLST/Enterobase)	2513	Yes	0-0	19-2320
34	SeqSphere	OAB	SKESA	Applied Maths (cgMLST/Enterobase)	2513	Yes	0-0	26-2315
80	SeqSphere	OAB	Skesa v2.4.0	Enterobase (cgMLST)	2513	Yes	0-0	26-2321
88	INNUca, chewBBACA and ReporTree	OAB	SPAdes v3.14.0[a]	INNUENDO wgMLST	7601	Yes	0-4	46-2785
90	SeqSphere	OAB	SKESA	Enterobase (cgMLST)	2513	Yes	0-0	26-2315
100	SeqSphere	OAB	SPAdes	SeqSphere Escherichia/Shigella cgMLST v1 scheme	2513	Yes	0-1	27-2321
123	SeqSphere	OAB	SPAdes	Enterobase (cgMLST)	2513	Yes	0-0	26-2321
124	BioNumerics	OAB	SPAdes	Applied Maths (cgMLST/Enterobase)	2506	Yes	0-0	26-2320
124*	Enterobase	A&M	SPAdes	Enterobase (cgMLST)	2513	Yes	0-1	30-2348
131	SeqSphere	OAB	SKESA	Enterobase (cgMLST)	3152	Yes	0-0	71-5243
133	BioNumerics	A&M	SPAdes	Applied Maths (cgMLST/Enterobase)	2513	Yes	0-0	27-200
134	SeqSphere	A&M	SPAdes v3.15.4	Enterobase (cgMLST)	2513	Yes	0-0	26-2319
135	SeqSphere	OAB	SPAdes	Enterobase (cgMLST)	2513	Yes	0-0	26-2319
136	SeqSphere	A&M	Unicycler	Enterobase (cgMLST)	2513	Yes	0-1	26-8631
138	chewbbaca	OAB	shovill v1.1.0	Enterobase (cgMLST)	2513	Yes	0-2	29-2280
139	Enterobase	A&M	SPAdes	Enterobase (cgMLST)	2513	Yes	0-3	30-2348
187	SeqSphere	OAB	SPAdes	Enterobase (cgMLST)	2513	Yes	0-0	27-2329
222	PHANtAsTiC pipeline	OAB	SPAdes v3.15[b]	Innuendo-curated Enterobase scheme	2360	Yes	0-7	32-2207

**: additional analysis*
A&M: Assembly- and mapping-based
OAB: Only assembly-based
OMB: Only mapping-based
[a]: Implemented in INNUca v4.2.2
[b]: Implemented in PHANtAsTiC pipeline
(See Annex 8, Table 26).

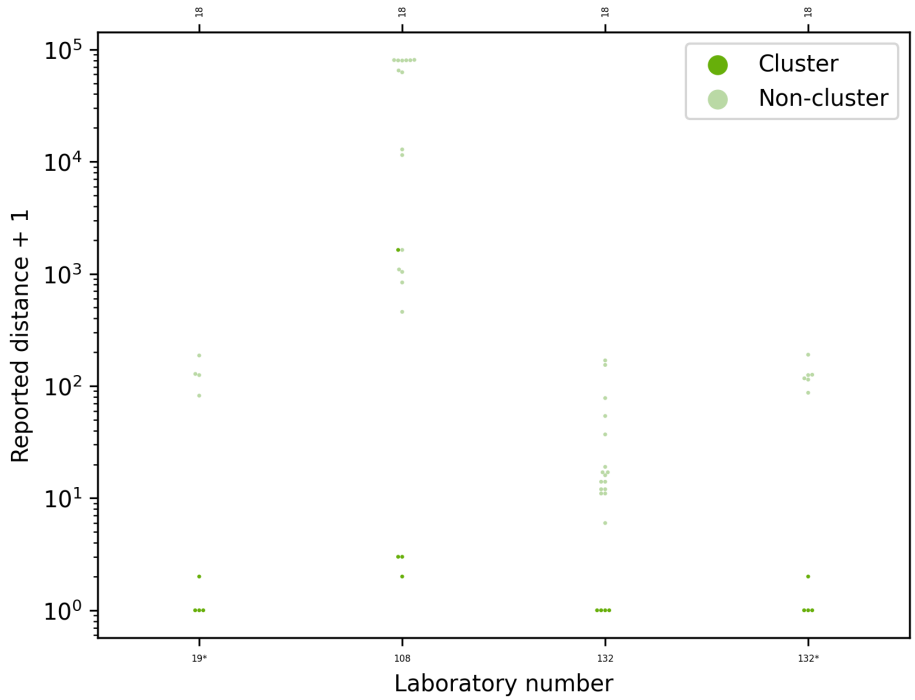
When observing the collected (main and additional) reported analyses, the majority of participant entries utilized the SeqSphere platform (56%, 10/18), followed by BioNumerics (17%, 3/18), Enterobase (11%, 2/18) and chewBBACA, the PHANtaAsTiC pipeline and a combination of INNUca, chewBBACA and ReporTree (each 6%, 1/18). Twelve of the 18 utilised SPAdes as the assembler (67%), with SKESA being the second most reported (22%, 4/18). The remaining two participants used the SPAdes optimisers shovill and Unicycler (each 6%, 1/18).

Reported allele-based clustering analyses predominantly used the Enterobase cgMLST scheme, whether directly reporting this, or as accessed through Applied Math applications (83%, 15/18). For all but two (Laboratories 124 and 131 reported 2506 and 3152, respectively), this resulted in 2513 loci for use in the cgMLST. The three remaining analysis pipelines used INNUENDO wgMLST (Laboratory 88, 7601 loci), SeqSphere Escherichia/Shigella cgMLST v1 scheme (Laboratory 100, 2513 loci) and an Innuendo-curated Enterobase scheme (Laboratory 222, 2360 loci). None of these differences in approach, however, had any apparent influence on successful cluster identification.

The number of loci included in the wg- or cgMLST schemes also did not seem to alter the resolution of the allele distance between cluster and non-cluster isolates. Allele distances within the identified clusters were reported to be between 0-7 (1.06 allele difference average, reported maximum distance), indicating different, successful clustering

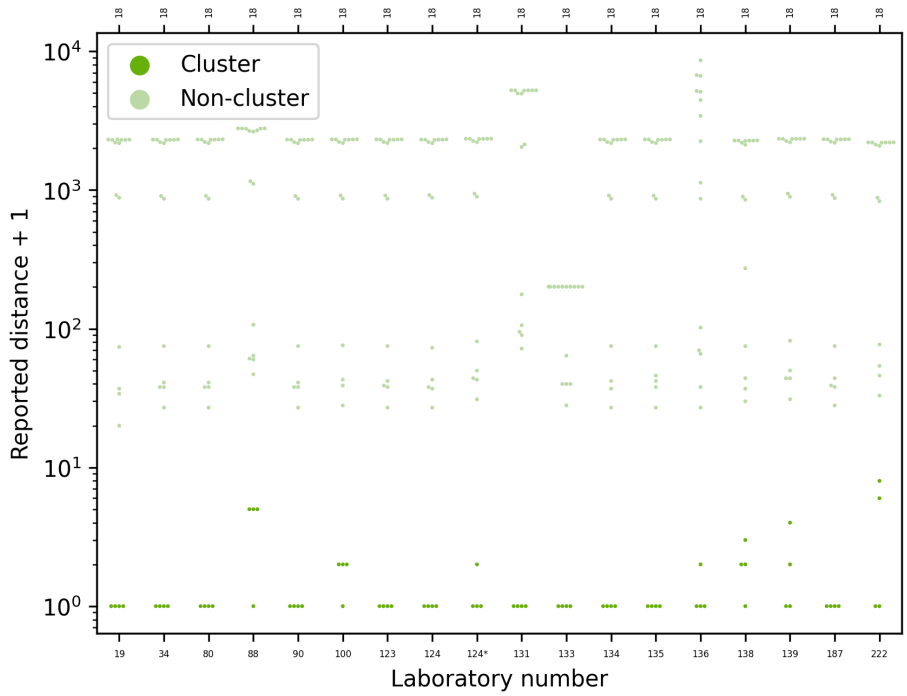
thresholds between participants (Table 8, Annex 8). Allele distances outside of the clusters were generally reported to be similar. However, a few notable differences were observed. Laboratory 133 reported a maximum difference of just 200 allele differences outside of the reported cluster. This was lower than the other participants by more than a factor of 10, but is likely due to a software-imposed cut-off at 200 rather than the actual maximum. At the other end of the spectrum, Laboratories 131 and 136 reported maximum allele differences outside of the clustering strains that exceed the numbers of loci in the Enterobase cgMLST scheme they both report using.

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



SNP: single nucleotide polymorphism; Participants were instructed to select Strain18 as reference (listed as '18' on the top scale). Dark green: reported cluster of closely related strains, Light green: not reported as part of cluster.

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



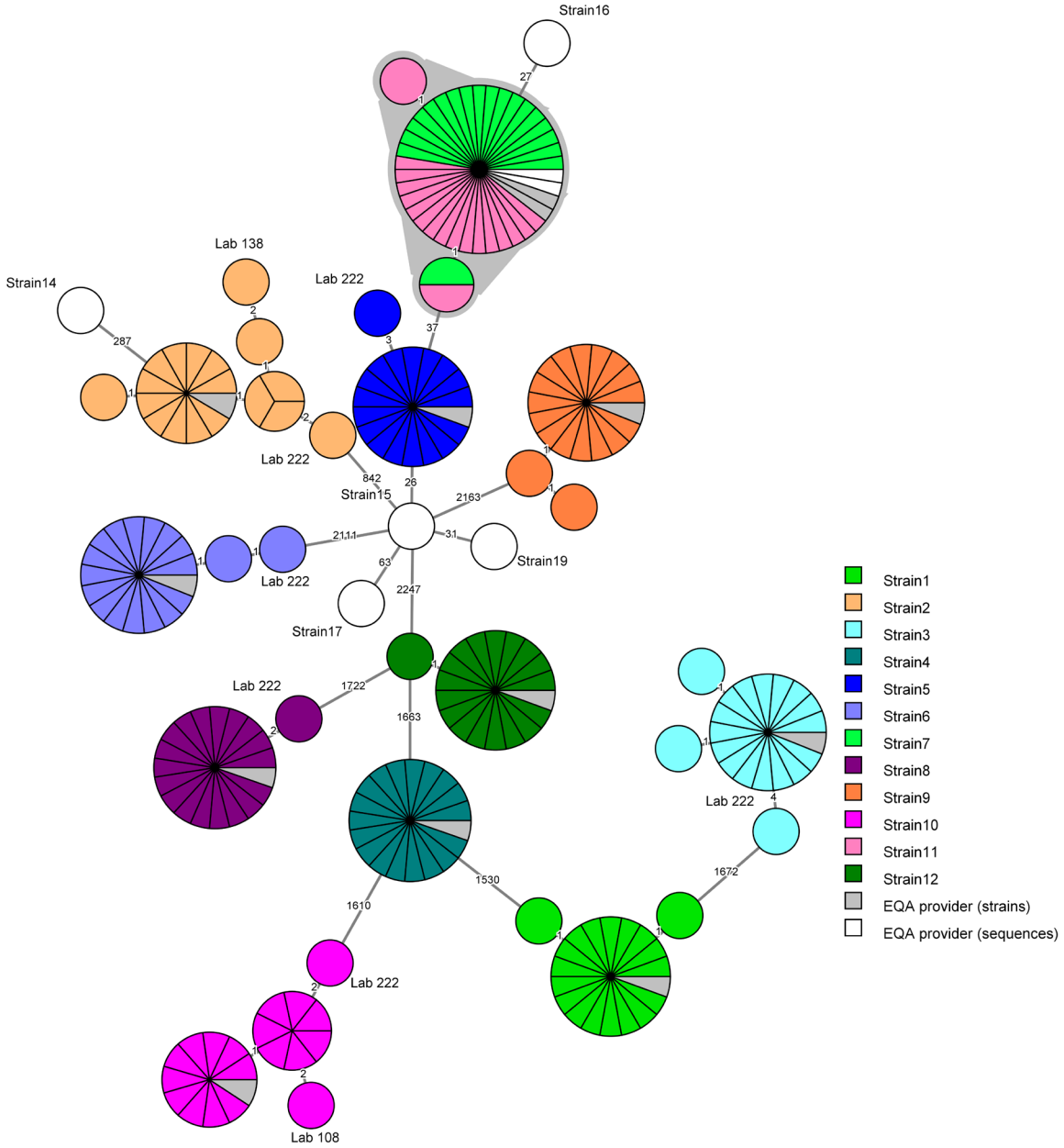
Participants were instructed to select strain18 as reference (listed as '18' on the top scale). Dark green: reported cluster of closely related strains, Light green: not reported as part of cluster.

3.4.1.4 Analysis of raw reads uploaded by participants

In addition to the reported cluster identification, participants submitted their FASTQ files to be evaluated by the EQA provider. The FASTQ files were then uploaded to an Applied Maths calculation engine for allele calling (Enterobase) [11] and evaluated by the EQA provider's in-house quality control (QC) pipeline [16].

The overall cgMLST analysis, shown in the minimum spanning tree (MST) based on submitted raw reads (FASTQ files) from the 19 laboratories, indicates a clear clustering of provided and submitted sequences for each test strain (Figure 11). Laboratories 108 and 222 were generally found to be >1 allele differences from each main strain cluster, likely due to differences in sequencing technology, as both used Ion Torrent sequencing, while the rest of the participants, as well as the EQA provider used Illumina sequencing.

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

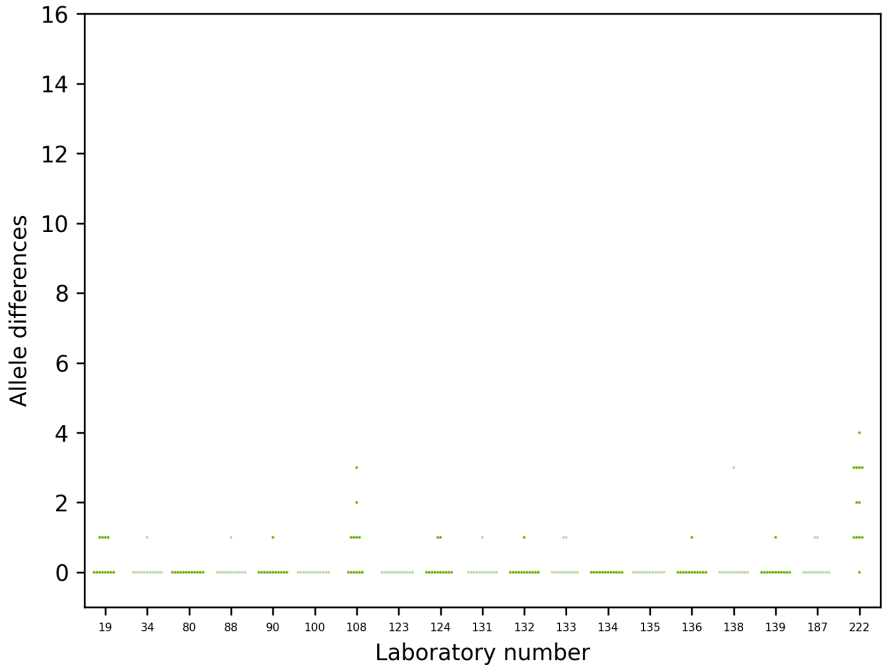


Minimum spanning tree (MST) in log scale of core genome multilocus sequence typing (cgMLST) [11] based on submitted raw reads (FASTQ files). Each of the strain1–12 test strains have a different colour. The EQA-provided sequences for strain1-strain12 from the EQA provider are in grey, and the provided sequences (strain13-20) are in white. Strain7, strain11 and strain18 were technical replicates. Supplied sequences that differ >1AD from EQA provider strains are labelled with laboratory number. A total of eight sequences were excluded, as the cgMLST loci percentage was below 95% (seven from laboratory 132, as well as one of the provided sequences modified for reduced quality (strain20)). Strain15 (modified to simulate contamination) was included in the MST analysis. Results from laboratories 108 and 222 were run in CE, using the Ion Torrent setup for allele calling.

Please note that the allele differences in Figure 11 do not exactly match those illustrated in the participant-specific, individual reports, and consequently there are discrepancies between these and Figure 11, although the same data are used. This discrepancy is caused by loci being dropped if they did not pass QC for all strains in the joint, inter-laboratory analysis. As a result, the joint analysis contains fewer loci.

In addition to the MST, for each laboratory, a separate cgMLST was performed on the submitted raw reads (FASTQ files), applying Applied Maths allele calling with the Enterobase scheme [11]. A hierarchical single linkage clustering was performed on the submitted data for each laboratory along with the EQA provider's reference strains (Stain1-12). Figure 12 shows the allele differences between each submitted sequence and the corresponding reference.

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



Allele difference from corresponding stain1-12 (EQA provider) based on submitted raw reads (FASTQ files) and analysed by EQA provider.

For 193 of 228 results (85%), no allele difference was identified between submitted sequences and the EQA provider reference sequences. For 25 results (11%, 25/228), a difference of one allele from the reference strain was calculated, and for 10 results (4%, 10/228), a difference of two to four alleles was observed. These differences were primarily reported by Laboratories 108 and 222, and are again likely attributed to differences in sequencing technologies.

Separately, the laboratories listed quantitative and qualitative QC parameters used to evaluate their data. As seen in Table 9, all laboratories have implemented QC thresholds for accepting the data, to some extent. All participants reported using confirmation of genus, and coverage with reported acceptance thresholds at >20->100X. Genome size thresholds were used by all but one participant (95%), while Q score parameters and the number of good cgMLST loci were reported by 79% of participants, each. The additional QC parameters reported by the participants are listed in Annex 9.

Table 9. Summary of selected QC parameters reported by participants

Laboratory	Confirmation of genus	Coverage	Q score (Phred)	Genome size	No. of good cgMLST loci
19	Kraken and Bracken analysis and <5% contamination with other genus	Minimum x 50 coverage	No	4640000-5560000	Minimum 95% core percent and maximum 30 loci with multiple consensus
34	KRAKEN	>75fold	Q>30	>5 kb	>90%
80	Kraken2/rMLST in In-house script and Mash in SeqSphere	>50	>=30	4.4 - 5.3	>90%
88	Kraken2 (as implemented in INNUca v4.2.2) and ConFindr.	INNUca v4.2.2 employs several coverage thresholds throughout the analysis (15x for the first estimated coverage; 30x for the assembly coverage).	INNUca v4.2.2 performs read quality control with FastQC and trimming/filtering of the reads with Trimmomatic (default settings).	INNUca v4.2.2 uses genome size as a QC criteria (we set 5.0 Mb as the expected genome size).	Allele calling was performed with chewBBACA v3.3.4. We excluded loci called in <90% of the samples and samples with <95% loci called with ReporTree v2.5.3 during the clustering analysis.
90	PubMLST rMLST	CGE KmerFinder	Ridom Mash Distance	40x	No
100	KmerFinder	30x	q20	4.5 to 5.5 Mb	90%
108	BLAST against database of reference genomes	Coverage >20x	No	4.8-6.0 Mbp	No
123	Contamination Check (Mash Screen) in SeqSphere	>50	>98	5.0-6.0	>98
124	Length GC% and in silico PCR *E.coli* det	>100 (acceptable >30 in BioNumerics)	Q30 > 60	3.9 Mb - 6.5 Mb	% alleles called available in BioNumerics (>80%)
131	Mash Screen (SeqSphere included)	50x<	30<	4.5-5.5 Mb	95%<
132	kraken2	Assembly coverage, expect above 10-15 on most contigs	20	Fastqc	No
133	In-built in BioNumerics	>30x	>30	Between 5.000.000bp - 5.800.000bp	At least 95% of good cgMLST loci
134	Mash screen score (implemented in SeqSphere)	50X	trimming based on the PHRED score at the 3´-end (selected value: 20)	expected genome size +/- 10%	<97% = failed; 97-98% = warning; >98% = good
135	Kraken2/Bracken	>30	>30	Between 4.6 - 5.8 Mb	>95%
136	K-mer	50x	No	5	99 % good targets
138	Kraken2 with database built from all refseq genomes; rMLST from Pubmlst	in house calculation. Our threshold for adequate coverage for *E.coli* is 60x	>=q30; fastp 0.22.0.	(4909000 < x ≤ 5493000)	<2% Loci missing
139	In house blastn based script	> 45 X	Discarding reads with Q scores < 15 on a minimum length of 50 bp	Range from 4.7 to 5.9 Mb	No
187	RefSeq Masher	95	32	/-20% genome size	95

Laboratory	Confirmation of genus	Coverage	Q score (Phred)	Genome size	No. of good cgMLST loci
222	No mismatches in the alignment with the 7 housekeeping genes of MLST panel (Warwick); Kmerfinder with only E. coli with Template Coverage >= 10	Minimum 50x average depth of coverage across the genome was considered as threshold value	No	No	Quality threshold for reliability of cluster analysis was set at at least 80% of loci found out of those part of the scheme (1880/2360)
% of laboratories using the QC parameter	**100%**	**100%**	**79%**	**95%**	**79%**

See Annex 9 for additional information.

For each laboratory, the submitted raw reads (FASTQ files) were evaluated by the EQA provider's in-house quality control pipeline [16]. An overview is given in Table 10. For the full QC evaluation of all strains, see Annex 10. According to the QC parameters, the sequencing quality was uniformly good, despite issued warnings. The majority of these warnings are issued due to genome sizes that exceed those expected by the Bifrost QC pipeline. Given the consistent violation of the size intervals, one can argue that these intervals are too restrictive, and that the issued warnings therefore should be disregarded. Only Laboratory 132 had sequences that failed in the EQA provider QC-pipeline (Annex 10, Table 38).

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

*Ranges**	*(Ec)*			*(5%)*	*(4.46-5.56)*	*(<250)*	*(>0)*	*(<1000)*	*(>50)*					
Lab No.	**Detected species**	**% Species 1**	**% Species 2**	**Unclassified reads (%)**	**Length at >25 x min.**	**Length (1-25) x min. coverage (kbp)**	**No. of contigs at 25 x min. coverage**	**Contigs at (1,25)X coverage**	**Average coverage**	**No. of reads (x 1000**	**Average read length**	**Average insert size**	**N50 (kbp)**	**QC status (Bifrost)**
19	Ec, Se	88.6-97.0	0.1-5.9	2.7-7.3	5.0-5.5	12.3-61.9	353.0-639.0	16.0-73.0	77.0-159.0	2811.0-5754.0	145.0-148.0	286.0-346.0	19.0-31.0	W
34	Ec	86.5-96.4	0.4-3.6	3.0-9.4	5.1-5.6	0.0-0.0	85.0-287.0	0.0-0.0	153.0-287.0	5606.0-10741.0	151.0-151.0	279.0-330.0	115.0-326.0	W
80	Ec	89.3-98.8	0.1-2.3	1.1-7.3	5.1-5.6	0.0-0.9	95.0-304.0	0.0-2.0	61.0-193.0	2319.0-7236.0	144.0-149.0	230.0-492.0	115.0-241.0	W
88	Ec, Se	87.9-97.6	0.1-6.1	2.0-9.8	5.1-5.5	2.7-43.9	85.0-350.0	1.0-34.0	74.0-121.0	2716.0-4328.0	146.0-149.0	386.0-470.0	51.0-150.0	W
90	Ec	87.7-96.8	0.1-1.5	2.6-12.0	5.1-5.6	0.0-5.6	79.0-261.0	0.0-5.0	67.0-151.0	1568.0-5727.0	143.0-238.0	170.0-375.0	101.0-314.0	W
100	Ec, Se	86.3-99.3	0.0-11.4	0.6-4.8	5.1-5.6	0.0-26.3	122.0-366.0	0.0-13.0	75.0-256.0	1546.0-6316.0	218.0-258.0	220.0-293.0	114.0-313.0	W
108	Ec	91.4-97.8	0.4-2.9	1.4-5.2	5.0-5.6	0.0-2.4	439.0-1913.0	0.0-11.0	98.0-119.0	1691.0-2000.0	293.0-316.0	0.0-0.0	5.0-29.0	W
123	Ec, Se, Sf	65.4-97.4	0.2-26.6	2.0-9.5	5.1-5.6	3.0-26.9	153.0-365.0	2.0-30.0	61.0-107.0	1317.0-2302.0	257.0-271.0	307.0-370.0	40.0-123.0	W
124	Ec	86.8-96.4	0.1-1.6	3.2-9.7	5.1-5.6	0.0-0.0	75.0-228.0	0.0-0.0	152.0-290.0	5544.0-10000.0	151.0-151.0	376.0-527.0	115.0-326.0	W
131	Ec	89.5-96.8	0.1-2.2	2.7-7.4	5.1-5.6	0.0-0.0	88.0-267.0	0.0-0.0	61.0-137.0	2251.0-4789.0	147.0-148.0	329.0-363.0	115.0-184.0	W
132	Ec, Se, Sf	78.4-98.5	0.0-15.0	1.3-12.0	0.4-5.3	141.1-4769.8	56.0-246.0	13.0-1222.0	18.0-62.0	698.0-2293.0	149.0-150.0	349.0-483.0	8.0-210.0	W
133	Ec, Se, Sf	79.6-99.5	0.0-9.0	0.4-6.9	5.1-5.6	0.0-61.2	73.0-245.0	0.0-20.0	44.0-81.0	813.0-1582.0	283.0-293.0	369.0-479.0	114.0-314.0	W
134	Ec	88.2-97.5	0.1-4.5	1.8-7.5	5.1-5.6	0.0-0.0	88.0-306.0	0.0-0.0	43.0-93.0	1679.0-3218.0	151.0-151.0	249.0-336.0	112.0-206.0	W
135	Ec	89.3-98.8	0.1-2.3	1.1-7.3	5.1-5.6	0.0-0.9	95.0-304.0	0.0-2.0	61.0-193.0	2319.0-7236.0	144.0-149.0	230.0-492.0	115.0-241.0	W
136	Ec	89.2-98.1	0.1-2.3	1.7-6.8	5.1-5.6	0.0-0.0	76.0-272.0	0.0-0.0	225.0-685.0	7847.0-27101.0	137.0-148.0	197.0-381.0	115.0-314.0	W
138	Ec	86.6-96.0	0.2-1.9	3.1-8.3	5.1-5.6	0.0-4.0	103.0-399.0	0.0-13.0	189.0-299.0	6941.0-11390.0	151.0-151.0	175.0-288.0	104.0-241.0	W
139	Ec	87.1-95.1	0.3-2.1	4.3-8.7	5.1-5.6	0.0-18.0	117.0-328.0	0.0-15.0	88.0-437.0	3283.0-15363.0	147.0-148.0	342.0-376.0	55.0-140.0	W
187	Ec, Se, Sf	65.8-98.0	0.0-12.2	1.8-17.5	5.1-5.6	0.0-0.0	88.0-336.0	0.0-0.0	51.0-198.0	1826.0-7426.0	146.0-149.0	228.0-432.0	114.0-188.0	W
222	Ec	92.6-98.4	0.4-1.3	0.9-4.2	5.0-5.5	0.0-0.0	359.0-1206.0	0.0-0.0	157.0-271.0	2643.0-4572.0	264.0-361.0	0.0-3.0	8.0-34.0	OK

**: indicative QC range; Ec: E. coli; Se: S. enterica; Sf: S. flexneri; W: One or more warnings were noted in the submitted sequences (see Annex 10).*

3.5 Feedback survey – evaluation of the EQA scheme

After the individual reports were sent to the participants, the EQA provider circulated a feedback survey to assess the STEC EQA scheme. The questionnaire contained questions related to accreditation, information on the individual report, actions taken if errors were detected, the usefulness of the QC evaluation of the participant-sequenced data, the relevance of including low-quality data, and suggestions for improvements. The survey response rate was 59% (13/22). The survey results are summarised in Table 11.

Based on the feedback-survey, we conclude that the assessment of the QC of the participants submitted sequences is being appreciated and needed. Two laboratories had feedback regarding reporting of data submission; streamline the reporting form for ease of use, particularly reporting the virulence gene determination and toxin subtype in the same step. Where another laboratory suggested to extend the QC evaluation to test for detection of STEC from mixed cultures, with different or same species as added complexity.

Table 11. Results of evaluation of the EQA scheme

Questions	Response (Yes)	Comments /actions
1) Used for accreditation/licensing purposes?	10/13 (77%)	One laboratory reported that their WGS STEC pipeline is not yet accredited, but during the future accreditation process these results will be included.
2) Satisfied with the format/comments?	13/13 (100%)	
3) Differed any of your analytical test results (*) with the expected results. Can you specify which corrective action(s), if any, was/were/will be taken	3/13 (23%)	
4) Usefulness of the manipulated sequences?	12/13 (92%)	One laboratory reported that if they had deviating results they will take a look on their procedures and see if they have to do some adjustments.
		One laboratory reported that they absolutely need this procedure for validating of their QC criteria
5) Usefulness of the QC-status of your submitted sequences?	13/13 (100%)	One laboratory reported that they received deviating results and will have a look at their procedures to see if they can do some adjustments.
6) Improvements/remarks		One laboratory reported that it would help if the reporting on the online form was not so extensive. It would e.g. be easier to be able to report virulence genes and toxin subtype in the same step.
		One laboratory suggested it would be an idea if the EQA could test for detection of mixed cultures, different species and same species.
		One laboratory suggested to try to make the results submission form easier to compile. The questions are too many and compiling it requires a lot of time.

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

4 Discussion

Based on the completed evaluation, most participants were satisfied with the format of the individual report and the additional feedback from the EQA provider. However, the suggestion of using a mixed culture would target the diagnostic laboratories rather than the reference laboratories in the EU. The inclusion of the modified sequences in the cluster analysis and the QC feedback of the uploaded sequences was well received by most of the participants. The suggestions are listed in Section 6, 'Recommendations'.

4.1 Serotyping

In EQA-13, 95% of the laboratories took part in the serotyping component. Of these, 24% provided phenotypic serotyping results (5/21), while 76% provided molecular serotyping results using WGS. The percentage of laboratories using phenotypic serotyping was the same as in EQA-12 (24%).

In EQA-13, 18 laboratories engaged in complete O:H serotyping, which marked a decrease from EQA-12 where there were 22 participants. Among these, 83% (15 of 18) correctly identified the serotype for all 12 test strains. This represents a slight decrease from EQA-12, where 72% (16 of 22) accurately assigned the serotypes for all 12 test strains for both O and H.

4.1.1 O group

When looking at the O group participation in previous EQAs, we observed an overall decrease from EQA-4 through EQA-10 (93%; 90%; 90%; 90%; 92%; 83% to 81%). However, in EQA-11, EQA-12 and EQA-13, we saw an increase in the participation of O-typing (96%, 96% and 95% laboratories).

One of the strains included in EQA-13 (Strain10), proved more challenging to O-type than the rest. Only one participant successfully typed it as O159, while two participants incorrectly typed it as O104, one typed it as O92, one typed it as O139, and 16 did not report a typing result for Strain10. Thus, when comparing the percentage of participants reporting all-correct O-typing results for all 12 EQA strains, this percentage would be (5%, 1/21), notably lower than that of EQA-12 (68%, 17/25). It should be noted, however, that participants, on average, correctly O-typed 88% of test strains. In fact, when excluding Strain10, each strain was successfully O-typed by 95% of participants. Given the near universal difficulties with Strain10, the EQA provider made the decision to treat the reported answer "not typeable" as a correct answer. This decision was made as the EQA provider acknowledges that the O159 antisera from SSI Diagnostica often has a low titre, which can influence the phenotypic result. Furthermore, only one sequence of O159 (EU294176) is included in the CGE tool, SerotypeFinder 2.0, which can influence the in-silico O grouping Following this, 76% of participants are noted as having correctly O typed all 12 strains. For one laboratory, phenotypic testing of strain1 (O187) gave an incorrect O103 as result. This is likely due to cross-reactivity between these two O-groups. For strain9 (O145), one lab reported the strain as O146, following phenotypic analysis. This could simply be a typo. Apart from these, all participants were able to successfully O-type Strain2 (O157) and Strain8 (O26), both of which are among the most commonly observed serogroups in the EU/EEA, as well as strains 5, 7 and 11 (O55) [1].

Similar to EQA-12 and EQA-11, not all the incorrect O group results were reported by laboratories using phenotypic methods. Laboratories 34 and 88 used a WGS based method and did not determine all O groups correctly (Annex 3). Six of the 31 (19%) incorrect results were reported as an incorrect O group, while the remaining (81%) were reported as "not typed". This is similar to the O grouping from EQA-12, where 31 incorrect results were reported, of which 10 (31%) were incorrect O-type and the remaining 69% were non-typeable/rough or not done.

4.1.2 H type

The average performance for correctly H-typing the 12 tests strains in EQA-13 was higher than any previous year. All of the 18 laboratories participating in the H-typing correctly identified all H-types (100%). Previous EQAs have had a slightly lower score (EQA-12 91%, EQA-11 84%, EQA-10 94%, EQA-9 94%, and EQA-8 92%). However, there was a decrease in H-typing participation (18 laboratories) compared to EQA-12 (22 laboratories) and EQA-11 (19 participants). The general performance for correctly reporting the H type, of all 12 test strains, was higher (100%) than the O grouping (76%). This might be explained by fewer participating laboratories and that the majority (17/18) used WGS-based methods.

4.1.2 OH serotyping

Complete O:H serotyping was performed by 18 (86%, 18/21) participants with an average overall score of 98%, and for each strain the score ranged from 83% for Strain10 (O159:H4) to 100% for Strain2 (O157:H7), Strain3 (O171:H2), Strain4 (O146:H21), Strain5 (O55:H7), Strain6 (O27:H30), Strain7 (O55:H7), Strain8 (O26:H11),

Strain11 (O55:H7) and Strain12 (O91:H14). The correct serotype of all 12 strains were reported by 83% (15/18) of the participants who performed the O:H serotyping (Figure 3, Annex 3).

The average percentage O:H serotyping in this EQA was, higher (98%) compared to EQA-12 (97%), EQA-11 (95%), EQA-10 (94%), EQA-9 (92%), EQA-8 (86%), EQA-7 (71%), and EQA-6 (78%). This year, the less common European serotype O159 was particularly difficult to identify, especially with phenotypic methods, as the O159 antisera from SSI Diagnostica has a low titre, and SerotypeFinder 2.0 includes only one O159 sequence (EU294176), limiting in silico O grouping.

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

4.2 Virulence profile determination

Between 19-22 (86-100%) participants took part in the detection of various virulence genes. The performance was generally strong, however with some variation between participants. The percentage of laboratories participating in the genotypical detection was generally higher than that of EQA-12: *stx1* (100%), *stx2* (100%), *eae* (100%), *aggR* (100%), *esta* (91%), and the combined participation for subtyping of *stx* genes was (86%). Only *esta* detection and *stx*-subtyping participation saw a decrease from EQA-12 (98% and 93%, respectively in EQA-12).

4.2.1 Detection of *aggR* and *esta*

The performance in detection of the EAEC *aggR* gene was high, with 91% of the participants correctly identifying *aggR* (20/22). This is a slightly lower correct detection rate compared to EQA-12 (98%) and EQA-11 (95%). Seven laboratories utilised another method than WGS to detect the *aggR* gene. Twenty laboratories participated in the detection of the *esta* gene. The performance was slightly lower than for the aggR gene, with 80% (16/20) of laboratories correctly identifying the *esta* gene. The average performance for *esta* was lower than in EQA-12 (98%) and EQA-11 (89%). This performance was attributed to three laboratories (108, 125 and 138) that could not identify the gene in strain1 and one laboratory (131) that wrongly identified the gene in strain10. Five laboratories utilised another method than WGS to detect the *esta* gene.

4.2.2 Detection of *eae*

In EQA-13, half of the supplied strains were positive for *eae*. This was correctly identified by all but three participants (86%). One of these laboratories used a WGS-based detection method, while the two other laboratories used another method for detection. As such, the performance was higher than that of the preceding EQA, where 79% of participants were able to identify all *eae*-positive strains. Overall, an average success rate of 98% was obtained in EQA-13, which is in line with those reported for EQAs 4-12 (96-99%). The detection errors were predominantly false negatives.

4.2.3 Detection of *stx1* and *stx2*

With a perfect participation rate (100%), the average successful identification level for *stx1* and *stx2* were 97% and 95%, respectively. While the participation rate was slightly higher than for the previous EQA, the success rates are in line with the results from EQA-12. For *stx1* detection, a total of nine errors were reported by three participants, while a total of 13 errors were reported from six participants for the *stx2* detection. There was an overlap between these groups, as participants reporting erroneous results for *stx1*, also did so for *stx2*.

4.2.4 Subtyping of *stx1* and *stx2*

Subtyping of Shiga toxin genes was performed by 86% of the EQA participants. For *stx1*, the success rate rounds up to 100%, as only one incorrect typing was reported for a total of 228 reported answers. For *stx2* subtyping, a total of six errors were reported, from four participants, resulting in a success rate of 97%. When combining subtyping results for *stx1* and *stx2*, 74% of laboratories participating in the subtyping successfully subtyped both genes in all strains. The average success rate for the combined subtyping was 97%, and as such the overall subtyping performance was slightly higher than that of EQA-12.

In the EQA-12 report, a notable increase in *stx2* subtyping was noted. For the current EQA, the level of reported errors appears to have improved. It should be noted that two of three laboratories that reported the majority of subtyping errors in EQA-12 did not participate in the subtyping section of EQA-13.

Since the establishment of the currently accepted Stx subtype taxonomy in 2012, six additional Stx subtypes have been proposed, Stx1e, Stx2h, Stx2i, Stx2k, Stx2l, Stx2m, Stx2n and Stx2o [18], some of which have already been discussed by the EFSA BIOHAZ Panel in the EFSA report [7]. The EQA provider is currently collaborating with the developers of the NCBI tool StxTyper to implement novel types into this bioinformatic tool.

4.3 Molecular typing-based cluster analysis

Nineteen of the 22 laboratories (86%) participating in EQA-13 performed cluster analyses. They all used WGS-derived data, not PFGE-derived data for their analyses. This is in line with the participation level (88%) and practices reported in EQA-12.

4.3.1 WGS-derived data

All participating laboratories (19) reported sequencing the supplied strains in-house, using commercial kits for library preparation. All but two laboratories reported using an Illumina sequencing platform (89%, 17/19), with NextSeq being the most widely used (53%, 10/19). Two laboratories reported using an Ion Torrent platform.

As part of the evaluation of WGS-derived data, EQA participants were asked to submit their raw sequencing reads for QC analyses using the Bifrost QC pipeline of the EQA provider as a common reference check. The submitted sequence data were generally of good quality. However, for all but one participating laboratory (95%), one or more warnings were noted by the QC tool. This is considerably higher than for EQA-12 (39%). Upon closer inspection, the vast majority of warnings were issued based on exceeding the maximum genome size threshold in Bifrost, followed, to a lesser degree, by indications of potential contamination. As warnings related to genome size were issued for 89% of participants, it appears that the QC thresholds for genome size are perhaps too restrictive in Bifrost. If increased by just 0.1MBp, most warnings would not have been issued. As such, the number of warnings issued is misleadingly high. For 10 of the 19 participants, warnings were issued for sequences, based on potential contamination. Issued when the combined percentages of primary detected species unclassified reads make up <95% of the total reads, this indicates that a secondary, identifiable species has likely contaminated the sequenced sample. Finally, sequences from three participants were issued due to sequencing coverages <50X. For two of these, the lowest reported coverages were 44X and 43X for which you can argue that a lower threshold of 50X may be restrictive, and that these coverages are sufficient for the analyses performed in this EQA. For the third participant, however, coverages of <30 were calculated for half of the strains (6/12), which is regarded as too low by the EQA provider.

As in previous years, the primary quality control parameters reported by participants in EQA-13 included a coverage threshold and verification of genus/species confirmation. Since EQA-9, the proportion of participants assessing genome size has remained above 71%, while the use of genus confirmation as a QC parameter has exceeded 91%. For two participating laboratories, the supplied QC parameters were inconclusive, as one reported tool was used to assess genome size rather than the requested size threshold, and one reporting using a Phred-score threshold of >98, which is unlikely, and could be a wrongly entered coverage threshold.

The performance of the cluster analysis was excellent, with all 19 laboratories (100%) correctly identifying the cluster of closely related strains, which is higher compared to EQA-12 (65%) and EQA-11 (80%). Even with the inclusion of the manipulated sequences, strain15 and strain20, all participants correctly identified the cluster of four closely related strains. However, approximately 1/3 of the participants did not identify the contamination in strain15.

Of the 19 laboratories, 17 (89%) reported using an allele-based method as the main analysis, and two (11%) reported using SNP analysis as their main method for cluster detection. Unlike in EQA-12, both participants using a SNP-based cluster detection method managed to correctly identify the cluster. When assessing the reported allele difference or SNP distances, the cgMLST approach showed very comparable results and, for all participants, a clear separation of the cluster and non-cluster strains.

There was a high level of uniformity among the reported allele differences reported by the laboratories using the SKESA assembler. These laboratories all reported 0-0 AD between strains in the cluster. SPAdes was the most popular assembler but the reported allele differences varied between 0-0 and 0-7 AD for the reported cluster. The laboratory (222) reporting 0-7 AD for strains in the cluster was the only laboratory using the PHANtAsTiC pipeline and 2360 loci in their analysis. This might explain the bigger difference in allele differences in the reported cluster compared to other laboratories that used SeqSphere, Bionumerics, chewBBACA or Enterobase in their analysis. The only two laboratories (222 and 88) reporting using the Innuendo scheme where the ones with the highest variation in allele differences in the cluster analysis; 0-4 and 0-7 AD. Laboratory 222 which had the biggest deviation from the allele differences by the EQA providers submitted Ion Torrent data. Therefore, the observed allelic differences (AD) may be artefacts of the method; however, the use of Ion Torrent data can complicate communication and investigation of multi-country outbreaks when relying solely on the allelic method. On a very positive note, both laboratories submitting Ion Torrent data identified the correct cluster in this EQA. This has proven to be a challenge in the previous EQAs.

From the additional analyses reported by other participants, the distances reported inside the cluster using SNP-based analyses (and identifying the correct cluster) were 0–1 (Reference-based method) or 0–1 (Assembly and mapping-based method, only one participant), thus showing no variation between the methods.

Two laboratories (138 and 187) have recently started using WGS-derived data, and EQAs are a good way to test the progress of this transition. Laboratory 138 provided good-quality data; however, the O-typing and the gene analyses posed some challenges. This emphasises the importance of understanding the pipeline and carefully evaluating the data. From the data visualised in Figure 9/10, there is a less clear separation between the cluster strains and the remaining strains for Laboratory 138, which successfully identified the cluster. Meanwhile, Laboratory 187 provided quality data with an average coverage between 51-198. However, they also faced some challenges in the gene analysis but correctly identified the cluster.

The submitted raw data indicate that when applying a standardised cgMLST analysis, minor random variations of a single allele are not uncommon, even with high sequencing coverage (Figure 12). In previous EQAs, this phenomenon has been particularly evident in laboratories submitting Ion Torrent data. However, unlike in EQA-12, both laboratories (88 and 222) that submitted Ion Torrent data in EQA-13 correctly identified the cluster. The EQA provider's analysis is not optimised for Ion Torrent, which can make accurate assembly challenging. Nevertheless, despite minor deviations in allelic differences, these discrepancies did not affect the overall analytical performance in this instance. However, reliance on Ion Torrent data can complicate communication and the investigation of multi-country outbreaks when using the allelic method exclusively.

In this EQA, the EQA provider included two modified strains, strain15 and strain20. Both were non-cluster sequences; strain15 was contaminated with approximately 9.3% *E. albertii* and strain20 had a low coverage. Both sequences were marked as QC-status C ('not acceptable quality') or B/C Quality only acceptable for outbreak situations (less good quality) by the EQA provider.

The contamination in strain15 was challenging to detect; only 32% of participants classified the strain as QC-status C ('not acceptable quality'), 32% as QC-status B ('less good quality'), and 37% as QC-status A ('acceptable quality'). In the assessment of the modified strains (Annex 11), only seven participants indicated that strain15 was either contaminated or likely contaminated. One of these laboratories identified contamination with *E. albertii* but removed the affected sequences using the INNUca pipeline before proceeding with further analysis of the strain. The majority of laboratories did not appear to detect the contamination and did not mention a possible contamination in their assessment of strain15. This result is consistent with findings from EQA-12, where one of the modified strains was contaminated with approximately 8% *S. sonnei*. In that instance, only 48% (11/23) of participants classified the strain as either QC-status B or C, while the remaining 52% (12/23) reported the sequence as being of acceptable quality (QC-status A). In contrast, in EQA-11, where 85% (17/20) of participants correctly identified contamination in a non-cluster sequence containing approximately 14% *E. albertii*, the higher contamination level may have made detection more straightforward. These findings suggest that laboratories may find it easier to identify contamination at 14% *E. albertii* than at lower levels, such as 9.3% *E. albertii* or 8% *S. sonnei*.

Nearly all of the participants (95%) correctly reported quality issues of strain20, (a non-cluster sequence with reduced coverage). Only one participant reported strain20 as A, 'acceptable quality', and used the sequence for analysis. All participants, except the laboratory that reported the strain as A, noted that low coverage in the assessment of strain20 (Annex 11).

The results of EQA-13 demonstrate continued high performance in WGS-based cluster detection. Despite variations in sequencing platforms and analytical pipelines, allele- and SNP-based methods produced highly comparable results. While minor discrepancies in allelic differences were observed, particularly among laboratories using Ion Torrent data, these did not impact the overall analytical performance. However, challenges remain in quality control, particularly regarding contamination detection, as seen with strain15.

The significantly higher number of QC warnings in this EQA – primarily due to genome size thresholds – suggests the need for a reassessment of threshold settings to ensure more meaningful reporting. Encouragingly, new participants transitioning to WGS demonstrated promising results, underscoring the value of EQAs in supporting methodological refinement. Moving forward, improving contamination detection and refining QC parameters will be critical to enhancing the robustness and comparability of WGS-based surveillance and outbreak investigations.

5 Conclusions

Twenty-two laboratories participated in the EQA-13 scheme, with 21 (95%) performing the serotyping part, 22 (100%) determining the virulence profile, and 19 (86%) engaging in cluster identification. Participation in the serotyping, virulence gene profile and cluster analysis was similar to EQA-12 (21/22, 95%, 25/26, 96% and 25/26, 96%). Similar to EQA-11 and EQA-12, this EQA incorporated cluster analysis based on molecular typing, utilising exclusively WGS-derived data since no participants submitted PFGE data this year. The last instance of PFGE reporting was in EQA-10, indicating a permanent shift in STEC 'finger-printing' from PFGE to WGS among Member States.

O:H serotyping was performed by 88% (21/22) of the participants, achieving an average score of 91%. Similar to previous EQAs, participation in O grouping exceeded that in H typing. Consistent with prior EQAs, not all laboratories exhibited the ability to determine all O groups. However, in this EQA-12 all laboratories correctly identified all H types. Generally, the more prevalent European serotypes generated higher scores compared to the less common ones, such as O187:H28 which posed greater challenges in identification, especially when participants utilised phenotypic methods. Further, it was clear that serotype O159 proved highly difficult, both in terms of conventional and WGS-based serotyping.

Once again, this year, the EQA provider included two other DEC pathotypes, EAEC (*aggR* gene), and ETEC (*esta* gene), testing the participating laboratories on their ability to detect STEC hybrid strains. The performance in detecting the aggR genes was relatively high (20/22, 90%), but lower then than the average score in EQA-12 where 98% correctly identified *aggR*. The average performance score for correctly identifying the *esta* gene was lower (80%) than in EQA-12 (98%). This performance discrepancy was attributed to three laboratories (108, 125 and 138) that couldn't identify the *esta* gene in strain1. All laboratories except three utilised a WGS-based method for the virulence gene profiling.

Detection of the eae gene had high participation rates, and average scores through the EQAs has always been 96% or above (EQA-4: 96%; EQA-5: 98%; EQA-6: 97%; EQA-7: 98%; EQA-8: 96%; EQA-9: 99%, EQA-10: 98%, and EQA-11: 97%; and EQA-12: 96%). However, in EQA-13 the average score was 86%, with 19/22 laboratories correctly identifying the gene in all strains.

Similarly, to previous EQAs, the participation in *stx1* and *stx2* gene detection and average scores for correctly identifying the genes were high, with an average score of 100% for *stx1* and 97% for *stx2.* Subtyping of *stx1* and *stx2* is valuable since specific subtypes (*stx*2a) have been associated with increased risk of HUS, hospitalisation, or bloody diarrhoea respectively [8]. The high participation rate of 100% (22/22) in the *stx1* and *stx2* detection is encouraging. The average score of laboratories that correctly performed the *stx* subtyping were 100% for *stx1*, 97% for *stx2*, and 97% combined *stx1* and *stx2*.

The incorporation of molecular typing-based cluster analysis in this EQA is up-to-date with the development of surveillance methods used by NPHRLs in Europe. Nineteen laboratories performed the cluster analysis, which is four less than EQA-12, and all 19 used WGS-derived data. Notably, no laboratory employed PFGE for cluster analysis while participating in this EQA.

Modifying genomes have been the practice by the EQA provider since EQA-10. As such, the strain sequence data were made accessible by the EQA provider, and participants were instructed to incorporate them into the cluster analysis while reporting characteristics and quality issues. It should be noted that contaminations with a different species can be more challenging to identify than low-quality sequences. Unlike EQA-11, where most participants identified the contamination (quality issue), in EQA-13, 63% of participants identified quality issues with strain15 but only 37% identified the contamination. In EQA-12, 48% of the participants reported issues with the quality of the contaminated strain. The cluster analysis performance was high, with all (100%) of laboratories correctly identifying the cluster of closely related strains. In addition, one laboratory overlooked the very low coverage of strain20. All in all, however, the results are encouraging.

Furthermore, 18 laboratories (18/19, 95%) reported using an allele-based method as the main analysis and one (5%) reported using SNP analysis. The use of a standard cgMLST scheme (e.g. Enterobase) gives a very high degree of homogeneity in the results, and allele-based methods seem to be useful for inter-laboratory comparability and communication about cluster definitions. SNP analyses can also provide valid cluster detection at the national level; however, the analysis pipeline needs to be carefully assessed.

The current EQA scheme for typing STEC is the 13th EQA organised for laboratories in FWD-Net. The molecular surveillance system implemented as part of EpiPulse Cases relies on the capacity of FWD-Net laboratories to produce analysable and comparable typing results into a central database. WGS-based typing for surveillance is increasingly used in the EU. Member States are asked to submit STEC WGS data in real-time to be accompanied by isolate metadata. ECDC coordinates centralised analysis of WGS STEC data when needed to support multi-country outbreak investigations.

6 Recommendations

6.1 Laboratories

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

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

6.2 ECDC and FWD-Net

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

6.3 EQA provider

The assessment of the provided genome sequences yielded positive results, with almost all participants successfully identifying the modifications introduced by the EQA provider, particular for strain20 with low coverage. The exception was the contamination with % *E. albertii* in strain15 where 63% of the participants identified the quality issue. Consequently, in subsequent EQA rounds, any EQA provider should increase the contamination load, following the approach employed in previous EQAs (e.g. introducing 14% contamination with *E. albertii* in EQA-11).

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

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

References

1. EFSA and ECDC (European Food Safety Authority and European Centre for Disease Prevention and Control), (2024). The European Union One Health 2023 Zoonoses report. EFSA Journal, 22(12), e9106. https://doi.org/10.2903/j.efsa.2024.9106
2. European Parliament and European Council. Regulation (EC) 2022/2370 amending Regulation (EC) No 851/2004 of the European Parliament and of the Council of 23November 2022 establishing a European centre for disease prevention and control – Article 5.8. Strasbourg: European Parliament and European Council; 2022. Available at: https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:32022R2370
3. European Parliament and European Council. Decision No 1082/2013/EU of the European Parliament and of the Council of 23 November 2022 on serious cross-border threats to health and repealing Decision No 1082/2013/EC (Text with EEA relevance). Strasbourg: European Parliament and European Council; 2013. Available at: http://publications.europa.eu/en/publication-detail/-/publication/8d817a1f-45fa-11e3-ae03-01aa75ed71a1
4. Bielaszewska M, Friedrich AW, Aldick T, Schürk-Bulgrin R, Karch H. Shiga Toxin Activatable by Intestinal Mucus in *Escherichia coli* Isolated from Humans: Predictor for a Severe Clinical Outcome. Clin Infect Dis. 2006 Nov 1;43(9):1160-7.
5. Friedrich AW, Bielaszewska M, Zhang WL, Pulz M, Kuczius T, Ammon A, et al. *Escherichia coli* Harboring Shiga Toxin 2 Gene Variants: Frequency and Association with Clinical Symptoms. J Infect Dis. 2002 Jan 1;185(1):74-84.
6. Persson S, Olsen KE, Ethelberg S, Scheutz F. Subtyping Method for *Escherichia coli* Shiga Toxin (Verocytotoxin) 2 Variants and Correlations to Clinical Manifestations. J Clin Microbiol. 2007 Jun;45(6):2020-4.
7. EFSA BIOHAZ Panel, Koutsoumanis K, Allende A, Alvarez-Ordóñez A, Bover-Cid S, Chemaly M, et al. Pathogenicity assessment of Shiga toxin-producing *Escherichia coli* (STEC) and the public health risk posed by contamination of food with STEC. EFSA Journal. 2020;18(1):5967, 105 pp. Available at: https://doi.org/10.2903/j.efsa.2020.5967
8. Scheutz F, Teel LD, Beutin L, Piérard D, Buvens G, Karch H, et al. Multicenter Evaluation of a Sequence-Based Protocol for Subtyping Shiga Toxins and Standardizing Stx Nomenclature. J Clin Microbiol. 2012 Sep;50(9):2951-63.
9. Boisen N, Østerlund M T, Joensen K G, Santiago A E, Mandomando I, Cravioto A el at. 2020, Redefining enteroaggregative *Escherichia coli* (EAEC): Genomic characterization of epidemiological EAEC strains, PloS Negl Trop Dis. 2020 Sep 8;14(9):e0008613. Doi: 10.1371/journal.pntd.0008613. eCollection 2020 Sep.
10. International Organization for Standardization (ISO). ISO/IEC 17043:2010 – Conformity assessment – General requirements for proficiency testing. Vernier: ISO; 2010. Available at: http://www.iso.org/iso/catalogue_detail.htm?csnumber=29366
11. Warwick Medical School. EnteroBase. Coventry: University of Warwick; 2018. Available at : http://enterobase.warwick.ac.uk
12. Sahl JW, Lemmer D, Travis J, Schupp JM, Gillece JD, Aziz M, et al. NASP: an accurate, rapid method for the identification of SNPs in WGS datasets that supports flexible input and output formats. Microb Genom. 2016 Aug 25;2(8):e000074.
13. Scheutz F, Fruth A, Cheasty T, Tschäpe H. Appendix 1 – O Grouping: Standard Operation Procedure (O SOP) and Appendix 2: and H Determination: Standard Operation Procedure (H SOP) – *Escherichia coli* O antigen grouping and H antigen determination. Copenhagen: Statens Serum Institut; 2002. Available at: http://www.ssi.dk/English/HealthdataandICT/National%20Reference%20Laboratories/Bacteria/~/media/49802860CB5E44D6A373E6116ABBDC0D.ashx
14. Scheutz F, Morabito S, Tozzoli R, Caprioli A. Identification of three vtx1 and seven vtx2 subtypes of verocytotoxin encoding genes of *Escherichia coli* by conventional PCR amplification. Copenhagen: Statens Serum Institut; 2002.
15. European Centre for Disease Prevention and Control (ECDC). Surveillance of National Reference Laboratory (NRL) capacity for six food- and waterborne diseases in EU/EEA countries – Campylobacteriosis, listeriosis, salmonellosis, Shiga toxin/ verocytotoxin–producing *Escherichia coli* (STEC/VTEC), shigellosis and yersiniosis. Stockholm: ECDC; 2012. Available at: http://ecdc.europa.eu/publications-data/survey-national-reference-laboratory-capacity-six-fwd-eueea-countries
16. Statens Serum Institut (SSI). Bifrost_QC [Internet; software package]. Copenhagen: SSI; 2019. Available at: https://github.com/ssi-dk/bifrost

Annex 1. List of participants

Table 12. Laboratories participating in EQA-13

Country	Laboratory	National institute
Austria	Reference Center for Escherichia coli including VTEC	Institute of Medical Microbiology and Hygiene Graz
Belgium	National Reference Centre STEC	Universitair Ziekenhuis Brussel
Bulgaria	NRL for Enteric Diseases	National Center of Infectious and Parasitic Disease
Germany	NRC Salmonella and other bacterial enterics	Robert Koch Institute
Denmark	Laboratory of Gastrointestinal Bacteria	Statens Serum Institut
Estonia	Laboratory of Communicable Diseases	Health Board
Spain	Laboratorio de Referencia e Investigación en Enfermedades Transmitidas por Agua y Alimentos	Instituto de Salud Carlos III
France	National Reference Center, Escherichia coli, Shigella, Salmonella	Institut Pasteur - CHU Robert Debré - APHP
Greece	Reference Centre for Salmonella, Shigella, Listeria, VTEC	University of West Attica
Croatia	National Reference Center for Salmonella and Dpt for intestinal pathogen diagnostics	Croatian Institute of Public Health
Hungary	FWD - Laboratory	National Center for Public Health and Pharmacy
Ireland	Public Health Laboratory Dublin	Health Service Executive
Iceland	Department of Clinical Microbiology	Landspitali - The University Hospital of Iceland
Italy	Microbiological Food Safety and Foodborne Disease Unit	Istituto Superiore di Sanità
Luxembourg	Pathogen Sequencing	Laboratoire National de Santé
Latvia	Laboratory Service, National Microbiology Reference Laboratory	Riga East University Hospital
The Netherlands	IDS - BVI	National Institute for Public Health and the Environment
Norway	National Reference Laboratory for Enteropathogenic Bacteria	Norwegian Institute of Public Health
Portugal	URGI	Instituto Nacional de Saúde Doutor Ricardo Jorge
Romania	Molecular Epidemiology for Communicable Diseases	Cantacuzino National Military Medical Institute for Research and Development
Sweden	Unit for Laboratory Surveillance of Bacterial Pathogens	Public Health Agency of Sweden
Slovenia	Department for Public Health Microbiology	National Laboratory for Health, Environment and Food

Annex 2. Participation overview EQA-12/-13

Table 13. Participation overview EQA-12/-13

	2022-2023 (EQA-12)				2023-2024 (EQA-13)			
Laboratory number	Participation (min. 1 part)	Serotyping	Virulence	Cluster WGS	Participation (min. 1 part)	Serotyping	Virulence	Cluster WGS
19	x	x	x	x	x	x	x	x
34	x	x	x	x	x	x	x	x
80	x	x	x	x	x	x	x	x
88	x	x	x	x	x	x	x	x
90	x		x	x	x	x	x	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
125#					x	x	x	
127	x	x	x	x				
128	x	x	x	x	x	x	x	
129	x	x	x	x				
130	x	x	x		x		x	
131	x	x	x	x	x	x	x	x
132	x	x	x	x	x	x	x	x
133	x	x	x	x	x	x	x	x
134	x	x	x	x	x	x	x	x
135	x	x	x	x	x	x	x	x
136	x	x	x	x	x	x	x	x
138	x	x	x	x	x	x	x	x
139	x	x	x	x	x	x	x	x
145*								
153	x	x	x	x				
187	x	x	x	x	x	x	x	x
222	x	x	x	x	x	x	x	x
230	x	x	x					
240	x	x						
Number of participants	26	25	25	23	22	21	22	19

** = Laboratory did not participate in EQA-12 or EQA-13*

 # = Laboratory did not participate in EQA-12

Annex 3. Serotyping result scores

Table 14. Results for O group typing

	Strain number												
Laboratory number	1	2	3	4	5	6	7	8	9	10	11	12	
EQA	O187	O157	O171	O146	O55	O27	O55	O26	O145	O159	O55	O91	Method
19	187	157	171	146	55	27	55	26	145	159	55	91	A
34	187	157	171	146	55	27	55	26	145	104	55	91	B
80	187	157	171	146	55	27	55	26	145	NT	55	91	B
88	187	157	171	146	55	27	55	26	145	92	55	91	B
90	187	157	171	146	55	27	55	26	145	NT	55	91	B
100	187	157	171	146	55	27	55	26	145	NT	55	91	A
108	187	157	171	146	55	27	55	26	145	NT	55	91	B
123	187	157	171	146	55	27	55	26	145	NT	55	91	B
124	187	157	171	146	55	27	55	26	145	NT	55	91	B
125	103	157	NT	NT	55	NT	55	26	145	NT	55	91	A
128	187	157	171	146	55	27	55	26	145	NT	55	91	A
131	NT	157	171	146	55	27	55	26	146	104	55	91	A
132	187	157	171	146	55	27	55	26	145	NT	55	91	B
133	187	157	171	146	55	27	55	26	145	NT	55	91	B
134	187	157	171	146	55	27	55	26	145	NT	55	91	B
135	187	157	171	146	55	27	55	26	145	NT	55	91	B
136	187	157	171	146	55	27	55	26	145	NT	55	91	B
138	NT	157	NT	NT	55	NT	55	26	145	139	55	NT	A
139	187	157	171	146	55	27	55	26	145	NT	55	91	B
187	187	157	171	146	55	27	55	26	145	NT	55	91	B
222	187	157	171	146	55	27	55	26	145	NT	55	91	B

n=21 participants
Purple shading: incorrect result
A: phenotypic serotyping, B: WGS-based serotyping
NT: non-typable

Table 15. Results for H typing

	Strain number												
Laboratory number	1	2	3	4	5	6	7	8	9	10	11	12	
EQA	H28	H-/H7	H-/H2	H21	H7	H30	H7	H-/H11	H-/H28	H4	H7	H14	Method
19	28	H-	H-	21	7	30	7	H-	H-	4	7	14	A
34	28	7	2	21	7	30	7	11	28	4	7	14	B
80	28	7	2	21	7	30	7	11	28	4	7	14	B
88	28	7	2	21	7	30	7	11	28	4	7	14	B
90	28	7	2	21	7	30	7	11	28	4	7	14	B
100	28	7	2	21	7	30	7	11	28	4	7	14	B
108	28	7	2	21	7	30	7	11	28	4	7	14	B
123	28	7	2	21	7	30	7	11	28	4	7	14	B
124	28	7	2	21	7	30	7	11	28	4	7	14	B
131	28	7	2	21	7	30	7	11	28	4	7	14	B
132	28	7	2	21	7	30	7	11	28	4	7	14	B
133	28	7	2	21	7	30	7	11	28	4	7	14	B
134	28	7	2	21	7	30	7	11	28	4	7	14	B
135	28	7	2	21	7	30	7	11	28	4	7	14	B
136	28	7	2	21	7	30	7	11	28	4	7	14	B
139	28	7	2	21	7	30	7	11	28	4	7	14	B
187	28	7	2	21	7	30	7	11	28	4	7	14	B
222	28	7	2	21	7	30	7	11	28	4	7	14	B

n=18 participants

Purple shading: incorrect result

A: phenotypic serotyping, B: WGS-based serotyping

Some H- results was accepted as correct results (Strain2, Strain3, Strain8, Strain9), when the EQA provider observed a tendency to be H- more than one during testing.

Annex 4. Virulence profiles result scores

Table 16. **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	-	-	-	-	-	-	-	-	-	+	-	-	B
90	-	-	-	-	-	-	-	-	-	+	-	-	A
100	-	-	-	-	-	-	-	-	-	+	-	-	A
108	-	-	-	-	-	-	-	-	-	+	-	-	B
123	-	-	-	-	-	-	-	-	-	+	-	-	B
124	-	-	-	-	-	-	-	-	-	+	-	-	B
125	-	-	-	-	-	-	-	-	-	-	-	-	A
128	-	-	-	-	-	-	-	-	-	+	-	-	A
130	-	-	-	-	-	-	-	-	-	+	-	-	A
131	-	-	-	-	-	-	-	-	-	+	-	-	A
132	-	-	-	-	-	-	-	-	-	+	-	-	B
133	-	-	-	-	-	-	-	-	-	+	-	-	B
134	-	-	-	-	-	-	-	-	-	+	-	-	B
135	-	-	-	-	-	-	-	-	-	-	-	-	B
136	-	-	-	-	-	-	-	-	-	+	-	-	B
138	-	-	-	-	-	-	-	-	-	+	-	-	A
139	-	-	-	-	-	-	-	-	-	+	-	-	B
187	-	-	-	-	-	-	-	-	-	+	-	-	B
222	-	-	-	-	-	-	-	-	-	+	-	-	B

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

Table 17. **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	-	+	-	-	+	-	+	+	+	-	+	-	B
90	-	+	-	-	+	-	+	+	+	-	+	-	A
100	-	+	-	-	+	-	+	+	+	-	+	-	A
108	-	+	-	-	+	-	+	+	+	-	+	-	B
123	-	+	-	-	+	-	+	+	+	-	+	-	B
124	-	+	-	-	+	-	+	+	+	-	+	-	B
125	-	+	-	-	-	-	-	-	+	-	+	-	A
128	-	+	-	-	+	-	+	+	+	-	+	-	A
130	-	+	-	-	+	+	+	-	+	-	+	-	A
131	-	+	-	-	+	-	+	+	+	-	+	-	A
132	-	+	-	-	+	-	+	-	+	-	+	-	B
133	-	+	-	-	+	-	+	+	+	-	+	-	B
134	-	+	-	-	+	-	+	+	+	-	+	-	B
135	-	+	-	-	+	-	+	+	+	-	+	-	B
136	-	+	-	-	+	-	+	+	+	-	+	-	B
138	-	+	-	-	+	-	+	+	+	-	+	-	A
139	-	+	-	-	+	-	+	+	+	-	+	-	B
187	-	+	-	-	+	-	+	+	+	-	+	-	B
222	-	+	-	-	+	-	+	+	+	-	+	-	B

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

Table 18. 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	+	-	-	-	-	-	-	-	-	-	-	-	B
90	+	-	-	-	-	-	-	-	-	-	-	-	A
100	+	-	-	-	-	-	-	-	-	-	-	-	A
108	-	-	-	-	-	-	-	-	-	-	-	-	B
123	+	-	-	-	-	-	-	-	-	-	-	-	B
124	+	-	-	-	-	-	-	-	-	-	-	-	B
125	-	-	-	-	-	-	-	-	-	-	-	-	A
131	+	-	-	-	-	-	-	-	-	+	-	-	A
132	+	-	-	-	-	-	-	-	-	-	-	-	B
133	+	-	-	-	-	-	-	-	-	-	-	-	B
134	+	-	-	-	-	-	-	-	-	-	-	-	B
135	+	-	-	-	-	-	-	-	-	-	-	-	B
136	+	-	-	-	-	-	-	-	-	-	-	-	B
138	-	-	-	-	-	-	-	-	-	-	-	-	A
139	+	-	-	-	-	-	-	-	-	-	-	-	B
187	+	-	-	-	-	-	-	-	-	-	-	-	B
222	+	-	-	-	-	-	-	-	-	-	-	-	B

n=20 participants *A: Other than WGS, B: WGS-based*
Purple shading: incorrect result

Table 19. 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	-	-	-	+	-	-	+	+	+	-	+	+	B
90	-	-	-	+	-	-	+	+	+	-	+	+	A
100	-	-	-	+	-	-	+	+	+	-	+	+	A
108	-	-	-	+	-	-	+	+	+	-	+	+	B
123	-	-	-	+	-	-	+	+	+	-	+	+	B
124	-	-	-	+	-	-	+	+	+	-	+	+	B
125	-	-	-	-	-	-	+	-	+	-	+	+	A
128	+	+	+	+	-	+	+	+	+	+	+	+	A
130	-	-	-	+	-	-	+	+	+	-	+	+	A
131	-	-	-	+	-	-	+	+	+	-	+	+	A
132	-	-	-	+	-	-	+	+	+	-	+	+	B
133	-	-	-	+	-	-	+	+	+	-	+	+	B
134	-	-	-	+	-	-	+	+	+	-	+	+	B
135	-	-	-	+	-	-	+	+	+	-	+	+	B
136	-	-	-	+	-	-	+	+	+	-	+	+	B
138	-	-	-	+	-	-	+	-	+	-	+	-	A
139	-	-	-	+	-	-	+	+	+	-	+	+	B
187	-	-	-	+	-	-	+	+	+	-	+	+	B
222	-	-	-	+	-	-	+	+	+	-	+	+	B

n=22 participants *A: Other than WGS*
Purple shading: incorrect result *B: WGS-based*

Table 20. 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	+	+	+	+	-	+	-	+	+	+	-	+	B
90	+	+	+	+	-	+	-	+	+	+	-	+	A
100	+	+	+	+	-	+	-	+	+	+	-	+	A
108	+	+	+	+	-	+	-	+	-	+	-	+	B
123	+	+	+	+	-	+	-	+	+	+	-	+	B
124	+	+	+	+	-	+	-	+	+	+	-	+	B
125	+	+	-	-	-	-	-	-	+	+	-	-	A
128	+	+	+	+	-	+	+	+	+	+	+	+	A
130	+	+	+	+	-	+	-	+	+	+	-	+	A
131	+	+	+	+	-	+	-	+	+	+	-	+	A
132	+	+	+	+	-	-	-	+	+	+	-	+	B
133	+	+	+	+	-	+	-	+	+	+	-	+	B
134	+	+	+	+	-	+	-	+	+	+	-	+	B
135	+	+	+	+	-	+	-	+	-	+	-	+	B
136	+	+	+	+	-	+	-	+	+	+	-	+	B
138	-	-	-	+	-	+	-	+	+	+	-	+	A
139	+	+	+	+	-	+	-	+	+	+	-	+	B
187	+	+	+	+	-	+	-	+	+	+	-	+	B
222	+	+	+	+	-	+	-	+	+	+	-	+	B

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

stx subtyping

Table 21. stx1

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

n=19 participants
Purple shading: incorrect result
-: negative for stx2
A: Other than WGS
B: WGS-based

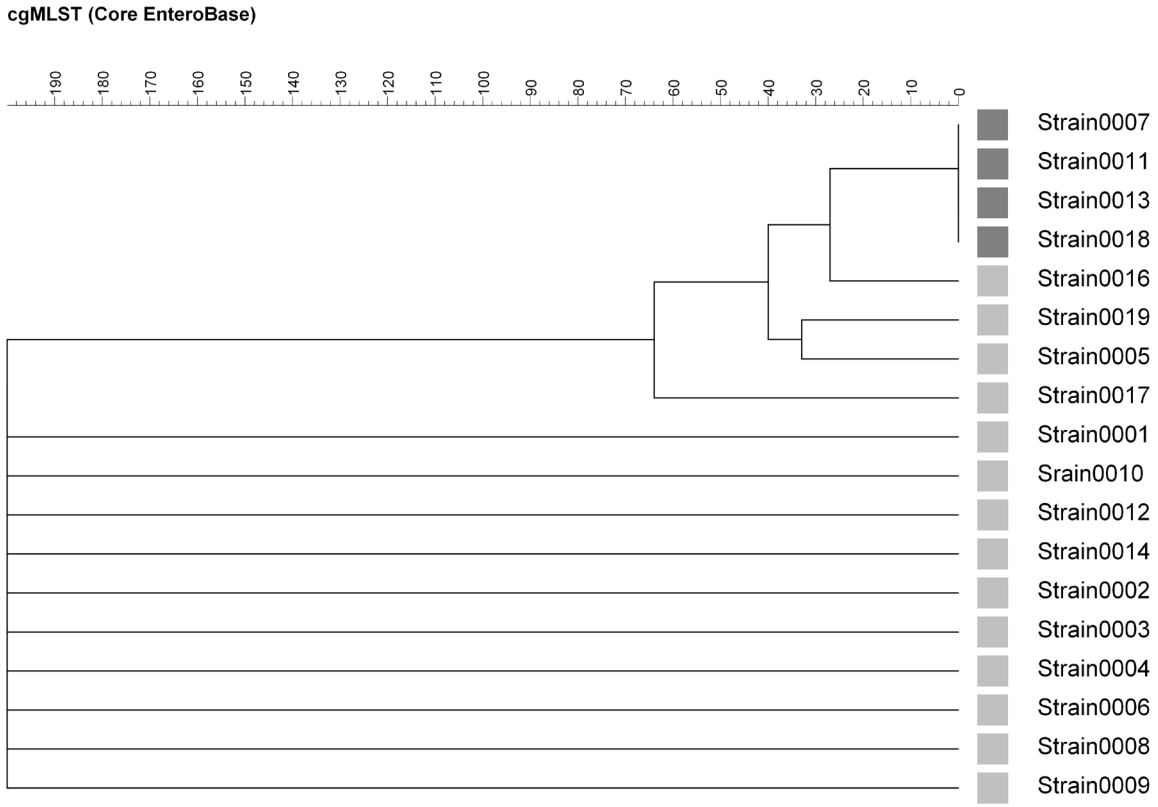
Table 22. stx2

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

n=19 participants *A: Other than WGS, B: WGS-based*
Purple shading: incorrect result *-: negative for stx2*

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

Figure 13. EQA provider's cluster analysis



Single linked dendrogram of core genome multilocus sequence typing (cgMLST) profiles of STEC EQA-13 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.
Strain7, strain11 and strain18 are technical duplicates.

Annex 6. Reported sequencing details

Table 23. 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	NextSeq
80	In own laboratory	Commercial kits	Xgen DNA libr prep EZ from Integrated DNA Technologies	NextSeq
88	In own laboratory	Commercial kits	Nextera XT DNA Library Preparation kit (Illumina)	NextSeq
90	In own laboratory	Commercial kits	Nextera XT DNA Library Preparation Kit	MiSeq
100	In own laboratory	Commercial kits	Illumina Library Prep kit	NextSeq
108	In own laboratory	Commercial kits	Ion Xpress™ Plus Fragment Library Kit for AB Library Builder™ System	Ion S5 XL System
123	In own laboratory	Commercial kits	Nextera XT Library Prep Kit (Illumina)	MiSeq
124	In own laboratory	Commercial kits	KAPA HyperPlus Kit	NovaSeq 6000
131	In own laboratory	Commercial kits	DNA Prep	NextSeq
132	In own laboratory	Commercial kits	Illumina DNA Prep	MiSeq
133	In own laboratory	Commercial kits	Illumina DNA Prep	NextSeq
134	In own laboratory	Commercial kits	DNA Prep Illumina	Mini Seq Illumina
135	In own laboratory	Commercial kits	Library preparation: Illumina DNA prep kit	NextSeq
136	In own laboratory	Commercial kits	Illumina DNA Prep, (M) Tagmentacion	NovaSeq
138	In own laboratory	Commercial kits	Illumina DNAPrep	NextSeq
139	In own laboratory	Commercial kits	Nextera XT	NextSeq
187	In own laboratory	Commercial kits	Illumina DNA prep	Mini Seq Illumina
222	In own laboratory	Commercial kits	NEBNext® Fast DNA Fragmentation & Library Prep Set for Ion Torrent, New England Biolabs	Ion GeneStudio S5 Prime System

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

Table 24. Reported cluster

Laboratory	Reported cluster	Corresponding to EQA provider strains	Correct
Provider		Strain7, Strain11, Strain13, Strain18	Yes
19	9847, 9665, 0013, 0018	Strain7, Strain11, Strain13, Strain18	Yes
34	9841, 9261, 0013, 0018	Strain7, Strain11, Strain13, Strain18	Yes
80	9171, 9973, 0013, 0018	Strain7, Strain11, Strain13, Strain18	Yes
88	9313, 9876, 0013, 0018	Strain7, Strain11, Strain13, Strain18	Yes
90	9034, 9176, 0013, 0018	Strain7, Strain11, Strain13, Strain18	Yes
100	9152, 9979, 0013, 0018	Strain7, Strain11, Strain13, Strain18	Yes
108	9036, 9809, 0013, 0018	Strain7, Strain11, Strain13, Strain18	Yes
123	9651, 9077, 0013, 0018	Strain7, Strain11, Strain13, Strain18	Yes
124	9446, 9009, 0013, 0018	Strain7, Strain11, Strain13, Strain18	Yes
131	9121, 9132, 0013, 0018	Strain7, Strain11, Strain13, Strain18	Yes
132	9981, 9224, 0013, 0018	Strain7, Strain11, Strain13, Strain18	Yes
133	9405, 9466, 0013, 0018	Strain7, Strain11, Strain13, Strain18	Yes
134	9740, 9314, 0013, 0018	Strain7, Strain11, Strain13, Strain18	Yes
135	9585, 9410, 0013, 0018	Strain7, Strain11, Strain13, Strain18	Yes
136	9195, 9962, 0013, 0018	Strain7, Strain11, Strain13, Strain18	Yes
138	9663, 9668, 0013, 0018	Strain7, Strain11, Strain13, Strain18	Yes
139	9749, 9928, 0013, 0018	Strain7, Strain11, Strain13, Strain18	Yes
187	9491, 9777, 0013, 0018	Strain7, Strain11, Strain13, Strain18	Yes
222	9377, 9143, 0013, 0018	Strain7, Strain11, Strain13, Strain18	Yes

Strain7, strain11 and strain18 are technical duplicates.

Annex 8. Reported results

Table 25. SNP distances

			Laboratory No.			
Strain ID	ST	Provider	19	132	132	108
Strain1	200	NA	NA	15	NA	80086
Strain2	11	NA	NA	18	NA	12848
Strain3	332	NA	NA	153	NA	80787
Strain4	442	NA	NA	53	NA	80410
Strain5	335	119	124	16	124	1092
Strain6	753	NA	NA	77	NA	64957
Strain7#	335	0	0	0	0	2
Strain8	21	NA	NA	13	NA	79694
Strain9	32	NA	NA	36	NA	62577
Strain10	678	NA	NA	13	NA	79899
Strain11#	335	0	0	0	0	2
Strain12	33	NA	NA	11	NA	80175
Strain13#	335	1	1	0	1	1
Strain14	11	NA	NA	168	NA	11443
Strain15	335	NA	NA	10	113	457
Strain16	335	78	81	11	86	1039
Strain17	335	178	186	16	189	1632
Strain18#	335	0	0	0	0	1632
Strain19	335	154	127	10	125	836
Strain20	335	NA	NA	5	116	NA

Table 26. Allelic differences

			Laboratory No.																	
Strain ID	ST	EQA	123	124	124	34	19	136	139	187	131	133	222	134	138	135	80	88	90	100
Strain1	200	2316	2316	2310	2340	2308	2320	6671	2341	2322	5243	200	2201	2312	2275	2313	2314	2783	2307	2315
Strain2	11	886	866	880	895	865	880	866	894	875	2045	200	832	865	852	865	865	1110	864	867
Strain3	332	2314	2309	2310	2343	2304	2300	8631	2342	2319	5234	200	2201	2308	2261	2308	2310	2785	2304	2310
Strain4	442	2326	2313	2310	2346	2309	2300	5119	2346	2324	5243	200	2196	2315	2278	2315	2316	2695	2307	2317
Strain5	335	40	37	37	42	37	36	37	43	37	94	39	53	36	36	37	37	60	37	38
Strain6	753	2186	2182	2180	2213	2176	2180	2250	2213	2194	4960	200	2081	2182	2121	2181	2181	2637	2176	2182
Strain7#	335	0	0	0	0	0	0	0	0	0	0	0	7	0	1	0	0	4	0	1
Strain8	21	233	2321	2320	2348	2315	2310	5184	2348	2329	5243	200	2207	2319	2275	2319	2321	2779	2315	2321
Strain9	32	2240	2223	2220	2260	2216	2200	4458	2263	2236	4977	200	2128	2221	2190	2222	2221	2672	2216	2223
Strain10	678	2322	2318	2310	2344	2313	2310	6746	2342	2326	5239	200	2203	2315	2280	2317	2318	2767	2313	2319
Strain11#	335	0	0	0	0	0	0	0	3	0	0	0	5	0	1	0	0	4	0	1
Strain12	33	2312	2306	2300	2335	2300	2300	3429	2336	2317	5216	200	2206	2304	2276	2304	2306	2781	2300	2306
Strain13#	335	0	0	0	1	0	0	0	1	0	0	0	0	0	2	0	0	4	0	1
Strain14	11	930	913	920	942	906	919	1130	943	924	2137	200	880	912	896	912	907	1158	906	914
Strain15	335	NA	38	36	43	37	NA	65	43	38	89	39	NA	NA	NA	45	37	59	37	NA
Strain16	335	26	26	26	30	26	19	26	30	27	71	27	32	26	29	26	26	46	26	27
Strain17	335	76	74	72	80	74	73	101	81	74	176	63	76	74	74	74	74	106	74	75
Strain18#¤	335	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0
Strain19	335	44	41	42	49	40	33	69	49	43	105	39	45	41	43	41	40	63	40	42
Strain20	335	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	272	NA	NA	NA	NA	NA

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

Annex 9. Reported QC parameters

Table 27. Reported QC parameters

Lab no.	1		2		3		4	
	Parameter	Threshold	Parameter	Threshold	Parameter	Threshold	Parameter	Threshold
19	N50	Available from QC analysis but no threshold	Number of contigs	Available from QC analysis but no threshold	Number of unidentified bases (N) or ambigiues sites	Available from QC analysis but no threshold		
34								
80	N50	>30 000	Tot. no of contigs	<1 000				
88	Inter- and intra-species contamination	INNUca (using default kraken parameters) and ConFindr (using default parameters for E. coli).	Number of contigs	INNUca v4.2.2 default parameters				
90	N50	>30 000	contamination check	<5% other species	total no. of contigs >=200 bases	<1 000 contigs		
100	N50	>50k	number of contigs	<500	contamination check with KmerFinder	most reads classified as E. coli (if more then 5% is other species, we consider the sample insufficient quality))	read length	coresponds to expected length of sequencing platform and kit
108								
123	Average Coverage	>50	assembly length	>5 000	N50	>50 000		
124	GC%	E. coli GC% +- 51%	N50	Threshold set in the quality control window of BioNumerics >52100	non-ACGT bases	Scatterplot (length vs non-ACGT)	Nr BAFPerfect	Scatterplot (length vs BAFPerfect)
131	N50	100kb<	contig count	<500				
132	Fastp filtering before assembly	We filter reads below 20 Phred before running SPAdes	Fastqc inspection	Fastqc results are inspected for anomalies, ie. GC content, size, amount of reads, overrepresentation etc.	Analysis failure on assembly	Poor assembly quality will result in insufficient coverage for MLST, AMR and cgMLST		
133	N50	ideally >70,000 but accepted if >30 000	number of contigs	ideally >=500	Nr of non ACGT	ideally <2 500		
134	Number of contigs	<700 = good; 700-1 000 = warning; > 1 000 = failed						
135	N50	>30 000	GC content	49.5 - 51.0	Number of contigs	<650	Contamination	<4%
136	N50	>30 000 bp	Total number of contigs	Less than 550				
138	N50	x > 72 925	GC%	50.3 < x ≤ 50.9	number of contigs >=0bp	x ≤ 605	rMLST_Support_%	>90% of alleles
139	General read quality control	fastq_info v2.0	Inter species contamination	Kraken2 (PlusPF-16 database), threshold 2%	N50	>20kb		

Annex 10. Calculated qualitative/quantitative parameters

Quality Assessment made by the SSI in-house quality control pipeline https://github.com/ssi-dk/bifrost [16].

Table 28. Laboratory 19

		Laboratory 19											
Parameter	Ranges*	9025	9084	9242	9269	9343	9455	9537	9665	9722	9845	9847	9968
Detected species	{Ec} or {Se} or {Sf}	Ec,	Ec, Se	Ec	Ec	Ec	Ec	Ec	Ec	Ec	Ec	Ec	Ec
% Species 1		88.8	95.7	95.4	91.5	91.5	94.7	94.6	97.0	96.9	91.6	97.0	88.6
% Species 2		5.9	0.9	1.0	3.9	1.4	0.5	0.8	0.1	0.1	0.6	0.1	2.2
% unclassified	{<100}	3.7	2.7	2.9	3.0	6.6	3.7	3.8	2.7	2.8	7.0	2.7	7.3
Length at >25X coverage (in Mbp)	{>4.64 ^ <5.56}	5.1	5.2	5.1	5.2	5.0	5.0	5.0	5.2	5.2	5.4	5.1	5.5
Length at [1,25]X coverage (in Kbp)	{<250}	23.9	61.9	51.1	47.9	16.7	16.7	13.9	49.7	12.3	40.6	14.4	43.0
Contigs at 25X coverage	{>0}	396	512	565	639	421	353	364	427	388	425	492	491
Contigs at [1,25]X coverage	{<1 000}	28	73	60	66	25	19	21	59	16	52	24	47
Average coverage	{>50}	98	98	89	108	125	114	120	77	159	90	113	99
# Reads (in thousands)		3 467	3 570	3 174	3 987	4 318	4 027	4 086	2 811	5 754	3 413	4 003	3 814
Average read length		148	147	148	148	147	145	148	146	145	147	148	147
Average insert size		330	320	336	325	311	299	346	286	291	313	340	312
N50 (in Kbp)		26	26	19	20	25	31	27	26	31	27	22	26
QC-status (Bifrost)		Warning	OK	OK	Warning	OK	OK	OK	OK	OK	OK	OK	OK

Ten strains passed the QC. The sum of "% Species 1" and "% unclassified" was below the expected threshold of 95% for strains 9025 and 9269.

Table 29. Laboratory 34

		Laboratory 34											
Parameter	Ranges*	9013	9261	9462	9619	9681	9767	9794	9840	9841	9854	9923	9957
Detected species	{Ec} or {Se} or {Sf}	Ec	Ec	Ec	Ec	Ec	Ec	Ec	Ec	Ec	Ec	Ec	Ec
% Species 1		90.1	96.1	92.9	93.7	94.7	92.7	90.7	95.2	96.4	86.5	95.9	91.6
% Species 2		3.6	0.6	1.0	0.5	0.8	1.7	0.5	0.5	0.4	1.5	0.6	0.9
Unclassified reads (%)	{<100}	4.5	3.1	4.7	4.4	3.4	3.8	8.1	3.2	3.0	9.4	3.2	6.2
Length at >25 x min. coverage (Mbp)	{>4.64 ^ <5.56}	5.4	5.3	5.1	5.1	5.3	5.2	5.6	5.4	5.3	5.6	5.3	5.1
Length [1-25] x min. coverage (kbp)	{<250}	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
No. of contigs at 25 x min. coverage	{>0}	287	130	85	109	197	123	156	198	133	243	95	117
No. of contigs [1-	{<1 000}	0	0	0	0	0	0	0	0	0	0	0	0

25] x min. coverage													
Average coverage	{>50}	178	163	217	231	230	287	277	181	166	174	153	220
No. of reads (x 1 000)		6 782	6 037	7 716	8 162	8 505	10 419	10 741	6 808	6 118	6 945	5 606	7 791
Average read length		151	151	151	151	151	151	151	151	151	151	151	151
Average insert size		307	308	307	319	320	307	311	326	312	279	330	325
N50 (kbp)		115	232	158	169	144	148	143	184	206	149	326	145
QC-status (Bifrost)		Warning	OK	OK	OK	OK	OK	Warning	OK	OK	Warning	OK	OK

Nine strains passed QC. The genome sizes of strains 9794 and 9854 exceeded the threshold in Bifrost. The sum of "% Species 1" and "% unclassified" was lower than the expected 95% for strain 9013.

Table 30. Laboratory 80

		Laboratory 80											
Parameter	Ranges*	9107	9149	9171	9175	9198	9295	9737	9770	9832	9853	9973	9999
Detected species	{Ec} or {Se} or {Sf}	Ec	Ec	Ec	Ec	Ec	Ec	Ec	Ec	Ec	Ec	Ec	Ec
% Species 1		90.1	92.2	94.7	89.2	86.1	93.4	92.4	89.7	94.3	90.2	94.8	94.4
% Species 2		2.8	0.8	0.7	0.8	1.9	0.8	0.6	1.2	0.7	2.5	0.6	0.7
Unclassified reads (%)	{<100}	4.9	5.2	4.2	9.1	9.1	4.3	5.2	7.5	4.6	5.0	4.1	3.9
Length at >25 x min. coverage (Mbp)	{>4.64 ^ <5.56}	5.4	5.1	5.3	5.6	5.6	5.3	5.1	5.1	5.3	5.2	5.3	5.4
Length [1-25] x min. coverage (kbp)	{<250}	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
No. of contigs at 25 x min. coverage	{>0}	371	120	177	220	327	263	145	152	133	165	178	256
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}	153	130	121	129	131	118	114	131	126	130	132	136
No. of reads (x 1 000)		5 965	4 725	4 590	5 107	5 287	4 551	4 097	4 720	4 738	4 835	5 012	5 203
Average read length		151	151	151	151	151	151	151	151	151	151	151	151
Average insert size		215	220	223	222	218	221	237	225	219	223	222	226
N50 (kbp)		104	140	184	136	139	141	148	135	206	135	175	119
QC-status (Bifrost)		OK	OK	OK	OK	Warning	OK	OK	OK	OK	OK	OK	OK

Eleven strains passed QC, with the genome size of strain 9198 exceeding the threshold in Bifrost.

Table 31. Laboratory 88

		Laboratory 88											
Parameter	Ranges*	9022	9277	9313	9351	9360	9530	9563	9698	9711	9751	9772	9876
Detected species	{Ec} or {Se} or {Sf}	Ec,	Ec, Se	Ec	Ec	Ec	Ec	Ec	Ec	Ec	Ec	Ec	Ec
% Species 1		89.0	97.6	97.5	87.9	88.5	92.7	96.8	94.3	95.5	96.2	90.9	97.6
% Species 2		6.1	0.1	0.1	1.6	2.4	2.3	0.6	0.8	0.5	0.7	0.4	0.1
Unclassified reads (%)	{<100}	3.4	2.0	2.2	9.8	7.2	2.6	2.0	4.2	2.9	2.3	7.9	2.1
Length at >25 x min. coverage (Mbp)	{>4.64 ^ <5.56}	5.2	5.2	5.3	5.1	5.5	5.4	5.4	5.1	5.1	5.3	5.5	5.3

Length [1-25] x min. coverage (kbp)	{<250}	13.2	26.8	20.9	9.2	43.9	14.9	21.7	2.7	19.9	21.6	15.1	21.0
No. of contigs at 25 x min. coverage	{>0}	138	140	151	149	277	350	232	85	179	247	187	163
No. of contigs [1-25] x min. coverage	{<1 000}	6	19	18	5	34	16	15	1	15	17	9	13
Average coverage	{>50}	94	76	99	121	74	116	84	91	96	78	97	86
No. of reads (x 1 000)		3 358	2 716	3 603	4 205	2 846	4 328	3 085	3 205	3 336	2 849	3 682	3 112
Average read length		149	149	149	149	149	149	149	146	149	149	149	149
Average insert size		440	436	450	432	470	415	452	386	452	454	468	442
N50 (kbp)		105	116	117	83	51	52	107	150	90	65	72	122
QC-status (Bifrost)		Warning	OK	OK	OK	Warning	OK	OK	OK	OK	OK	OK	OK

Ten strains passed the QC. Strain 9022 exhibited a high percentage in Species 2, which may indicate a contamination, and the genome size of strain 9360 exceeded the threshold in Bifrost.

Table 32. Laboratory 90

		Laboratory 90											
Parameter	**Ranges***	**9034**	**9176**	**9235**	**9246**	**9332**	**9704**	**9817**	**9911**	**9937**	**9952**	**9970**	**9991**
Detected species	{Ec} or {Se} or {Sf}	Ec	Ec	Ec	Ec	Ec	Ec	Ec	Ec	Ec	Ec	Ec	Ec
% Species 1		96.3	96.8	92.4	89.2	92.3	87.7	93.9	96.1	92.7	94.5	95.4	94.9
% Species 2		0.3	0.2	1.5	1.4	0.3	0.1	1.3	0.4	1.2	0.9	1.0	0.5
Unclassified reads (%)	{<100}	3.1	2.7	5.3	7.4	6.6	12.0	3.4	2.6	5.0	3.6	2.9	3.5
Length at >25 x min. coverage (Mbp)	{>4.64 ^ <5.56}	5.4	5.4	5.1	5.6	5.6	5.3	5.2	5.4	5.5	5.1	5.4	5.1
Length [1-25] x min. coverage (kbp)	{<250}	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	5.0	5.6	0.0
No. of contigs at 25 x min. coverage	{>0}	119	125	114	241	142	135	104	199	261	79	218	110
No. of contigs [1-25] x min. coverage	{<1 000}	0	0	0	0	0	0	0	0	0	3	5	0
Average coverage	{>50}	110	67	71	111	118	151	68	96	98	116	127	117
No. of reads (x 1 000)		2 600	1 570	1 605	3 011	2 959	5 727	1 568	2 373	2 368	2 573	3 031	2637
Average read length		233	233	233	213	229	143	235	225	238	236	230	233
Average insert size		363	351	350	278	334	170	375	312	372	375	337	352
N50 (kbp)		271	232	135	169	139	314	149	245	127	168	101	127
QC-status (Bifrost)		OK	OK	OK	Warning	Warning	OK	OK	OK	OK	OK	OK	OK

Ten strains passed QC, with the genome sizes of strains 9246 and 9332 exceeding the threshold in Bifrost.

Table 33. Laboratory 100

		Laboratory 100											
Parameter	Ranges*	9012	9152	9191	9347	9523	9561	9659	9708	9759	9864	9943	9979
Detected species	{Ec} or {Se} or {Sf}	Ec	Ec	Ec	Ec	Ec	Ec,	Ec, Se	Ec	Ec	Ec	Ec	Ec
% Species 1		99.3	99.3	98.3	97.9	97.0	86.3	95.1	91.0	98.1	92.3	94.8	98.8
% Species 2		0.0	0.0	0.7	0.2	0.9	11.4	0.1	3.3	0.6	1.0	0.4	0.0
Unclassified reads (%)	{<100}	0.6	0.6	0.7	1.3	1.6	1.7	4.3	2.1	0.9	4.8	4.3	1.1
Length at >25 x min. coverage (Mbp)	{>4.64 ^ <5.56}	5.3	5.3	5.3	5.1	5.1	5.2	5.5	5.4	5.4	5.6	5.1	5.3
Length [1-25] x min. coverage (kbp)	{<250}	0.0	0.0	26.3	0.5	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}	122	167	251	166	125	172	251	366	291	345	172	204
No. of contigs [1-25] x min. coverage	{<1 000}	0	0	13	1	0	0	0	0	0	0	0	0
Average coverage	{>50}	95	93	91	75	106	169	169	219	202	256	223	207
No. of reads (x 1 000)		1 953	1 969	1 921	1 546	2 207	3 593	3 948	4 796	4 632	6 316	4 633	5 117
Average read length		258	255	257	249	246	248	240	252	238	232	247	218
Average insert size		293	286	287	273	268	270	258	276	248	239	263	220
N50 (kbp)		313	196	141	148	165	135	114	115	143	161	114	159
QC-status (Bifrost)		OK	OK	OK	OK	OK	Warning	OK	Warning	OK	Warning	OK	OK

Nine strains passed the QC. Strain 9561 exhibited a high percentage in Species 2, which may indicate contamination. The sum of "% Species 1" and "% unclassified" was lower than the expected 95% for strain 9708. The genome size of strain 9864 exceeded the threshold in Bifrost.

Table 34. Laboratory 108

		Laboratory 108											
Parameter	Ranges*	9019	9036	9186	9327	9488	9685	9689	9701	9734	9760	9809	9898
Detected species	{Ec} or {Se} or {Sf}	Ec	Ec	Ec	Ec	Ec	Ec	Ec	Ec	Ec	Ec	Ec	Ec
% Species 1		94.0	97.8	97.8	95.1	93.0	93.7	96.0	97.2	97.1	91.4	97.7	93.6
% Species 2		0.4	0.5	0.4	1.0	1.5	2.5	1.1	0.7	0.5	1.2	0.5	2.9
Unclassified reads (%)	{<100}	4.7	1.4	1.5	2.1	4.8	2.2	2.1	1.5	1.7	5.2	1.6	2.3
Length at >25 x min. coverage (Mbp)	{>4.64 ^ <5.56}	5.5	5.3	5.2	5.1	5.0	5.2	5.0	5.2	5.4	5.6	5.3	5.3
Length [1-25] x min. coverage (kbp)	{<250}	0.0	0.0	1.2	0.0	0.0	0.0	2.4	1.8	0.0	0.0	0.0	0.9
No. of contigs at 25 x min. coverage	{>0}	512	493	1113	450	447	447	1456	1913	489	581	439	1553
No. of contigs [1-25] x min. coverage	{<1 000}	0	0	3	0	0	0	11	8	0	0	0	4
Average coverage	{>50}	112	114	98	119	119	110	104	99	112	108	112	105
No. of reads (x 1 000)		1 999	2 000	1 691	2 000	2 000	1 906	1 748	1 723	2 000	2 000	2 000	2 000

Average read length		316	309	309	309	306	307	307	315	306	310	304	293
Average insert size		0	0	0	0	0	0	0	0	0	0	0	0
N50 (kbp)		24	23	8	23	25	25	6	5	29	25	25	6
QC-status (Bifrost)		OK	OK	OK	OK	OK	OK	OK	OK	OK	Warning	OK	OK

Eleven strains passed the QC, with the genome size of strain 9760 exceeding the threshold in Bifrost. Some QC values may be unreliable due to assembly issues for Ion Torrent data (contigs, average insert size, N50).

Table 35. Laboratory 123

		Laboratory 123											
Parameter	Ranges*	9077	9170	9254	9464	9485	9570	9596	9612	9618	9648	9651	9916
Detected species	{Ec} or {Se} or {Sf}	Ec	Ec	Ec	Ec	Ec	Ec	Ec,	Ec, Se	Ec	Ec	Ec,	Ec, Sf
% Species 1		96.2	97.4	91.8	95.5	94.7	94.9	65.4	88.1	88.9	74.9	97.2	95.3
% Species 2		0.2	0.2	0.3	0.6	0.4	0.9	26.6	1.3	1.3	14.3	0.2	0.7
Unclassified reads (%)	{<100}	3.2	2.0	7.2	2.8	4.0	2.9	6.2	9.5	7.5	6.8	2.3	3.1
Length at >25 x min. coverage (Mbp)	{>4.64 ^ <5.56}	5.3	5.3	5.5	5.4	5.1	5.1	5.2	5.1	5.6	5.4	5.3	5.4
Length [1-25] x min. coverage (kbp)	{<250}	11.4	5.3	15.1	6.2	5.0	7.8	12.2	4.1	26.9	20.5	13.4	3.0
No. of contigs at 25 x min. coverage	{>0}	265	153	265	251	174	194	188	176	348	365	192	213
No. of contigs [1-25] x min. coverage	{<1 000}	14	2	13	4	7	7	12	5	30	15	10	5
Average coverage	{>50}	107	93	82	84	81	97	88	94	61	70	73	72
No. of reads (x 1 000)		2 302	1 952	1 795	1 812	1 580	1 934	1 812	1 890	1 317	1 479	1 508	1 503
Average read length		257	257	262	261	269	262	262	262	271	268	265	266
Average insert size		307	308	327	318	351	323	333	323	370	339	332	332
N50 (kbp)		60	123	52	97	58	68	68	64	40	49	103	105
QC-status (Bifrost)		OK	OK	OK	OK	OK	OK	Warning	OK	Warning	Warning	OK	OK

Nine strains passed QC. The genome size of strain 9618 exceeded the threshold in Bifrost, and strains 9596 and 9648 exhibited high percentages in Species 2, which may indicate a contamination.

Table 36. Laboratory 124

		Laboratory 124											
Parameter	Ranges*	9009	9040	9145	9298	9446	9526	9662	9778	9821	9900	9921	9972
Detected species	{Ec} or {Se} or {Sf}	Ec	Ec	Ec	Ec	Ec	Ec	Ec	Ec	Ec	Ec	Ec	Ec
% Species 1		96.0	96.4	92.2	94.6	95.0	93.3	94.4	91.7	86.8	94.3	93.2	92.8
% Species 2		0.1	0.1	0.2	0.9	0.2	1.2	0.6	0.6	1.6	1.4	0.5	1.4
Unclassified reads (%)	{<100}	3.7	3.2	6.9	3.8	4.6	4.7	4.3	6.4	9.7	3.6	5.2	5.1
Length at >25 x min. coverage (Mbp)	{>4.64 ^ <5.56}	5.4	5.3	5.6	5.4	5.3	5.1	5.4	5.2	5.6	5.4	5.1	5.1
Length [1-25] x min. coverage (kbp)	{<250}	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0

No. of contigs at 25 x min. coverage	{>0}	111	79	146	171	111	75	173	100	179	228	96	107
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}	275	280	263	273	273	290	270	152	258	268	287	290
No. of reads (x 1 000)		10 000	10 000	10 000	10 000	10 000	10 000	10 000	5 544	10 000	10 000	10 000	10 000
Average read length		151	151	151	151	151	151	151	151	151	151	151	151
Average insert size		478	470	376	487	493	501	503	527	505	449	509	432
N50 (kbp)		233	326	143	136	271	169	238	148	168	115	175	167
QC-status (Bifrost)		OK	OK	Warning	OK	OK	OK	OK	OK	Warning	OK	OK	OK

Ten strains passed QC, with the genome sizes of strains 9145 and 9821 exceeding the threshold in Bifrost.

Table 37. Laboratory 131

		Laboratory 131											
Parameter	Ranges*	9024	9058	9120	9121	9132	9398	9420	9498	9633	9636	9902	9951
Detected species	{Ec} or {Se} or {Sf}	Ec	Ec	Ec	Ec	Ec	Ec	Ec	Ec	Ec	Ec	Ec	Ec
% Species 1		95.3	92.4	94.9	96.7	96.8	96.8	94.9	96.2	92.8	93.9	89.5	92.7
% Species 2		1.0	2.2	0.4	0.1	0.1	0.1	0.7	0.5	0.6	0.9	1.3	0.2
Unclassified reads (%)	{<100}	3.2	4.0	4.0	3.1	2.9	2.9	3.3	2.7	5.8	4.5	7.4	6.7
Length at >25 x min. coverage (Mbp)	{>4.64 ^ <5.56}	5.4	5.4	5.1	5.3	5.3	5.3	5.2	5.4	5.1	5.1	5.6	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}	207	267	125	128	126	101	121	194	123	88	224	154
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}	93	82	81	61	72	94	109	118	137	121	111	109
No. of reads (x 1 000)		3 449	3 104	2 852	2 251	2 645	3 427	3 930	4 376	4 789	4 215	4 300	4 203
Average read length		147	147	147	147	147	147	147	148	147	147	147	147
Average insert size		341	329	334	355	345	336	332	344	348	352	357	363
N50 (kbp)		138	115	148	184	184	184	146	184	135	158	150	139
QC-status (Bifrost)		OK	OK	OK	OK	OK	OK	OK	OK	OK	OK	Warning	Warning

Ten strains passed QC, with the genome sizes of strains 9902 and 9951 exceeding the threshold in Bifrost.

Table 38. Laboratory 132

		Laboratory 132											
Parameter	Ranges*	9224	9283	9372	9397	9459	9483	9486	9715	9739	9906	9946	9981
Detected species	{Ec} or {Se} or {Sf}	Ec	Ec	Ec	Ec	Ec	Ec	Ec	Ec	Ec,	Ec, Sf	Ec	Ec,
% Species 1		98.5	95.5	97.4	90.6	95.9	89.8	96.9	98.5	80.8	85.8	78.4	98.1
% Species 2		0.0	0.6	0.6	0.2	0.7	1.4	0.5	0.0	7.7	1.7	15.0	0.0
Unclassified reads (%)	{<100}	1.3	3.3	1.7	8.8	2.5	7.1	2.0	1.3	6.0	12.0	5.2	1.7

Length at >25 x min. coverage (Mbp)	{>4.64 ^ <5.56}	3.5	1.2	3.1	0.4	4.9	0.7	0.4	3.0	5.3	4.8	4.9	5.0
Length [1-25] x min. coverage (kbp)	{<250}	1 831.6	3791.7	2 284.6	4624.4	239.5	4 760.8	4 769.8	2 227.8	141.1	280.7	395.2	388.1
No. of contigs at 25 x min. coverage	{>0}	105	56	131	97	117	116	74	77	246	113	99	109
No. of contigs [1-25] x min. coverage	{<1 000}	122	305	113	1222	13	498	350	133	13	21	45	29
Average coverage	{>50}	31	23	26	20	46	18	20	26	62	61	54	46
No. of reads (x 1 000)		1 110	773	945	740	1578	698	729	917	2 293	2 134	1 926	1 672
Average read length		149	150	149	150	149	150	150	150	150	150	150	149
Average insert size		349	403	440	450	368	396	483	430	406	400	409	411
N50 (kbp)		132	51	89	8	119	31	49	141	104	135	135	210
QC-status (Bifrost)		Warning	Fail	Warning	Fail	Warning	Fail	Fail	Warning	Warning	OK	Warning	Warning

One strain passed the QC. Nine of the strains that failed or triggered warnings were associated with an average coverage below 50, as well as a low number of reads. Strains 9739 and 9946 exhibited high percentages in Species 2, which may indicate a contamination.

Table 39. Laboratory 133

		Laboratory 133											
Parameter	**Ranges***	**9177**	**9201**	**9312**	**9336**	**9350**	**9405**	**9466**	**9587**	**9611**	**9725**	**9851**	**9984**
Detected species	{Ec} or {Se} or {Sf}	Ec	Ec,	Ec, Se	Ec	Ec	Ec	Ec	Ec	Ec	Ec	Ec,	Ec, Sf
% Species 1		92.2	89.0	91.2	98.5	99.5	99.4	99.1	92.3	96.5	79.6	97.4	97.8
% Species 2		1.3	9.0	1.6	0.4	0.0	0.0	0.1	0.4	0.7	9.0	0.4	0.7
Unclassified reads (%)	{<100}	4.6	1.2	6.3	0.7	0.4	0.5	0.7	6.9	1.4	6.2	1.8	0.7
Length at >25 x min. coverage (Mbp)	{>4.64 ^ <5.56}	5.6	5.2	5.1	5.4	5.3	5.4	5.3	5.6	5.1	5.5	5.1	5.3
Length [1-25] x min. coverage (kbp)	{<250}	4.8	0.0	0.0	4.2	0.0	0.0	0.0	3.4	0.0	4.3	0.0	61.2
No. of contigs at 25 x min. coverage	{>0}	180	95	116	172	76	102	120	135	109	245	73	159
No. of contigs [1-25] x min. coverage	{<1 000}	2	0	0	1	0	0	0	1	0	4	0	20
Average coverage	{>50}	59	69	51	62	78	49	44	81	71	78	74	55
No. of reads (x 1 000)		1 171	1 250	926	1 184	1 433	895	813	1 582	1 296	1 501	1 311	1 038
Average read length		289	291	291	291	292	293	291	293	283	291	291	291
Average insert size		429	434	441	453	453	477	460	479	369	446	443	465
N50 (kbp)		161	135	135	236	314	269	184	114	127	114	168	149
QC-status (Bifrost)		Warning	Warning	OK	OK	OK	Warning	Warning	Warning	OK	Warning	OK	OK

Six strains passed the QC. The genome sizes of strains 9177 and 9587exceeded the threshold in Bifrost. Strains 9201 and 9725 showed high percentages in Species 2, which may indicate a contamination. Strains 9405 and 9466 had an average coverage below 50.

Table 40. Laboratory 134

		Laboratory 134											
Parameter	Ranges*	9115	9125	9214	9314	9437	9541	9598	9616	9740	9748	9799	9971
Detected species	{Ec} or {Se} or {Sf}	Ec	Ec	Ec	Ec	Ec	Ec	Ec	Ec	Ec	Ec	Ec	Ec
% Species 1		91.4	96.1	89.8	97.4	91.7	94.9	88.2	95.8	97.5	94.0	91.0	95.8
% Species 2		1.3	0.9	2.5	0.5	0.6	1.1	1.9	0.3	0.1	2.6	4.5	0.3
Unclassified reads (%)	{<100}	6.4	2.2	3.7	1.8	6.9	3.0	7.5	3.1	2.0	2.2	3.0	3.5
Length at >25 x min. coverage (Mbp)	{>4.64 ^ <5.56}	5.1	5.4	5.4	5.3	5.6	5.1	5.6	5.1	5.3	5.4	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}	124	210	306	150	181	88	262	119	137	231	133	112
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}	73	87	68	88	82	93	43	93	64	68	67	51
No. of reads (x 1 000)		2 517	3 183	2 631	3 185	3 148	3 194	1 679	3 218	2 334	2 604	2 392	1 824
Average read length		151	151	151	151	151	151	151	151	151	151	151	151
Average insert size		308	290	263	294	288	307	299	311	325	249	297	336
N50 (kbp)		145	138	113	163	136	159	112	148	184	146	144	206
QC-status (Bifrost)		OK	OK	Warning	OK	Warning	OK	Warning	OK	OK	OK	Warning	OK

Eight strains passed the QC. The sum of "% Species 1" and "% unclassified" were lower than the expected 95% for strains 9214 and 9799. The genome sizes of strains 9437 and 9598 exceeded the threshold in Bifrost. The average coverage was below 50 for strain 9598.

Table 41. Laboratory 135

		Laboratory 135											
Parameter	Ranges*	9018	9215	9410	9521	9585	9589	9595	9620	9640	9646	9674	9750
Detected species	{Ec} or {Se} or {Sf}	Ec	Ec	Ec	Ec	Ec	Ec	Ec	Ec	Ec	Ec	Ec	Ec
% Species 1		94.6	93.8	98.8	96.0	97.4	97.3	94.7	92.0	96.7	92.6	96.1	89.3
% Species 2		0.5	2.3	0.1	0.8	0.1	0.1	1.1	0.2	0.5	1.1	0.8	1.7
Unclassified reads (%)	{<100}	4.1	2.3	1.1	2.5	2.3	2.4	3.1	7.3	2.2	5.6	2.6	7.2
Length at >25 x min. coverage (Mbp)	{>4.64 ^ <5.56}	5.1	5.1	5.3	5.3	5.3	5.3	5.2	5.6	5.4	5.1	5.4	5.6
Length [1-25] x min. coverage (kbp)	{<250}	0.0	0.9	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}	128	95	134	170	160	114	165	185	236	136	304	199
No. of contigs [1-25] x min. coverage	{<1 000}	0	2	0	0	0	0	0	0	0	0	0	0
Average coverage	{>50}	188	166	160	105	178	170	193	181	176	187	190	61
No. of reads (x 1 000)		6 698	5 860	5 773	3 844	6 565	6 225	7 103	7 037	6 602	6 603	7 236	2 319

Average read length		146	148	149	149	146	146	144	146	146	146	146	149
Average insert size		273	297	322	492	270	281	230	272	267	277	274	458
N50 (kbp)		148	166	184	144	163	241	127	133	143	135	115	144
QC-status (Bifrost)		OK	OK	OK	OK	OK	OK	OK	Warning	OK	OK	OK	Warning

Ten strains passed QC, with the genome sizes of strains 9620 and 9750 exceeding the threshold in Bifrost.

Table 42. Laboratory 136

		Laboratory 136											
Parameter	Ranges*	9195	9273	9294	9395	9412	9567	9575	9666	9761	9858	9939	9962
Detected species	{Ec} or {Se} or {Sf}	Ec	Ec	Ec	Ec	Ec	Ec	Ec	Ec	Ec	Ec	Ec	Ec
% Species 1		96.5	92.5	97.6	96.0	93.9	95.3	95.3	92.6	89.2	95.9	96.9	98.1
% Species 2		0.1	1.3	0.1	0.4	2.3	0.8	1.4	0.3	2.1	0.8	0.6	0.1
Unclassified reads (%)	{<100}	3.4	5.6	2.2	2.8	2.7	3.1	2.4	6.6	6.8	2.6	2.0	1.7
Length at >25 x min. coverage (Mbp)	{>4.64 ^ <5.56}	5.3	5.1	5.3	5.1	5.4	5.1	5.2	5.6	5.6	5.4	5.4	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}	190	114	101	118	272	76	120	174	246	238	205	130
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}	685	225	265	431	255	302	287	283	493	364	371	363
No. of reads (x 1 000)		27 101	7 847	9 632	15 139	9 579	10 458	10 299	10 846	19 096	13 735	13 793	13215
Average read length		137	148	147	147	148	148	148	148	148	144	147	148
Average insert size		197	369	316	282	351	381	321	307	308	249	305	338
N50 (kbp)		184	135	314	168	115	166	148	143	165	136	236	206
QC-status (Bifrost)		OK	OK	OK	OK	OK	OK	OK	Warning	Warning	OK	OK	OK

Ten strains passed QC, with the genome sizes of strains 9666 and 9761 exceeding the threshold in Bifrost.

Table 43. Laboratory 138

		Laboratory 138											
Parameter	Ranges*	9150	9192	9264	9305	9352	9364	9388	9442	9443	9627	9663	9668
Detected species	{Ec} or {Se} or {Sf}	Ec	Ec	Ec	Ec	Ec	Ec	Ec	Ec	Ec	Ec	Ec	Ec
% Species 1		90.4	94.3	86.6	96.0	92.9	93.2	91.6	92.0	93.6	95.2	95.8	96.0
% Species 2		0.8	0.8	1.9	0.2	1.3	1.0	1.6	1.2	0.4	0.5	0.5	0.3
Unclassified reads (%)	{<100}	7.9	3.5	8.3	3.5	3.7	4.5	4.4	5.8	4.6	3.1	3.2	3.4
Length at >25 x min. coverage (Mbp)	{>4.64 ^ <5.56}	5.5	5.3	5.6	5.3	5.4	5.1	5.2	5.1	5.1	5.4	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	4.0	0.0	0.0
No. of contigs at 25 x min. coverage	{>0}	230	241	339	109	399	103	186	139	120	257	182	153
No. of contigs [1-	{<1 000}	0	0	0	0	0	0	0	0	0	13	0	0

25] x min. coverage													
Average coverage	{>50}	237	272	189	192	282	198	189	220	195	213	299	203
No. of reads (x 1 000)		9 468	10 357	7 898	7 005	11 390	7 054	7 638	7 799	6 941	8 190	11 361	7 528
Average read length		151	151	151	151	151	151	151	151	151	151	151	151
Average insert size		216	218	198	284	190	258	175	258	288	234	218	266
N50 (kbp)		136	141	165	241	104	148	148	135	151	184	184	232
QC-status (Bifrost)		OK	OK	Warning	OK	OK	OK	OK	OK	OK	OK	OK	OK

Eleven strains passed the QC. The genome size of strain 9264 exceeded the threshold in Bifrost and the sum of "% Species 1" and "% unclassified" was lower than the expected 95%.

Table 44. Laboratory 139

		Laboratory 139											
Parameter	Ranges*	9158	9158	9158	9158	9158	9158	9158	9158	9158	9158	9158	9158
Detected species	{Ec} or {Se} or {Sf}	Ec	Ec	Ec	Ec	Ec	Ec	Ec	Ec	Ec	Ec	Ec	Ec
% Species 1		92.1	92.1	92.1	92.1	92.1	92.1	92.1	92.1	92.1	92.1	92.1	92.1
% Species 2		1.4	1.4	1.4	1.4	1.4	1.4	1.4	1.4	1.4	1.4	1.4	1.4
Unclassified reads (%)	{<100}	5.4	5.4	5.4	5.4	5.4	5.4	5.4	5.4	5.4	5.4	5.4	5.4
Length at >25 x min. coverage (Mbp)	{>4.64 ^ <5.56}	5.1	5.1	5.1	5.1	5.1	5.1	5.1	5.1	5.1	5.1	5.1	5.1
Length [1-25] x min. coverage (kbp)	{<250}	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
No. of contigs at 25 x min. coverage	{>0}	117	117	117	117	117	117	117	117	117	117	117	117
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}	437	437	437	437	437	437	437	437	437	437	437	437
No. of reads (x 1 000)		15 363	15 363	15 363	15 363	15 363	15 363	15 363	15 363	15 363	15 363	15 363	15 363
Average read length		147	147	147	147	147	147	147	147	147	147	147	147
Average insert size		342	342	342	342	342	342	342	342	342	342	342	342
N50 (kbp)		119	119	119	119	119	119	119	119	119	119	119	119
QC-status (Bifrost)		OK	OK	OK	OK	OK	OK	OK	OK	OK	OK	OK	OK

Eleven strains passed the QC, with the genome size of strain 9369 exceeding the threshold in Bifrost.

Table 45. Laboratory 187

		Laboratory 187											
Parameter	Ranges*	9272	9272	9272	9272	9272	9272	9272	9272	9272	9272	9272	9272
Detected species	{Ec} or {Se} or {Sf}	Ec	Ec	Ec	Ec	Ec	Ec	Ec	Ec	Ec	Ec	Ec	Ec
% Species 1		98.0	98.0	98.0	98.0	98.0	98.0	98.0	98.0	98.0	98.0	98.0	98.0
% Species 2		0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1
Unclassified reads (%)	{<100}	1.8	1.8	1.8	1.8	1.8	1.8	1.8	1.8	1.8	1.8	1.8	1.8
Length at >25 x min. coverage (Mbp)	{>4.64 ^ <5.56}	5.3	5.3	5.3	5.3	5.3	5.3	5.3	5.3	5.3	5.3	5.3	5.3

		Laboratory 187											
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}	92	92	92	92	92	92	92	92	92	92	92	92
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}	88	88	88	88	88	88	88	88	88	88	88	88
No. of reads (x 1 000)		3 149	3 149	3 149	3 149	3 149	3 149	3 149	3 149	3 149	3 149	3 149	3 149
Average read length		149	149	149	149	149	149	149	149	149	149	149	149
Average insert size		367	367	367	367	367	367	367	367	367	367	367	367
N50 (kbp)		184	184	184	184	184	184	184	184	184	184	184	184
QC-status (Bifrost)		OK	OK	OK	OK	OK	OK	OK	OK	OK	OK	OK	OK

Nine strains passed the QC. The genome size of strain 9418 exceeded the threshold in Bifrost. Strains 9549 and 9586 showed high percentages in Species 2, which may indicate contamination.

Table 46. Laboratory 222

		Laboratory 222											
Parameter	**Ranges***	**9097**	**9097**	**9097**	**9097**	**9097**	**9097**	**9097**	**9097**	**9097**	**9097**	**9097**	**9158**
Detected species	{Ec} or {Se} or {Sf}	Ec	Ec	Ec	Ec	Ec	Ec	Ec	Ec	Ec	Ec	Ec	Ec
% Species 1		96.5	96.5	96.5	96.5	96.5	96.5	96.5	96.5	96.5	96.5	96.5	92.1
% Species 2		0.8	0.8	0.8	0.8	0.8	0.8	0.8	0.8	0.8	0.8	0.8	1.4
Unclassified reads (%)	{<100}	2.2	2.2	2.2	2.2	2.2	2.2	2.2	2.2	2.2	2.2	2.2	5.4
Length at >25 x min. coverage (Mbp)	{>4.64 ^ <5.56}	5.1	5.1	5.1	5.1	5.1	5.1	5.1	5.1	5.1	5.1	5.1	5.1
Length [1-25] x min. coverage (kbp)	{<250}	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
No. of contigs at 25 x min. coverage	{>0}	359	359	359	359	359	359	359	359	359	359	359	117
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}	189	189	189	189	189	189	189	189	189	189	189	437
No. of reads (x 1 000)		3 694	3 694	3 694	3 694	3 694	3 694	3 694	3 694	3 694	3 694	3 694	15 363
Average read length		264	264	264	264	264	264	264	264	264	264	264	147
Average insert size		0	0	0	0	0	0	0	0	0	0	0	342
N50 (kbp)		32	32	32	32	32	32	32	32	32	32	32	119
QC-status (Bifrost)		OK	OK	OK	OK	OK	OK	OK	OK	OK	OK	OK	OK

All strains passed the QC. Some QC values may be unreliable due to assembly issues for Ion Torrent data (contigs, average insert size, N50).
Quality assessment made by the EQA provider in-house quality control pipeline.
**: indicative QC ranges; Ec. E. coli, Se: S. enterica Sf: S.flexneri (listed if >5%).*

Annex 11. Accessing provided sequences

Table 47. Participants' description of strain15

Lab ID	Sero / Stx sub	ST	Cluster	QC Status	Description Strain15
EQA provider		**335**	**No**	**C**	**A sequence contaminated with app. 9.3% *E. albertii*, "% Species 2" = 9.4.**
19			No	C	The sample is contaminated with Escherichia albertii. The sample needs to be restreaked for pure culture and resequenced.
34			No	C	looks like mixed culture, we found O55 and O131 for serotype and eaeA types gamma1 and omicron
80			No	A	Good targets: 98.7% -KRAKEN: E. coli -Coverage: 65 -Size: 5.3 -N50: 85099 -Tot no of contigs: 315
88			No	B	QC failed due to the detection of an inter-species contamination with Escherichia albertii. Reads identified as belonging to the contaminating species by Kraken2 were removed from the Fastq files and then the INNUca pipeline was able to assemble the genome of Escherichia coli. The final assembly passed all the downstream QC criteria (contamination check, genome coverage and >95% loci called), and, for this reason, was used for outbreak investigation. Of note, all the allelic differences with the index case were visually inspected and confirmed in IGV.
90			No	A	QC parameters within the accepted values
100			No	C	Genome length is too big (6,3 Mbp), number of contigs is too large (1014). Possible contamination with E. albertii.
108			No	C	Genome size too high, contamination.
123			No	B	% good targets E. coli cgMLST:99,2 (our treshold:98%) species match: E.coli (no evidence for contamination) GC content: 50,4 Genome size:5,3 Av.Coverage:47 (our treshold: 50) N50: 239507 (our treshold: 50 000)
124			No	A	Strain 15 has a genome size of 5.3 Mb, 51 GC%, a relatively high number of N bases and an average coverage of 57. 95% alleles were called.
131			No	A	N50 a bit low
132			No	B	Fair amount of reads with poor Phred score (14.5% below 20). Assembly filters below Phred 20 and contig coverage was around 16. MLST was 100%, cgMLST 95.58%. Sufficient HQ reads for acceptable assembly.
133			No	A	All main QC criteria are satisfied (Avg Quality >30; Avg Read Coverage >30x; N50>30 000; Nr Contigs <500; Length: 5 000 000 – 5 800 000; Core Percentage >=95%)
134			No	C	number of contigs > 1 000 Genome size = 6.8 Mb coverage <50
135			No	B	Coverage 27, is too low (criterium >=30)
136			No	A	Values in acceptable range
138			No	B	Warning: contaminated with Escherichia albertii approx. 10% 1) contigs >=0bp - 1140 (x ≤ 605) 2) average coverage - 43 (x > 50) 3) GC% - 49.75 (50.3 < x ≤ 50.9) 4) N50 – 132 711 (x > 72925) 5) assembly length - 6 518 392 (4 909 000 < x ≤ 5 493 000)
139			No	B	Sequence quality within acceptable ranges regarding average phred score of the reads, genome length, N50, number of contigs. Average genome coverage below routine thresholds.
187			No	A	good coverage
222			No	C	Sample contaminated with Escherichia albertii

Table 48. Participants' description of strain20

Lab ID	Sero / Stx sub	ST	Cluster	QC Status	Description Strain20
EQA provider	**-**	**335**	**No**	**C**	**A nonCluster sequence (ST335) with low coverage,"Length at (1,25)X coverage (in Kbp)" and "Contigs at (1,25)X coverage" are high, where "Length at >25X coverage (in Mbp)" and "Contigs at 25X coverage" are low.**
19			No	C	The read coverage is too low, resulting in poor assembly (many contigs and small genome size) and low core% in the cgMLST analysis and is therefore not suitable for cluster analysis.
34			No	C	only 50% of cgMLST gene content covered avarage coverage about 14fold
80			No	C	Good targets: 50.4% (<90%) -KRAKEN: E. coli -Coverage: 15 (<50) -Size: 4.4 -N50: 1792 (<30 000) -Tot no of contigs: 3 739 (>1 000)
88			No	C	QC failed due to low coverage. It was not possible to assemble the genome of this sample as the depth of coverage was lower than allowed by INNUca (<15x for first estimated coverage).
90			No	C	Very few reads in the fastq files
100			No	C	to low avg. coverage (only 14x), number of contigs is too high (1 021), N50 is quite low (9 129), GC content is higher than expected (51,23), largest contig is only 35 973 bp.
108			No	C	Coverage too low
123			No	C	% good targets E. coli cgMLST:94,4 (our treshold:98%) species match: E.coli (no evidence for contamination) GC content: 51,1 Genome size:5,1 Av.Coverage:12 (our threshold: 50) N50: 9695 (our threshold: 50 000) Although this strain shows 2 AD to strain 0019 it presumably belongs to an outbreak because the 2 genes difference could be due to the missing 145 values in the analysis (only 94,4% good targets). In addition, this strain has the same complex type as strain 0019 (CT 49953).
124			No	C	Strain 20 has a genome size of 5.1 Mb, 52 GC%, a relatively high number of N bases and an average coverage of 14 (<30). 67% alleles were called (<80).
131			No	C	cgMLST Perc. Good Targets: 50,4 % (too low) Avg. Coverage: 14 x (too low) Appr. Genome Size: 4,4 Mb (smaller than expected) N50: 1792 (too short) Top Species (Match Identity): Escherichia coli (0.99) Contig Count (Assembled): 3739 (too high)
132			No	A	Sufficient HQ reads.
133			No	C	Majority of main QC criteria are not satisfied (Avg Read Coverage <30, N50 <30, Nr Contigs >500, Core Percentage <95% (only 68%). Not acceptable for analysis
134			No	C	number of contigs > 1 000 Coverage <50 (14) Number of targets <97% (64.6%)
135			No	C	Coverage 7, too low. Too few alleles called <90%. Low number of reads
136			No	C	Low genome size (4.7 Mb), low genome fraction (86.58%), Low N50 (9.7 Kbp), High conting number (753). Indicates DNA fragmentation
138			No	C	Fail: 1) contigs >=0bp - 1281 ($x \leq 605$) 2) average coverage - 14 ($x > 50$) 3) GC% - 51.43 ($50.3 < x \leq 50.9$) 4) N50 - 6063 ($x > 72925$) 5) assembly length - 4823558 ($4909000 < x \leq 5493000$)
139			No	C	Sequence quality below acceptable ranges regarding average phred score of the reads, genome length, N50, number of contigs, and average genome coverage. We can't determine whether this strain belongs to the cluster. Our default answer is no.
187			No	C	missing genome more than 20%
222			No	C	Very low coverage (14x)

-: no reported data/analysis performed

Annex 12. Word format of the online form

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

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

STEC EQA-13 2024-2025

Dear Participant,

Welcome to the thirteenth External Quality Assessment (EQA-13) scheme for typing of STEC in 2024-2025.

NOTE: New virulence gene *esta* (STa).
If you are using WGS, please read the WGS part of the submission protocol thoroughly before starting your analysis. This year, you are required to use a specific strain/sequence when reporting allele differences/SNP distances.

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

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

The available options in this participation form include:

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

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

1. Country

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

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

2. Institute name

3. Laboratory name

4. Laboratory ID

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

5. E-mail

6. STEC EQA-13 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.

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

7. Serotyping and virulence gene determination of STEC

8. Submitting results

(State one answer only)

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

9. Submitting results - Serotyping

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

10. Results for serotyping (O Group)

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

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

11. Please specify the method used:

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

Method:

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

12. Results for serotyping (H Type)

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

H type:
Strain 1 ___
Strain 2 ___
Strain 3 ___
Strain 4 ___
Strain 5 ___
Strain 6 ___
Strain 7 ___
Strain 8 ___
Strain 9 ___
Strain 10 ___
Strain 11 ___
Strain 12 ___

13. Please specify the method used:

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

Method:

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

14. Submitting results – Virulence gene determination

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

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

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

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

17. Results for virulence gene determination

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

	eae	*aagR*	*esta (Sta)*	*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

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

19. Results for subtyping

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

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

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

All isolates have to be subtyped regardless of the results of the initial screening. 'ND' will by default be evaluated as an incorrect result.

	stx2a	stx2b	stx2c	stx2d	stx2e	stx2f	stx2g	stx2a; stx2b	stx2a; stx2c	stx2a; stx2d	stx2a; stx2e	stx2a; stx2g	stx2b; stx2c	stx2b; stx2d	stx2b; stx2g	stx2c; stx2d	stx2c; stx2e	stx2c; stx2g	stx2d; stx2e	stx2d; stx2g	stx2e; stx2f	stx2a; stx2b; stx2c	stx2a; stx2c; stx2d	stx2b; stx2c; stx2d	stx2a; stx2b; stx2c; stx2d	Negative	ND
Strain 1	▯	▯	▯	▯	▯	▯	▯	▯	▯	▯	▯	▯	▯	▯	▯	▯	▯	▯	▯	▯	▯	▯	▯	▯	▯	▯	▯
Strain 2	▯	▯	▯	▯	▯	▯	▯	▯	▯	▯	▯	▯	▯	▯	▯	▯	▯	▯	▯	▯	▯	▯	▯	▯	▯	▯	▯
Strain 3	▯	▯	▯	▯	▯	▯	▯	▯	▯	▯	▯	▯	▯	▯	▯	▯	▯	▯	▯	▯	▯	▯	▯	▯	▯	▯	▯
Strain 4	▯	▯	▯	▯	▯	▯	▯	▯	▯	▯	▯	▯	▯	▯	▯	▯	▯	▯	▯	▯	▯	▯	▯	▯	▯	▯	▯
Strain 5	▯	▯	▯	▯	▯	▯	▯	▯	▯	▯	▯	▯	▯	▯	▯	▯	▯	▯	▯	▯	▯	▯	▯	▯	▯	▯	▯
Strain 6	▯	▯	▯	▯	▯	▯	▯	▯	▯	▯	▯	▯	▯	▯	▯	▯	▯	▯	▯	▯	▯	▯	▯	▯	▯	▯	▯
Strain 7	▯	▯	▯	▯	▯	▯	▯	▯	▯	▯	▯	▯	▯	▯	▯	▯	▯	▯	▯	▯	▯	▯	▯	▯	▯	▯	▯
Strain 8	▯	▯	▯	▯	▯	▯	▯	▯	▯	▯	▯	▯	▯	▯	▯	▯	▯	▯	▯	▯	▯	▯	▯	▯	▯	▯	▯
Strain 9	▯	▯	▯	▯	▯	▯	▯	▯	▯	▯	▯	▯	▯	▯	▯	▯	▯	▯	▯	▯	▯	▯	▯	▯	▯	▯	▯
Strain 10	▯	▯	▯	▯	▯	▯	▯	▯	▯	▯	▯	▯	▯	▯	▯	▯	▯	▯	▯	▯	▯	▯	▯	▯	▯	▯	▯
Strain 11	▯	▯	▯	▯	▯	▯	▯	▯	▯	▯	▯	▯	▯	▯	▯	▯	▯	▯	▯	▯	▯	▯	▯	▯	▯	▯	▯
Strain 12	▯	▯	▯	▯	▯	▯	▯	▯	▯	▯	▯	▯	▯	▯	▯	▯	▯	▯	▯	▯	▯	▯	▯	▯	▯	▯	▯

21. Submitting Cluster results

- ❑ Cluster analyses based on PFGE and/or WGS
- ❑ Did not participate in the Cluster part – Go to 116

22. Submitting Cluster analysis results

- ❑ 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

- ❑ Cluster analysis based on WGS data – Go to 27
- ❑ Do not wish to submit any cluster results based on WGS data – Go to 116

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

- ❑ 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

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

32. Reference genome used:

Preferable use EQA strain0018 (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

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

38. If another tool is used please enter here:

39. Please indicate allele calling method:

- ❑ 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

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

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

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

Cluster detected by analysis on data derived from WGS

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

44. Please list the ID for the strains included in the cluster of closely related strains detected by WGS: please use semicolon (;) to separate the ID´s.
This includes the 12 test strains and the 8 provided sequences (20 in total). For the provided sequences write the numbers like: 0013, 0014, 0015, 0016 ect.

45. Report the ID, part of the cluster (yes/no), and SNP distance/allele difference

Please use 9999 for not analyzed

	ID	Cluster (Yes/No)	AD/SNP
Strain1	_____	☐(Yes) ☐(No)	_____
Strain2	_____	☐(Yes) ☐(No)	_____
Strain3	_____	☐(Yes) ☐(No)	_____
Strain4	_____	☐(Yes) ☐(No)	_____
Strain5	_____	☐(Yes) ☐(No)	_____
Strain6	_____	☐(Yes) ☐(No)	_____
Strain7	_____	☐(Yes) ☐(No)	_____
Strain8	_____	☐(Yes) ☐(No)	_____
Strain9	_____	☐(Yes) ☐(No)	_____
Strain10	_____	☐(Yes) ☐(No)	_____
Strain11	_____	☐(Yes) ☐(No)	_____
Strain12	_____	☐(Yes) ☐(No)	_____

46. For each ID report: part of the cluster (yes/no), QC status (A/B/C), QC comment and SNP distance/allele difference

QC status:
Please select the QC status that fits with your assessment of the strain
A = Acceptable quality, B = Quality only acceptable for outbreak situations (less good quality), C = Not acceptable quality - strain not analyzed

Distance:
Please use 9999 for not analyzed

	Cluster (Yes/No)	QC (A/B/C)	QC comment	AD/SNP
Strain0013	☐(Yes) ☐(No)	☐(A) ☐(B) ☐(C)	_____	_____

Strain0014	☐(Yes) ☐(No)	☐(A) ☐(B) ☐(C)	_____	_____
Strain0015	☐(Yes) ☐(No)	☐(A) ☐(B) ☐(C)	_____	_____
Strain0016	☐(Yes) ☐(No)	☐(A) ☐(B) ☐(C)	_____	_____
Strain0017	☐(Yes) ☐(No)	☐(A) ☐(B) ☐(C)	_____	_____
Strain0018	☐(Yes) ☐(No)	☐(A) ☐(B) ☐(C)	_____	_____
Strain0019	☐(Yes) ☐(No)	☐(A) ☐(B) ☐(C)	_____	_____
Strain0020	☐(Yes) ☐(No)	☐(A) ☐(B) ☐(C)	_____	_____

47. (Optional) Would you like to add additional information for the strains? e.g. serotype or sequence type (ST)

☐ Yes
☐ No – Go to 48

	Serotype	**Subtype**	**Sequence type (ST)**
Strain1	_____	_____	_____
Strain2	_____	_____	_____
Strain3	_____	_____	_____
Strain4	_____	_____	_____
Strain5	_____	_____	_____
Strain6	_____	_____	_____
Strain7	_____	_____	_____
Strain8	_____	_____	_____
Strain9	_____	_____	_____

Strain10	_____	_____	_____
Strain11	_____	_____	_____
Strain12	_____	_____	_____
Strain0013	_____	_____	_____
Strain0014	_____	_____	_____
Strain0015	_____	_____	_____
Strain0016	_____	_____	_____
Strain0017	_____	_____	_____
Strain0018	_____	_____	_____
Strain0019	_____	_____	_____
Strain0020	_____	_____	_____

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

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

- ❑ Yes – Go to 49
- ❑ No – Go to 86

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

- ❑ SNP-based – Go to 51
- ❑ Allele-based – Go to 58
- ❑ Other – Go to 50

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

51. Please report the used SNP-pipeline

(reference if publicly available or in-house pipeline)

52. Please select the approach used for the SNP analysis

- ❑ Reference-based – Go to 53
- ❑ Assembly-based – Go to 56

53. Reference genome used:

(preferable use EQA strain 0018, downloaded sequences as reference). Otherwise indicate Multi-locus Sequence Type (e.g. ST8) and isolate ID

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

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

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

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

58. Please select tool used for the allele analysis

- ❑ BioNumerics – Go to 60
- ❑ SeqSphere – Go to 60
- ❑ Enterobase – Go to 60
- ❑ Other – Go to 59

59. If another tool is used please list here:

60. Please indicate allele calling method:

- ❑ Assembly-based and mapping-based – Go to 61
- ❑ Only assembly-based – Go to 61
- ❑ Only mapping-based – Go to 62

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

62. Please select scheme used for the allele analysis

- ❑ Applied Maths (wgMLST) – Go to 64
- ❑ Applied Maths (cgMLST/Enterobase) – Go to 64
- ❑ Enterobase (cgMLST) – Go to 63
- ❑ Other – Go to 63

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

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

65. Additional analysis on data derived from WGS

66. Results for the additional cluster analysis.

Reporting allele differences/SNP distances to strain 0018 (as downloaded sequence) (e.g. SNP- or Allele-based)
Please use 9999 for not analysed.

	Distance/difference (e.g. SNP/allele) to the strain 0018 (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)	___

67. 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

- ❑ Yes – Go to 68
- ❑ No – Go to 86

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

- ❑ SNP-based – Go to 70
- ❑ Allele-based – Go to 77
- ❑ Other – Go to 69

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

70. Please report the used SNP-pipeline

Reference if publicly available or in-house pipeline.

71. Please select the approach used for the SNP analysis

- ❑ Reference-based – Go to 72
- ❑ Assembly-based – Go to 75

72. Reference genome used:

(preferable use EQA strain 0018, downloaded sequences as reference). Otherwise indicate Multi-locus Sequence Type (e.g. ST8) and isolate ID

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

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

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

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

77. Please select tool used for the allele analysis

- ❑ BioNumerics – Go to 79
- ❑ SeqSphere – Go to 79
- ❑ Enterobase – Go to 79
- ❑ Other – Go to 78

78. If another tool is used please enter here:

79. Please indicate allele calling method:

- ❑ Assembly-based and mapping-based – Go to 80
- ❑ Only assembly-based – Go to 80
- ❑ Only mapping-based – Go to 80

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

81. Please select scheme used for the allele analysis

- ❑ Applied Maths (wgMLST) – Go to 83
- ❑ Applied Maths (cgMLST/Enterobase) – Go to 83
- ❑ Enterobase (cgMLST) – Go to 83
- ❑ Other – Go to 82

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

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

84. Third analysis on data derived from WGS

85. Results for the third cluster analysis

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

	Distance/difference (e.g. SNP/allele) to the strain 0018 (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) ___

86. Additional questions to the WGS part

87. Where was the sequencing performed

- ❑ In own laboratory
- ❑ Externally

88. Protocol used to prepare the library for sequencing:

- ❑ Commercial kits – Go to 89
- ❑ Non-commercial kits – Go to 91

89. Please indicate name of commercial kit:

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

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

233. The sequencing platform used

- ❑ Ion Torrent PGM – Go to 94
- ❑ Ion Torrent Proton – Go to 94
- ❑ Ion S5 XL System – Go to 94
- ❑ Ion Genestudio S5 system – Go to 94
- ❑ Genome Sequencer Junior System (454) – Go to 94
- ❑ Genome Sequencer FLX System (454) – Go to 94
- ❑ Genome Sequencer FLX+ System (454) – Go to 94
- ❑ PacBio RS II – Go to 94
- ❑ PacBio RS – Go to 94
- ❑ HiScanSQ – Go to 94
- ❑ HiSeq 1000 – Go to 94
- ❑ HiSeq 1500 – Go to 94
- ❑ HiSeq 2000 – Go to 94
- ❑ HiSeq 2500 – Go to 94
- ❑ HiSeq 4000 – Go to 94
- ❑ Genome Analyzer lix – Go to 94
- ❑ MiSeq – Go to 94
- ❑ MiSeq Dx – Go to 94
- ❑ MiSeq FGx – Go to 94
- ❑ ABI SOLiD – Go to 94
- ❑ NextSeq – Go to 94
- ❑ MinION (ONT) – Go to 94
- ❑ Mini Seq Illumina – Go to 94
- ❑ Other – Go to 93

93. If another platform is used please list here:

94. Criteria used to evaluate the quality of sequence data

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

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

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

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

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

- ❑ Yes
- ❑ No – Go to 97

96. Procedure used to evaluate confirmation of genus:

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

- ❑ Yes
- ❑ No – Go to 99

98. Procedure or threshold used for coverage:

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

- ❑ Yes
- ❑ No – Go to 101

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

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

- ❑ Yes
- ❑ No – Go to 105

102. Procedure or threshold used for genome size:

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

- ❑ Yes
- ❑ No – Go to 105

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

105. 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).

106. Other criteria used to evaluate the quality of sequence data –

additional criteria 1:

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

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

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

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

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

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

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

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

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

116. Comment(s):

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

117. Please remember to upload your raw reads to the sFPT site:

https://sit-ftp.statens-it.dk/

Code: EQA_STEC13_upload

Have you remembered to upload your raw reads?

❑ Yes

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

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

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

❑ Yes

Thank you for your participation

Thank you for filling out the Submission form for the STEC EQA-13.

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

Remember to press "Finish" to complete submission.

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

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

European Centre forDisease Prevention and Control (ECDC)

Gustav III:s Boulevard 40
16973 Solna, Sweden

Tel. +46 858 60 10 00
ECDC.info@ecdc.europa.eu

www.ecdc.europa.eu

Publications Office
of the European Union