ecdc
EUROPEAN CENTRE FOR DISEASE PREVENTION AND CONTROL

TECHNICAL REPORT

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

www.ecdc.europa.eu

ECDC TECHNICAL REPORT

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

This report was commissioned by the European Centre for Disease Prevention and Control (ECDC), coordinated by Taina Niskanen (ECDC Food- and Waterborne Diseases and Zoonoses Programme) and produced by Susanne Schjørring, Gitte Sørensen, Kristoffer Kiil, Flemming Scheutz, Malgorzata Ligowska-Marzeta, and Eva Møller Nielsen of the Foodborne Infections Unit at Statens Serum Institut, Copenhagen, Denmark.

Suggested citation: European Centre for Disease Prevention and Control. External quality assessment scheme for typing of Shiga toxin-producing *Escherichia coli*. Stockholm: ECDC; 2019.

Stockholm, February 2019

ISBN 978-92-9498-308-4
DOI 10.2900/028998
Catalogue number TQ-02-19-159-EN-N

Contents

Figures

Tables

Annexes

Abbreviations

aaiC	Chromosomal gene marker for enteroaggregative *E. coli*
AEEC	Attaching and effacing *E. coli*
aggR	Gene encoding the master regulator in enteroaggregative *E. coli*
BN	BioNumerics
bp	Base pairs
cgMLST	Core genome multilocus sequence type
eae	CVD434. *E. coli* attaching and effacing gene probe *ehxA*
EAEC	Enteroaggregative *E. coli*
EFSA	European Food Safety Authority
eltA G119.	Heat labile enterotoxin (LT), almost identical to cholera toxin
EQA	External quality assessment
estA DAS101.	Heat stable enterotoxin (estAp; porcine variant: STp. estAh; human variant: STh; STIa)
EURL	European Union Reference Laboratory
FWD	Food- and waterborne diseases
FWD-Net	Food- and Waterborne Diseases and Zoonoses Network
GFN	Global Foodborne Infections Network, Food Safety, WHO
HUS	Haemolytic uraemic syndrome
ND	Not done
NSF	Non-sorbitol fermenter
NT	Non-typeable
PFGE	Pulsed-field gel electrophoresis
PH NRL	Public health national reference laboratories
QC	Qualitative control
SNP	Single nucleotide polymorphism
SOP	Standard operating procedure
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 type
WGS	Whole genome sequencing

Executive summary

This report presents the results of the eighth round of the external quality assessment (EQA-8) scheme for typing of Shiga toxin-producing *Escherichia coli* (STEC) organised for laboratories 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 this EQA under a framework contract with ECDC. EQA-8 contains serotyping, detection of virulence genes and molecular typing-based cluster analysis.

Human STEC infection is a zoonotic disease with an EU notification rate of 1.82 cases per 100 000 population in 2016. The number of STEC cases in the EU has increased since 2015 and the most commonly reported STEC O group was O157 (38.6% of cases with known serogroup).

Since 2007, ECDC's FWD programme has been responsible for EU-wide surveillance of STEC, including facilitating, detecting and investigating food-borne outbreaks. Surveillance data, including basic typing parameters and molecular typing data for the isolated pathogen, are reported by Member States to The European Surveillance System (TESSy). The surveillance system relies on the capacity of laboratories providing data to FWD-Net to produce comparable typing results. In order to ensure the EQA is linked to the development of surveillance methods used by public health national reference laboratories (PH NRLs) in Europe, EQA-8 contains a molecular typing-based cluster analysis using either pulsed-field gel electrophoresis (PFGE) and/or whole genome sequencing (WGS)-derived data, while the quality assessment of PFGE performed in previous years has been excluded.

The objectives of the EQA are to assess the quality and comparability of typing data reported by PH NRLs participating in FWD-Net. Test isolates for the EQA were selected to cover strains currently relevant to public health in Europe and represent a broad range of clinically relevant types for STEC. Two separate sets of 12 test isolates were selected for serotyping/virulence profiling and molecular typing-based cluster analysis respectively. Twenty-seven laboratories registered and 25 completed the exercise, representing a decrease in participation of 17% from the previous assessment (EQA-7). This decrease in the number of participants may have been caused by adding molecular typing-based cluster analysis (using PFGE and/or WGS without a standard protocol) or removing two independent methods that were covered in previous assessments: quality assessments of PFGE and phenotypic analysis.

The full O:H serotyping was performed by 60% (15/25) of participating laboratories, with an average score of 86%. In general, the more common European serotypes generated the highest scores, e.g. 100% for both O157:H7 isolates, while the less frequent O187:H28 obtained an average score of only 53%. Notably, not all laboratories demonstrated the capacity to determine all O groups and H types and participation in H typing was low (15/25). Capacity building, as well as use of a wider range of antisera, would be beneficial, as 57% (13/23) of participating laboratories still perform traditional serotyping with the use of antisera. The majority of the incorrect results were reported as not done (ND) or non-typeable (NT) for both O grouping (78%; 46/59) and H typing (86%; 12/14).

The quality of the virulence profile results was generally good, with high average scores for *eae* (99%), *vtx1* (99%) and *vtx2* (98%), similar to previous EQAs. Seventeen participants (85%) identified the two enteroaggregative *E. coli* (EAEC) isolates by correctly reporting the presence of the *aaiC* and *aggR* genes. Subtyping of *vtx1* and *vtx2* obtained a combined average score of only 77%, much lower than previous EQAs (EQA-4: 90%; EQA-5: 92%, EQA-6: 91%; EQA-7: 90%). The majority of incorrect results were caused by the unexpected reporting of ND in the *stx* negative isolates (35/38 for *stx1* and 15/32 for *stx2*). Almost half of the 'true' *stx2* errors (6/13) were reported for one specific isolate harbouring *stx2a* and *stx2d* or *stx2e* (using WGS).

Out of the 25 laboratories participating in EQA-8, 18 (72%) performed molecular typing-based cluster analysis using any method. The idea of the cluster analysis part of the EQA was to assess the PH NRL's ability to identify a cluster of genetically closely related isolates given that a multitude of different laboratory methods and analytical methods are used as the primary cluster detection approach in Member States. This part of the EQA was atypical in the sense that the aim was to assess the participants' ability to reach the correct conclusion, i.e. to correctly categorise the cluster test isolates, not to follow a specific procedure.

The cluster of closely related isolates contained four ST21 isolates that could be identified by both PFGE and WGS-derived data. The expected cluster was based on a predefined categorisation by the organiser. Notably, half the laboratories (9/18) used PFGE for cluster analysis and only two also reported cluster analysis based on WGS data.

Eleven laboratories performed cluster analysis using WGS-derived data. Performance was high, with 10 (91%) of participants correctly identifying the cluster of closely related isolates. In this EQA, participants were free to choose their preferred analytical method for the WGS-based cluster identification. An allele-based method was preferred since 72% (8/11) used core genome or whole genome multilocus sequence type (cgMLST/wgMLST) compared to 27% (3/11) using single nucleotide polymorphism (SNP) for the main reported cluster analysis.

Allele- and SNP-based methods seemed equally suitable for cluster identification. In general, for both cgMLST and wgMLST, the reported results were at a comparable level despite analysis with different schemes. The reported SNP results showed more variability. 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 for comparison and communication between laboratories. This issue is further complicated as many laboratories still use PFGE and will probably not switch to WGS in the near future. In this EQA, 39% (7/18) of participants in cluster analysis only used PFGE and three did not identify the correct cluster.

1 Introduction

1.1 Background

ECDC is an independent EU agency with a mandate to operate dedicated surveillance networks. Its mission is to identify, assess and communicate current and emerging threats to human health from communicable diseases. ECDC fosters the development of sufficient capacity within the EU/EEA's network for diagnosis, detection, identification and characterisation of infectious agents that may threaten public health. ECDC maintains and extends such cooperation and supports the implementation of quality assurance schemes [1].

External quality assessments (EQAs) are an essential part of quality management and use an external evaluator to assess the performance of participating laboratories on test samples supplied specifically for the purpose.

ECDC's disease networks organise a series of EQAs for EU/EEA countries. The aim of EQAs is to identify needs for improvement in laboratory diagnostic capacities relevant to epidemiological surveillance of communicable diseases as in Decision No 1082/2013/EU [2] and ensure reliability and comparability of results generated by laboratories across all EU/EEA countries.

The main purposes of EQA schemes are to:

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

Since 2012, the unit of Foodborne Infections at Statens Serum Institut (SSI), Denmark has been the EQA provider for the three lots covering typing of *Salmonella enterica* ssp. *enterica*, Shiga toxin/verocytotoxin-producing *Escherichia coli (E. coli;* STEC/VTEC) and *L. monocytogenes*. In 2016, SSI was granted the new round of tenders (2017–2020) for all three lots. For lot 2 (STEC) from 2017, the EQA scheme no longer covers assessment of PFGE quality. However, it still covers serotyping, virulence profiling and molecular typing-based cluster analysis. This report presents the results of the eighth EQA scheme (STEC EQA-8).

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 development of the disease. A large number of serotypes of *E. coli* have been recognised as Stx producers. Notably, the majority of reported human STEC infections are sporadic cases. Symptoms associated with STEC infection in humans vary from mild diarrhoea to life-threatening haemolytic uraemic syndrome (HUS), which is clinically defined by a combination of haemolytic anaemia, thrombocytopenia and acute renal failure.

In 2016, the EU notification rate of STEC infections was 1.82 cases per 100 000 population. The total number of confirmed STEC infection cases was 6 378, an increase from 2015 (n=5 929). Ten deaths due to STEC infection were reported, resulting in an EU case fatality of 0.3%. As in previous years, the most commonly reported STEC O group was O157 (38.6% of cases with known serogroup). O group O157 was followed by O26 [3].

ECDC's FWD programme is responsible for EU-wide surveillance of FWDs and facilitating detecting and investigating food-borne outbreaks since 2007. One of the key objectives for the FWD programme is improving and harmonising the surveillance system in the EU to increase scientific knowledge of aetiology, risk factors and burden of FWDs and zoonoses. Surveillance data, including some basic typing parameters for the isolated pathogen, are reported by Member States to TESSy. In addition to the basic characterisation of the pathogens isolated from infections, there is public health value to use more discriminatory typing techniques in the surveillance of food-borne infections. Since 2012, ECDC has enhanced surveillance incorporating molecular typing data ('molecular surveillance'). Three selected FWD pathogens were included: *Salmonella enterica* ssp. *enterica, L. monocytogenes* and STEC/VTEC. The overall aims of integrating molecular typing into EU level surveillance are to:

- foster 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 emergence of new evolving pathogenic isolates
- support investigations to trace-back 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 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.

EQA schemes have targeted PH NRLs already expected to be performing molecular surveillance at the national level.

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 activable *stx2d* subtype in *eae*-negative STEC appear to be highly associated with the serious sequela HUS [4–6]. Other specific subtypes of Stx1 and Stx2 are primarily associated with milder course of disease without HUS [4–6].

Understanding the epidemiology of *stx* subtypes is therefore important to reduce the risk of STEC infection and for the surveillance of STEC.

The recommended method for *stx* subtyping is a specific PCR [7]. STEC serotype O157:[H7] may be divided into two groups: non-sorbitol fermenters (NSF) and a highly virulent variant of O157-fermenting sorbitol (SF). STEC EQA-8 included O:H serotyping, detection of virulence genes (*eae*, *stx1* and *stx2,* including subtyping of *stx* genes) and genes for enteroaggregative *E. coli* (EAEC) and molecular typing-based cluster analysis.

1.4 Objectives

1.4.1 Serotyping

The objectives of STEC EQA-8 were to assess the ability to assign correct O groups and H types by using either serological (somatic 'O' and flagellar 'H' antigens) or molecular typing methods (PCR or WGS).

1.4.2 Virulence profile

The objectives of the virulence gene determination of STEC EQA-8 were to assess the ability to assign the correct virulence profile.

The presence/absence of *stx1*, *stx2*, *eae, aaiC* and *aggR* genes and subtyping of stx genes (*stx1a, stx1c* and *stx1d* and *stx2a* to *stx2g*).

1.4.3 Molecular typing-based cluster analysis

The objective of the molecular typing-based cluster analysis of STEC EQA-8 was to assess the ability to detect a cluster of closely related isolates. Laboratories could perform the analyses using PFGE or derived data from WGS.

2 Study design

2.1 Organisation

STEC EQA-8 was funded by ECDC and arranged by SSI following ISO/IEC 17043:2010 [8]. EQA-8 included serotyping, virulence gene determination and a molecular typing-based cluster analysis and was carried out between November 2017 and March 2018.

Invitations were emailed to ECDC contact points in FWD-Net (30 countries) by 1 November 2017 with a deadline to respond by 7 November 2017. In addition, invitations were sent to EU candidate and potential candidate countries Albania, Bosnia and Herzegovina, the former Yugoslav Republic of Macedonia, Kosovo[i], Montenegro, Serbia and Turkey.

Twenty-seven PH NRLs in EU/EEA and EU candidate countries accepted the invitation to participate and 25 submitted results (Annex 1). EQA test isolates were sent to participants on 4 December 2017. In Annex 2, participation details in EQA-7 and EQA-8 are listed to give an overview of the trend in the number of participants. In addition, 18 laboratories not funded by ECDC participated in the EQA. Participants were asked to submit their results to an SFTP-site and complete the online form by 23 February 2018 (Annex 14). Due to delays during shipping and busyness in the laboratories, the deadline was extended to 23 March 2018.

The EQA submission protocol, invitation letter and an empty submission form were available on the online site.

2.2 Selection of test isolates

Forty-seven test isolates were selected to fulfil the following criteria:

- represent commonly reported strains in Europe
- remain stable during the preliminary test period at the organising laboratory; and
- include closely related isolates.

The 47 selected isolates were analysed using the methods used in the EQA before and after having been re-cultured 10 times. All candidate isolates remained stable using these methods and the final test isolates were selected. The 12 test isolates (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). The 12 test isolates for cluster analysis were selected to include isolates with different or varying relatedness and different multilocus sequence types (ST16, 17, 21 and 29). One set of technical duplicates was included in the cluster test isolates. (Annexes 5–8, 10–11). Using either PFGE or WGS-derived data, the cluster of closely related isolates consisted of four STEC ST21 isolates (one technical duplicate). The characteristics of all the STEC test isolates are listed as 'Original/REF' in Annexes 3–11.

Table 1. Characterisation of test isolates

Parts	Number of test isolates	Characterisation	Annexes
Serotyping	12#	O8:H9, O91:H14, O63:H6, O104:H4, O111:H8/H-, O121:H19, O126;H27/H-, O146:H28/H-, O154:H31, O157:H7 (x2), O187:H28	3
Virulence profile	12#	*aaiC aggR* (x2), *eae stx*2a, *stx*2g, *eae stx2a stx*2c, *eae stx*2f, *stx*1a *stx*2b (x2) *stx2*a *stx2*d or stx2e, *stx*1d, *eae stx*1a, *eae* stx1a stx2a	4
Cluster analysis	12	ST16, ST17, ST21 (x9) (O26:H11, *stx1a*), and ST29	5–13

#: same 12 isolates.

2.3 Carriage of isolates

All test isolates were blinded and shipped on 4 December 2017. Letters stating the unique isolate IDs were included in the packages and distributed individually to the participants by email on 5 December 2017 as an extra precaution. Sixteen participants received the isolates within one day, nine within three days and two nine and 16 days after shipment respectively. No participants reported damage to the shipment or errors in the unique isolate IDs.

In January 2018, instructions for the submission of results procedure were emailed to the participants. This included the links to the online uploading site and submission form.

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

2.4 Testing

In the serotyping part, 12 STEC isolates were tested to assess the participants' ability to obtain the correct serotype. The participants could perform conventional serological methods according to suggested protocol [9] or molecular-based serotyping (PCR or WGS). The serotypes were submitted in the online form.

The same set of isolates for serotyping analysis was used for 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 *aaiC* and *aggR* (two genes related to EAEC), *eae* and *stx1* and *stx2,* as well as subtyping of *stx* genes *stx1 (stx1a, stx1c or stx1d)* and *stx2 (stx2a - stx2g)* according to suggested protocol [10]. The results were submitted in the online form.

In the molecular typing-based cluster analysis part, participants could perform the laboratory part using PFGE or WGS-derived data. Participants were instructed to report the IDs of isolates included in the cluster of closely related isolates by method. If PFGE analysis was conducted, the participant reported the total number of bands and number of shared bands with a selected cluster representative isolate.

Laboratories performing WGS could use their own analysis pipeline for cluster analysis, e.g. single nucleotide polymorphism analysis (SNP-based) or whole genome multilocus sequence typing (wgMLST)/cgMLST (allele-based) and were asked to submit the isolates identified as a cluster of closely related isolates based on the analysis used. Laboratories could report results from up to three analyses (1 main and 0–2 additional), but the detected cluster had to be based on results from the main analysis. The laboratories reported SNP distance or allelic differences between a selected cluster isolate and each test isolate and uploaded the raw reads (FASTQ files) to an SFTP site.

2.5 Data analysis

As the participating laboratories submitted their results, the serotype, virulence profile and cluster analysis results, as well as the participants' uploaded raw reads, were imported to a dedicated STEC EQA-8 BioNumerics (BN) database. The EQA provider reported to participants if errors in the submission process were identified, thereby obtaining analysable results. The EQA provider was in contact with five participants in order to ensure sequences were uploaded to the SFTP site. One participant had difficulties achieving high enough quality of the PFGE gel in order to make the cluster analysis and data were not submitted.

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

Virulence profile results were evaluated according to the percentage of correct results, generating a score from 0–100% for *eae*, *aaiC*, *aggR*, *stx1*, *stx2*, subtyping of *stx1* and *stx2* and combined subtype.

Molecular typing-based cluster analysis was evaluated according to correct or incorrect identification of the expected cluster of closely related isolates based on a pre-defined categorisation by the organiser. The EQA provider's PFGE results were based on *Xba*I profiles [11]. The EQA provider's WGS-derived cluster analysis was based on allele-based cgMLST [12] and SNP analysis (NASP) [13]. The correct number of closely related STEC isolates (4) could be identified by both PFGE and WGS-derived data. The cluster contained four ST21 isolates: REF14, REF15, REF19 and REF23 (REF19 and REF23 were technical duplicates). The EQA provider found at most two allele differences or four SNPs between any two isolates in the cluster. The rest of the cluster test isolates were an additional five ST21s, one ST16, one ST17 and one ST29.

Individual evaluation reports were distributed to participants in early July 2018 and certificates of attendance in October 2018. If WGS data were used, the evaluation report included a quality assessment made by the EQA provider's in-house quality control pipeline (e.g. coverage, N50, sequence length and number of contigs). The evaluation report did not include an evaluation based on quality thresholds.

3 Results

3.1 Participation

Laboratories could participate either in the full EQA scheme or one part only (serotyping, virulence profile or molecular typing-based cluster analysis). Of the 27 participants who signed up, 25 completed and submitted their results. The majority of participants (76%; 19/25) completed the EQA with analysis in each of the three parts. In total, 23 (92%) participants participated in serotyping, all participated in the detection of one or more of the virulence genes and 18 (72%) in cluster analysis. (Table 2).

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

	Serotyping[1]	Virulence profile[2]	Cluster analysis[3]
Number of participants	23	25	18
% of participants	92*	100*	72*

[1]: O grouping and/or H typing
[2]: detection of at least one gene (aaiC, aggR, eae, stx1 and stx2) and/or subtyping of stx1 *and* stx2
[3]: molecular typing-based cluster analyses based on PFGE or WGS-derived data
**: percentage of the total number (25) of participating laboratories.*

O grouping results were provided by 23 participants (92%) and H typing only provided by 15 (60%). The majority 13/23 (57%) used phenotypic method (Annex 3). All participants (100%) participated in the detection of virulence genes (*eae* and *stx1* and *stx2*), while most (80%, 20/25) also participated in detection of enteroaggregative genes *aaiC* and *aggR* as well as the *stx* subtyping. Most participants (62%, 11/18) reported cluster analysis using WGS-derived data, while nine (50%) reported using PFGE data. Two 18 participants (11%) submitted cluster data based on both PFGE and WGS (Table 3).

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

	Serotyping n=23		Virulence profile n=25					Cluster analysis n=18		
	O group	H type	*aaiC*	*aggR*	*eae*	*stx1* and *stx2*	*stx* subtyping	PFGE	WGS	Both
Number of participants	23[#]	15	20	20	25	25	20	9	11	2
Percentage of participants^	100%	65%	80%	80%	100%	100%	80%	50%	61%	11%
Percentage of participants *	92%	60%	80%	80%	100%	100%	80%	36%	44%	8%

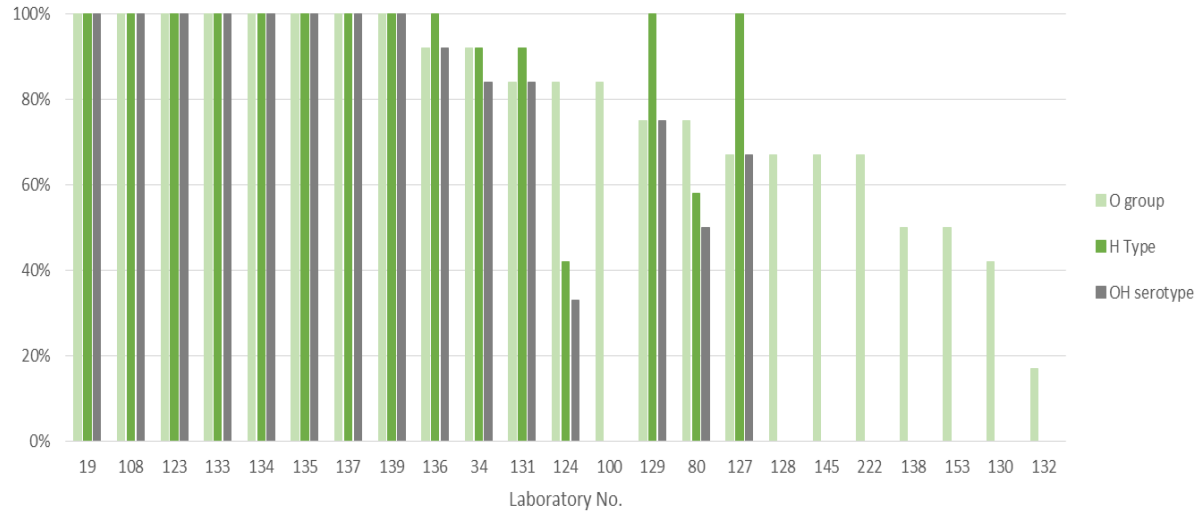
^: percentage of participants in respective part of EQA
**: percentage of total number of participating laboratories (25)*
[#]: phenotypic (n=13)/PCR-based (N=4)/WGS-based (n=6).

3.2 Serotyping

Twenty-three (92%) laboratories performed O grouping and eight (35%) of the 23 were able to type all 12 test isolates correctly, giving an average score of 79% (Figure 1). Eight laboratories (35%) reported the correct O group for the rare O group O187 (isolate REF3) and 13 correctly reported O154 (isolate REF10) (Figure 2). The highest performances were displayed for the two O157 isolates (100%), O91, O104, O111 and O121 (87%; Figure 2), some included in the minimum requirements of ECDC [14]. One laboratory detected O157 only, generating incorrect (non-O157) results for the 10 other isolates (Annex 3). Fifteen (65%) laboratories reported an incorrect O group for one or more isolates, primarily by reporting 'non-typable' (NT) or 'not done' (ND: 78%; 46/59, Annex 3). In total, the average score was 79% (Figure 1).

Fifteen (60%) laboratories performed H typing, while 65% of participants performed O grouping. The general performance for H typing was higher than O grouping, with the majority (60%; 9/15) of participants correctly H typing all 10 test isolates, resulting in an average score of 92% (Figure 1). Only two laboratories reported an incorrect H type, while the rest obtained incorrect H type results by reporting NT, H or ND results (86%; 12/14; Annex 3).

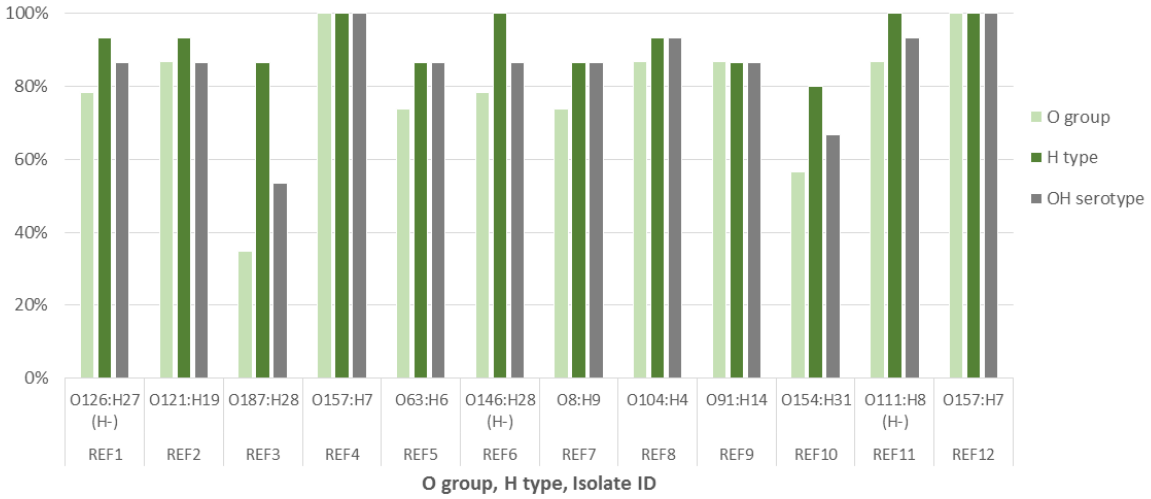
Figure 1. Participant percentage scores for O grouping and H typing



Arbitrary numbers represent participating laboratories.
Bars represent number of correctly assigned O groups (light green), n=23, H types (dark green), n=15 and combined O:H serotype (grey), n=15.

Complete O:H serotyping was performed by 15 (65%) participants with an average score of 86%, ranging from 53% (8/15) for O187:H28 to 100% (15/15) for isolate O157:H7 of the participants reporting the correct serotype (Figure 2).

Figure 2. Average percentage test isolate score for serotyping of O and H



Bars represent the percentage of laboratories correctly assigning O groups (light green): n=23.
H types (dark green): n=15.
Combined O:H serotypes (grey): n=15.
Average scores: O group, 79%; H type, 92% and combined O:H serotype, 86%.

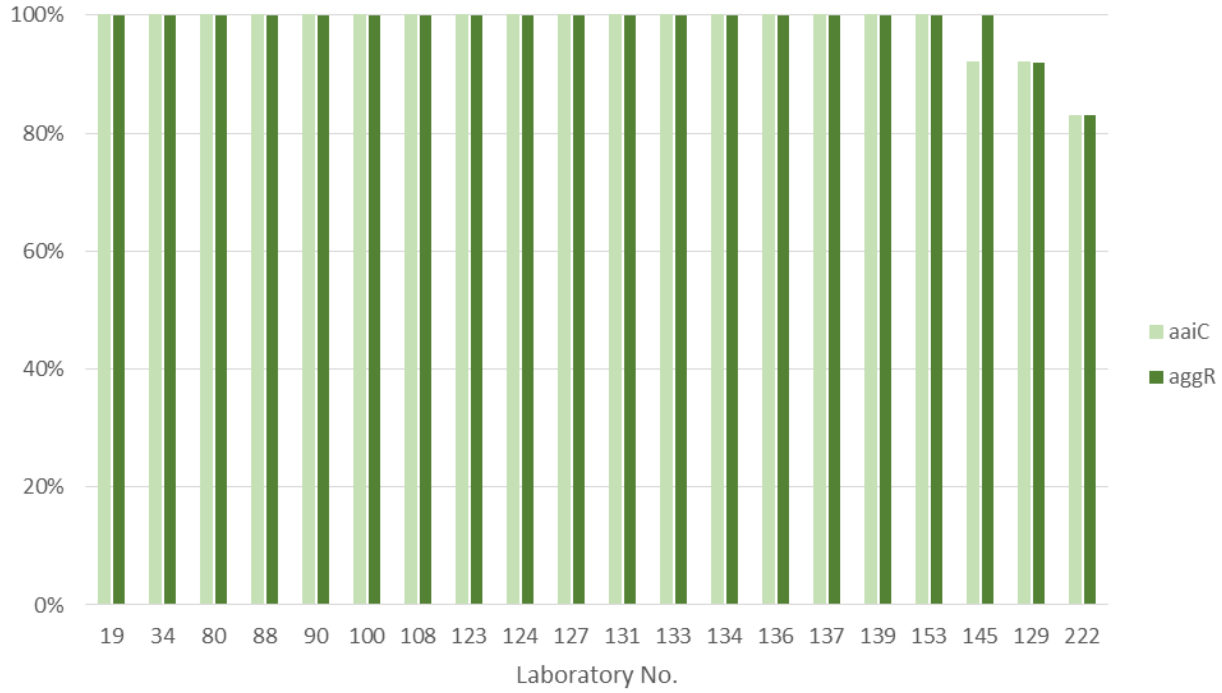
3.3 Virulence profile

Between 20–25 laboratories submitted results for each of the virulence genes, consisting of detection of EAEC *(aaiC* and *aggR*) and virulence genes (*eae*, *stx1* and *stx2*) and subtyping of *stx1* and *stx2* genes. All 25 participants submitted results for *eae* and *stx* genes. Twenty-five laboratories (100%) submitted subtyping results of *stx1* and *stx2* genes and 20 (80%) reported results for EAEC genes *aaiC* and *aggR*.

3.3.1 Detection of EAEC genes (*aaiC* and *aggR*)

The performance of the 20 laboratories reporting genotyping results for EAEC, *aaiC* (85%; 17/20) and *aggR* (90%; 18/20) was high (Figure 2). One laboratory was responsible for four of the seven errors. Further analysis of results from this laboratory revealed the swapping of isolates REF3 and REF8. One laboratory failed only to detect *aaiC,* but reported *aggR* correctly. The average scores were 98% and 99% respectively (Annex 4).

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

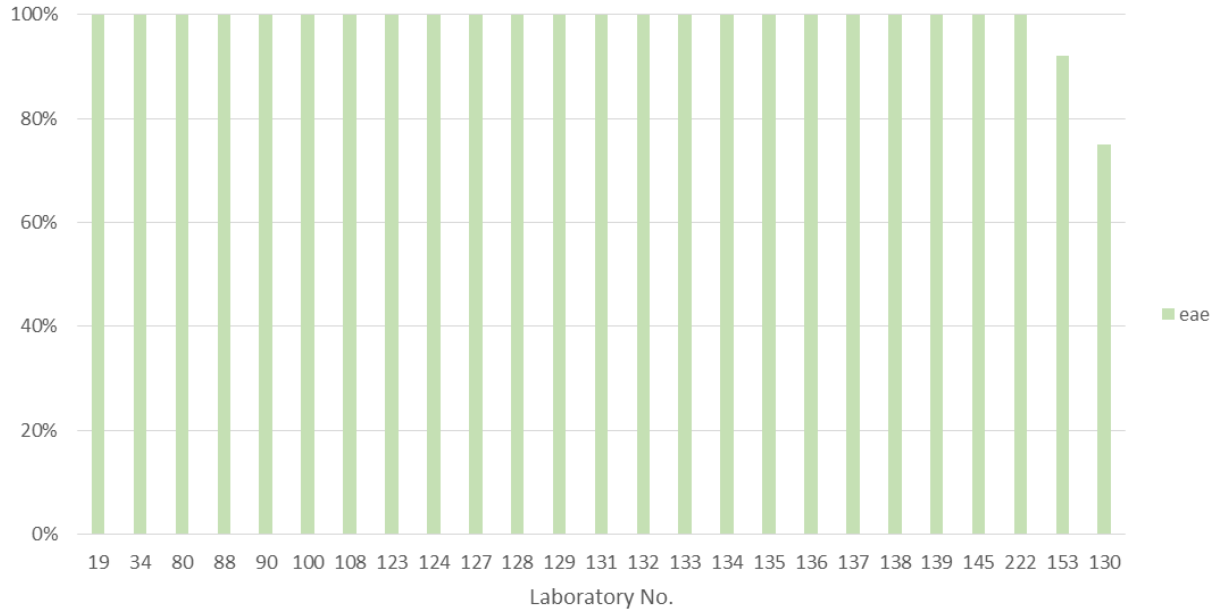


Arbitrary numbers represent participating laboratories.
Bars represent number of correct genotyping of aaiC *(light green) and* aggR *(dark green): n=20.*

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

Detection of virulence genes *eae*, *stx1* and *stx2* was performed by 25 (100%) laboratories with a generally high performance (Figures 4–5). For *eae* detection, 23 (92%) laboratories obtained a 100% score (Figure 4). Two laboratories (130 and 153) reported incorrect *eae* results. In total, *eae* was only misidentified four times: three false negative results were reported for isolate REF2, REF5 and REF11 and one false positive for isolate REF10 (Annex 4).

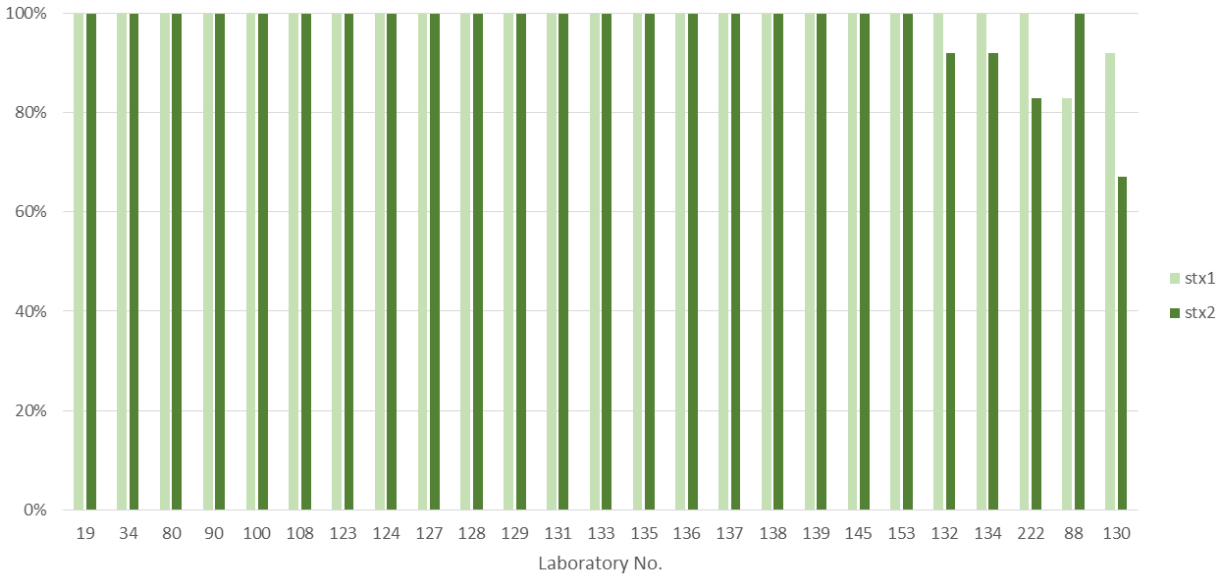
Figure 4. Participant percentage scores for genotyping of *eae*



Arbitrary numbers represent participating laboratories.
Bars represent number of correct genotyping of eae *(light green): n=25.*

The performance of detection of *stx*1 and *stx*2 genes was high; 23 (92%) laboratories reported 100% correct *stx1* results and 20 (80%) laboratories reported 100% correct *stx2* results (Figure 5). All of the three incorrect *stx1* and the eight incorrect *stx2* results were reported by five laboratories. Most were false negatives and two laboratories had false positive results in REF8 (Annex 4).

Figure 5. Participant percentage scores for detection of *stx1* and *stx2*



Arbitrary numbers represent participating laboratories.
Bars represent number of correct genotyping of stx1 *(light green) and* stx2 *(dark green): n=25.*
Average scores: stx, *99%;* stx2, *98%.*

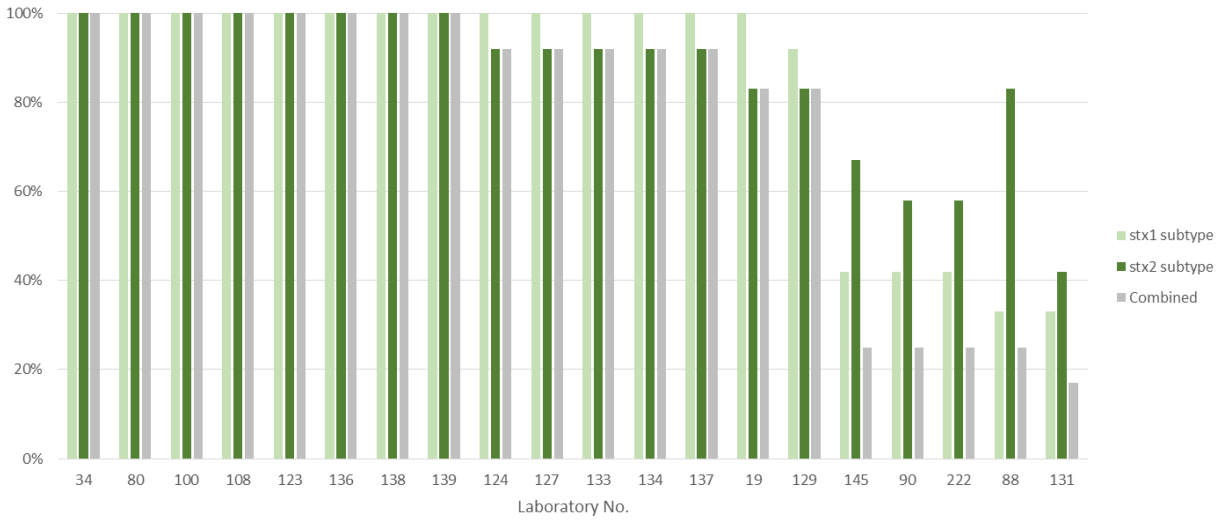
3.3.3 Subtyping of stx1 and stx2

Subtyping of *stx1* and *stx2* was performed by 20 laboratories. Fourteen (70%; 14/20) subtyped *stx1* correctly and eight (40%; 8/20) reported correct *stx2* subtype for all ten test isolates (Figure 6). In addition to the 20 laboratories performing complete subtyping of all *stx1* and *stx2* subtypes, two additional participants (laboratory 128 and 135) performed correct detection of *stx2f* only, these results were not included in the analysis (Annex 4).

Six laboratories (30%) reported an incorrect subtype of *stx1* for one or more isolates, primarily by reporting ND (not done) instead of correctly tested negative (92%; 35/38, Annex 4). Laboratories were not allowed to only report results for selected test isolates for a particular test, so reporting an ND was considered an incorrect result if the laboratory reported results of other isolates for that test. In total, the average score was 84% (Figure 6). The true mis-subtyping *stx1* results (n=3) were reported by two laboratories (Annex 4). Laboratory 88 incorrectly reported *stx1a* for a negative isolate (REF9) and vice versa for another test isolate (REF8). Laboratory 131 incorrectly reported *stx1c and stx1d* for the only *stx1d* positive isolate (REF10).

Twelve laboratories (60%) reported an incorrect subtyping of *stx2* for one or more isolates, primarily by reporting ND instead of negative (59%; 19/32, Annex 4). In total, the average score was 87% (Figure 6). The true number of instances of mis-subtyping *stx2* was 13, of which 6 consisted of reporting either *stx2a* or *stx2d* instead of both *stx2a* and *stx2d* or *stx2e* if using WGS (REF7). Laboratory 222 had clearly swapped the results of the isolates REF3 and REF8, not reporting *stx2g* and *stx2* negative, but the opposite, also seen in the serotype results. Two laboratories (19 and 131) incorrectly reported *stx2c* or *stx2a, stx2c* or *stx2d* for the isolate REF4 (*stx2a* and *stx2c* positive). The additional incorrectly determined *stx2* results were reported by laboratory 131 for REF2 and REF12. Both reported *stx2a* and *stx2c*, but the correct result was only *stx2a*.

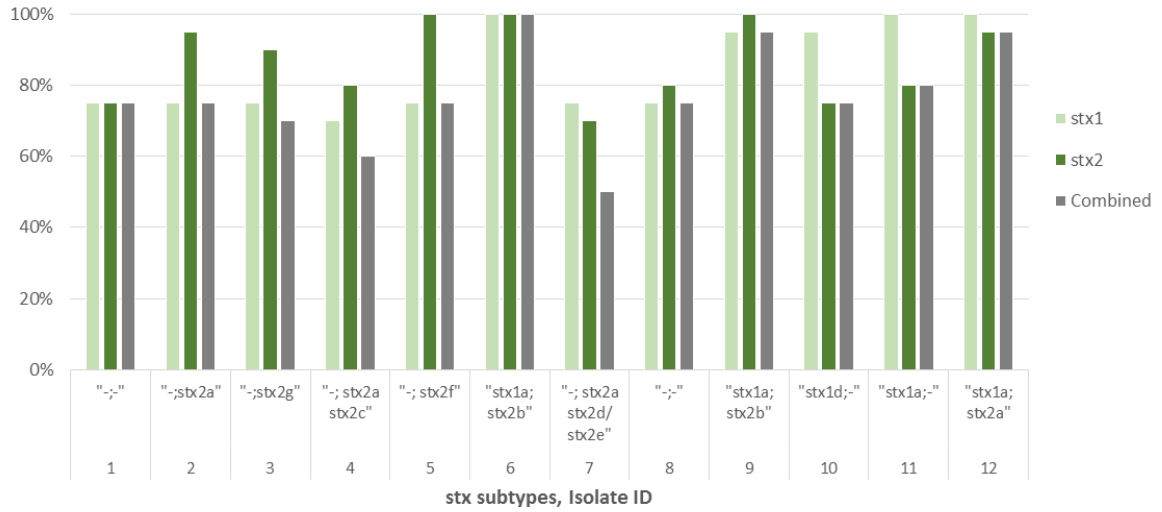
Figure 6. Participant percentage scores for subtyping of *stx1* and *stx2*



Arbitrary numbers represent participating laboratories.
Bars represent number of correct subtyping of stx1 (light green), stx2 (dark green), combined stx1 and stx2 (grey), n=20.
Reporting ND (not done) evaluated as incorrect.

Only 40% (8/20) of the participants were able to correctly *stx2* subtype all 10 test isolates (Figure 6). Among the *stx2* subtypes, *stx2f, stx2b and stx2a* generated the highest scores and *stx2a, stx2d* (*stx2e* if using WGS) generated the lowest (Figure 7).

Figure 7. Average percentage test isolate 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=20
Average scores: stx1, 84%; stx2, 87% and combined stx1 and stx2, 77%.

Incorrect subtype results were reported 70 times, the majority of which (54/70) were due to reporting ND instead of negative. The incorrect results of *stx2* subtyping shown in Table 4 are divided in three categories:; false negatives (4/32), incorrect subtype of *stx2* (11/32) or ND (19/32). Laboratory 131 reported three of the incorrect *stx2* subtypes and four of the ND results. Laboratory 129 reported two of the false negative *stx2* subtyping results, but had no errors in the determination of *stx2*.

Table 4. Incorrect *stx2* subtype results

		Incorrect subtype results			
Isolate ID	**EQA provider**	**False negative**	**Incorrect**	**Total true errors**	**Errors by reporting ND#**
REF1	-				5
REF2	*stx2a*		*stx2a + stx2c* (1)	1	
REF3	*stx2g*	2 (ND)		2	
REF4	*stx2a stx2c*	1 (ND) 1	*stx2c* (1) *stx2a + stx2c + stx2d* (1)	4	
REF5	*stx2f*				
REF6	*stx2b*				
REF7	*stx2a stx2d/ stx2e*		*stx2a* (5) *stx2d* (1)	6	
REF8	-		*stx2g* (1)	1	3
REF9	*stx2b*				
REF10	-				5
REF11					4
REF12	*stx2a*		*stx2a + stx2c* (1)	1	
Total				**15**	**17**

ND#: not done.

3.4 Molecular typing-based cluster analysis

In this part of the EQA, participants should have correctly identified a cluster of closely related isolates among 12 test isolates by using either PFGE and/or WGS-derived data. The cluster test isolates were pre-categorised by the EQA provider.

The EQA provider's PFGE results were based on an *XbaI* profile. The EQA provider's cluster analysis of WGS-derived data was based on allele-based (cgMLST [15]) and SNP analysis (NASP [13]). The correct number of closely related isolates (4) could be identified by both PFGE and WGS-derived data. The cluster contained four O26:H11 (*stx1a*), ST21 isolates: REF14, REF15, REF19 and REF23 (REF19 and REF23 were technical duplicates). The EQA provider found at most two allele differences or four SNPs between any two isolates in the cluster. The rest of the cluster test isolates were an additional five ST21s, one ST16, one ST17 and one ST29 (Annexes 5–13).

3.4.1 PFGE-derived data

Of the 25 participants in the EQA, nine (36%) performed cluster analysis using PFGE-derived data. Six (67%) correctly identified the cluster of closely related isolates defined by a pre-categorisation from the EQA provider among the 12 cluster test isolates. Table 5 shows the overview of the isolate each participant included or excluded in cluster identification. Laboratory 100 reported only the two technical duplicates to be a cluster. Laboratory 124 included one additional isolate, REF21, as part of the cluster of closely related isolates, accepting a profile with a two band differences. Laboratory 132 included eight of the nine ST21s into the cluster, including several band differences instead of only four correct isolates.

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

		Laboratory number								
Isolate ID	ST	19	90	100	123	124	127	132	136	222
REF13	21	No	No	No	No	No	No	Yes	No	No
REF14‡	21	Yes	Yes	No	Yes	Yes	Yes	Yes	Yes	Yes
REF15‡	21	Yes	Yes	No	Yes	Yes	Yes	Yes	Yes	Yes
REF16	21	No	No	No	No	No	No	Yes	No	No
REF17	21	No	No	No	No	No	No	Yes	No	No
REF18	17	No	No	No	No	No	No	No	No	No
REF19#‡	21	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
REF20	21	No	No	No	No	No	No	No	No	No
REF21	21	No	No	No	No	Yes	No	Yes	No	No
REF22	16	No	No	No	No	No	No	No	No	No
REF23#‡	21	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
REF24	29	No	No	No	No	No	No	No	No	No
Cluster-identified conclusion		Yes	Yes	No	Yes	No	Yes	No	Yes	Yes

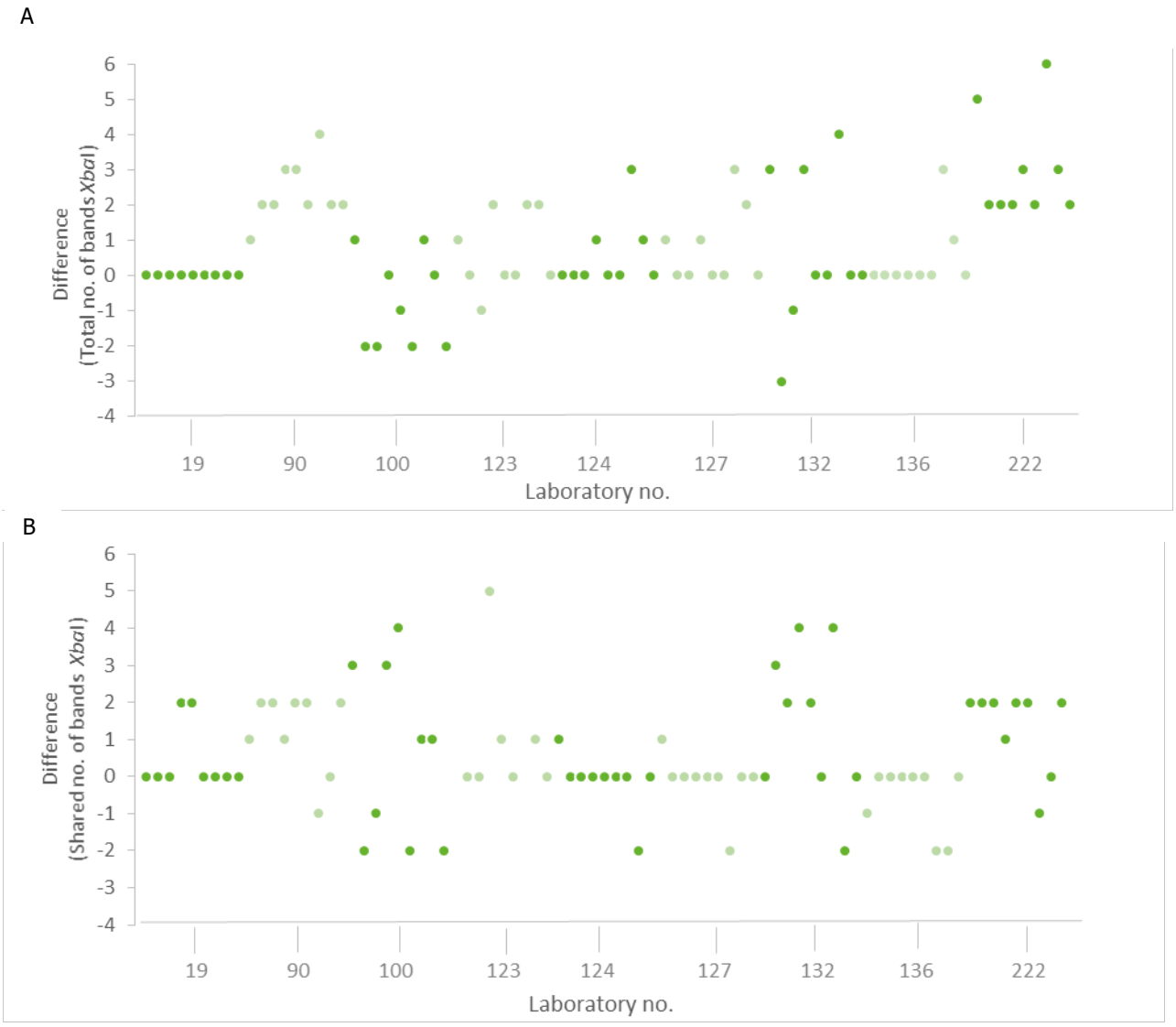
‡: closely related isolates
#: technical duplicate isolates (Annex 7).

For each isolate, participants were instructed to report the total number of bands in the *Xba*I profile. The number of bands shared between each test isolate and the selected cluster representative was also reported (Figure 8, Annexes 7–8).

Figure 8A shows the difference between the number of bands reported by the participants and the number observed by the EQA provider for *XbaI*. The PFGE profile of *E. coli* contains a large number of bands within the region of 200–350 kb, which make the cluster analysis based on PFGE harder to interpret compared with other species such as *Salmonella* or *Listeria*. This is illustrated in Figure 8, which shows a high number of variations. Five of the nine participants identified 19 bands in the profile of the cluster isolates, but laboratory 132 reported 22 and 20 bands in REF14 and REF15 respectively instead of 19 bands. Laboratories 90 and 222 only counted 17 bands in the cluster isolates, which also can be seen in Figure 8A, as they did not have any numbers of bands that corresponded to the EQA provider's results.

Figure 8B shows the difference between the participants' reported number of shared bands with a selected cluster representative and the number observed by the EQA provider for *Xba*I. Laboratories 90, 100, 132 and 222 produced several results with 2 or more band differences from the expected result, while laboratories, 100 and 132 did not identify the correct cluster.

Figure 8. Difference between reported number of bands (A) and shared bands (B) for each isolate to selected isolate



Data from all nine O26:H11, ST21 isolates: REF13, REF14, REF15, REF16, REF17, REF19, REF20, REF21 and REF23. Laboratory 123 only reported data for seven isolates (Annexes 7–8).

3.4.2 WGS-derived data

3.4.2.1 Reported results from participants

Eleven participants (44%) performed cluster analysis using WGS-derived data. Only one laboratory reported using external assistance for sequencing. Different sequencing platforms were listed among the participants: 1 MiniSeq, 5 MiSeq, 2 HiSeq, 2 NextSeq and 1 Ion Torrent. All reported using commercial kits for library preparation. Out of the 11 participants, nine (82%) used Illumina's Nextera kit. Four participants reported volume changes from the manufacturer protocol (Annex 9).

Performance was high in cluster analysis with WGS-derived data. Ten participants (91%) correctly identified the cluster of closely related isolates defined by pre-categorisation from the EQA provider among the 12 test isolates (Table 6). One laboratory included all ST21 and the ST29 isolates as being in the cluster of closely related isolates. After the individual evaluation reports were distributed, the laboratory explained that the question of listing the closely related isolates was misinterpreted as reporting which isolates were analysed.

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

		Laboratory number										
Isolate ID	**ST**	**19**	**34**	**80**	**108**	**123**	**129**	**133**	**134**	**135**	**137**	**139**
REF13	21	No	No	No	No	No	No	No	No	No	No	Yes
REF14‡	21	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
REF15‡	21	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
REF16	21	No	No	No	No	No	No	No	No	No	No	Yes
REF17	21	No	No	No	No	No	No	No	No	No	No	Yes
REF18	17	No	No	No	No	No	No	No	No	No	No	No
REF19‡#	21	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
REF20	21	No	No	No	No	No	No	No	No	No	No	Yes
REF21	21	No	No	No	No	No	No	No	No	No	No	Yes
REF22	16	No	No	No	No	No	No	No	No	No	No	No
REF23‡#	21	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
REF24	29	No	No	No	No	No	No	No	No	No	No	Yes
Main analysis		Allele (cgMLST)	Allele (cgMLST)	Allele (cgMLST)	SNP	Allele (cgMLST)	Allele (cgMLST)	Allele (wgMLST)	Allele (cgMLST)	Allele (wgMLST)	SNP	SNP
Additional analysis		SNP	SNP									
Identified cluster		Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	No

‡: closely related isolates
#: technical duplicate isolates
ST: sequence type
Allele: allele-based analysis.
SNP: single-nucleotide polymorphism (Annex 10).

Laboratories were instructed to report the data analysis used for cluster identification and select a representative isolate in the cluster for reporting SNP distance or allelic differences between the selected isolate and each test isolate included in analysis. Laboratories could report results from up to three analyses (1 main and 0–2 additional), but the detected cluster had to be based on results from the main analysis.

Out of the five participants using SNP, only three used SNP as the main analysis for cluster detection, while two reported SNP as an additional analysis. All five used a reference-based approach with different ST21 isolates as reference. Three used Burrows-Wheeler Aligner (BWA), one used CLC and one used Bowtie2 as the read mapper, but different variant callers were used (Table 7).

Table 7. Results of SNP-based cluster analysis

Lab	**SNP-based**					
	Approach	**Reference**	**Read mapper**	**Variant caller**	**Distance within cluster**	**Distance outside cluster**
Provider	Reference-based	ST21 (REFXX)	BWA	GATK	0–3	256–444
19*	Reference-based	ST21_9523	BWA	GATK	0–4	238–374
34*	Reference-based	NC_013361 O26:H11 str11368	BWA	VarScan (vers.2.3)	0–4	35-71 (241–1194)
108	Reference-based	In-house strain in resp ST	CLC assembly cell v4.4.2	CLC assembly cell v4.4.2	0–5	321–397
137	Reference-based	CC21 ST21, ST29 11368_O26 CC17 ST17 12009 CC11 ST11 Sakai CC23 ST23 CP004009 CC122 ST583 ERR1010184 CC33 ST33 177412_H15400864401-1	BWA v0.7.12	GATK v2.6.5	0–5	294–548
139	Reference-based	O26_NC_013361	Bowtie2	SAMtools	0–988	None reported

**: additional analysis*
¤: reported distance to ST21 (non-ST21) isolates (Annex 11).

All eight participants that used allele-based analysis selected this method as the main analysis for cluster detection (Table 7). Four of eight (50%) used an assembly-based allele calling method, one used a mapping-based method and three laboratories used both mapping- and assembly-based allele calling (Table 8).

Table 8. Results of allele-based cluster analysis

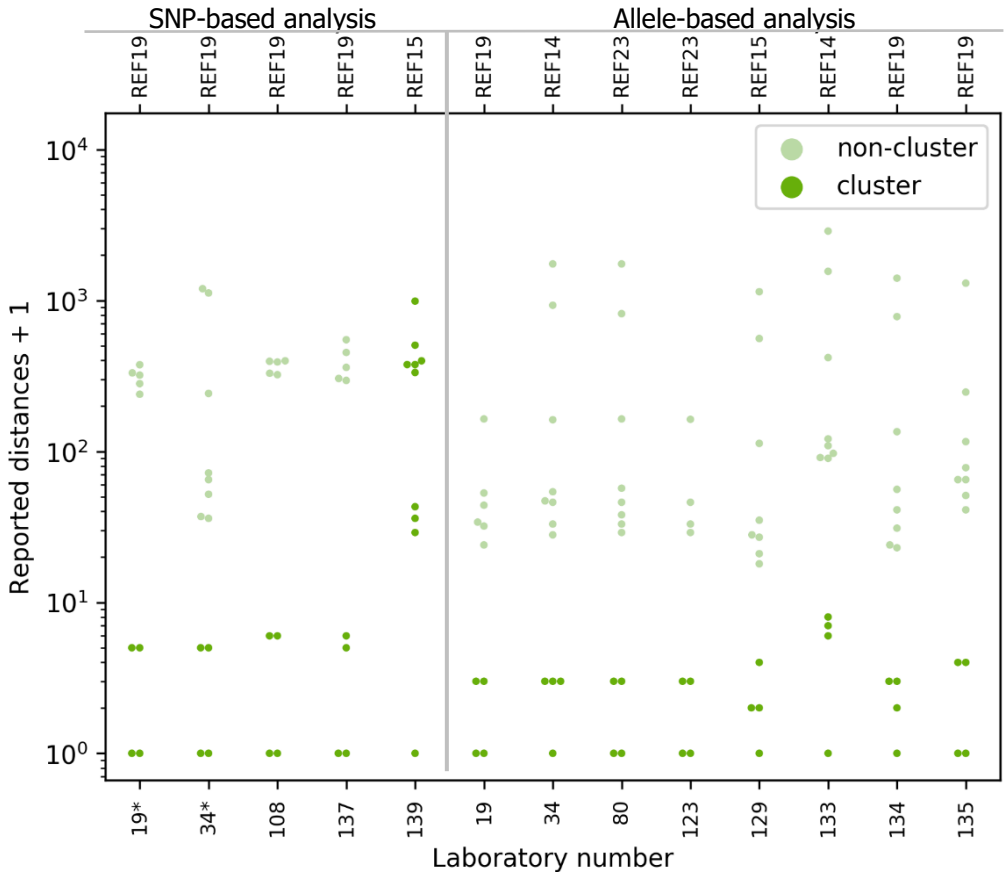
Laboratory	Allele-based analysis					
	Approach	Allelic calling method	Assembler	Scheme	Difference within cluster	Difference outside cluster¤
Provider	BioNumerics	Assembly- and mapping-based	SPAdes	Applied Maths (cgMLST/Enterobase)	0–2	29–54 (163–1745)
19	BioNumerics	Assembly- and mapping-based	SPAdes	Applied Maths (cgMLST/Enterobase)	0–2	23–52 (163)
34	SeqPhere	Mapping-based only	-	Enterobase (cgMLST)	0–2	27–53 (161–1744)
80	SeqPhere	Assembly-based only	Velvet	Enterobase (cgMLST)	0–2	28–56 (163–1744)
123	SeqPhere	Assembly-based only	SPAdes	Enterobase (cgMLST)	0–2	28–45 §(162)
129	SeqPhere	Assembly-based only	Velvet	cgMLST:The Ridom SeqSphere+ software's target definer was utilised to identify 1802 target loci shared by the reference strain NC_000913.3 (*Escherichia coli* strain K–12) and 38 additional complete query genomes obtained from GenBank. Default filter setting	0–3	17–34 (112–1141)
133	BioNumerics	Assembly- and mapping-based	SPAdes	Applied Maths (wgMLST)	0–7	89–120 (417–2878)
134	SeqPhere	Assembly- and mapping-based	Velvet	cgMLST defined in house of 2 128 targets	0–2	22–55 (134–1403)
135	SeqPhere	Assembly-based only	CLC Genomics Workbench	Resulting targets: 3 199 targets were defined for cgMLST (3 122 565 bases), 1 304 targets were used as accessory targets (1 113 123 bases) Reference Genome: CP000247.1, 4938920 bases, 4 685 genes (*Escherichia coli* 536, complete genome)	0–3	40-115 (64–1300)

¤: reported differences to ST21 (non-ST21; Annex 11)
§: difference not reported for REF16 and 17.

All eight laboratories using allele-based methods identified the correct cluster of four closely related isolates (Table 8–9). Four laboratories, 19, 34, 80 and 123, performed cgMLST using the same scheme as the EQA provider (cgMLST/Enterobase [12]) and two, 129 and 134, used an in-house cgMLST scheme. Two additional laboratories performed wgMLST; laboratory 133 used the wgMLST scheme by Applied Maths and laboratory 135 used an in-house wgMLST scheme. Seven laboratories reported allele differences of 0–3 within the cluster and laboratory 133 using wgMLST reported an allele difference within the cluster of 7 at most (Figure 9, Table 8). The allele differences reported depended on the isolate selected as cluster representative. Five laboratories selected REF19 or REF23 (technical duplicates), two used REF14 and one used REF15 (Figure 9).

Five other test isolates (REF13, REF16, REF17, REF20, and REF21) were also ST21, but not pre-defined by the EQA provider as part of the cluster. Based on cgMLST, six laboratories reported allele differences to the selected cluster isolate at 17–56 for this group of isolates and based on wgMLST, laboratories 133 and 135 reported allele differences at 40–120. Only six of eight laboratories reported results for the three non-ST21 isolates (REF18, REF22 and REF24). Based on cgMLST, the reported differences were 112–1744 and 417–2878 based on wgMLST by laboratory 133 (Table 8, Annex 11). Laboratory 135 reported 64–1300 by wgMLST, caused by a very low allele difference for REF 18. Evaluation of the laboratory's raw data in BN using wgMLST scheme disclosed a similar number of allelic differences as the EQA provider's data (data now shown), which suggests a reporting error by the laboratory (Annex 11).

Figure 9. Reported SNP distances or allelic differences for each test isolate to selected cluster representative isolate



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

Of the five laboratories performing SNP analysis (three as main analysis and two as additional), laboratories 19, 34, 108 and 137 identified the correct cluster of closely related isolates and reported SNP distances within the cluster from 0–5 (Figure 9). Laboratory 139 performed only SNP analysis and could not identify the correct cluster, but reported all ST21 and the ST29 isolates as part of the cluster of closely related isolates with a SNP difference within the cluster from 0–988. After the individual evaluation report was distributed, the laboratory explained it had misinterpreted the question of closely related isolates and instead reported all isolates analysed by the SNP analysis. However, the reported SNP distances were still higher than expected (0–42 within the correct cluster).

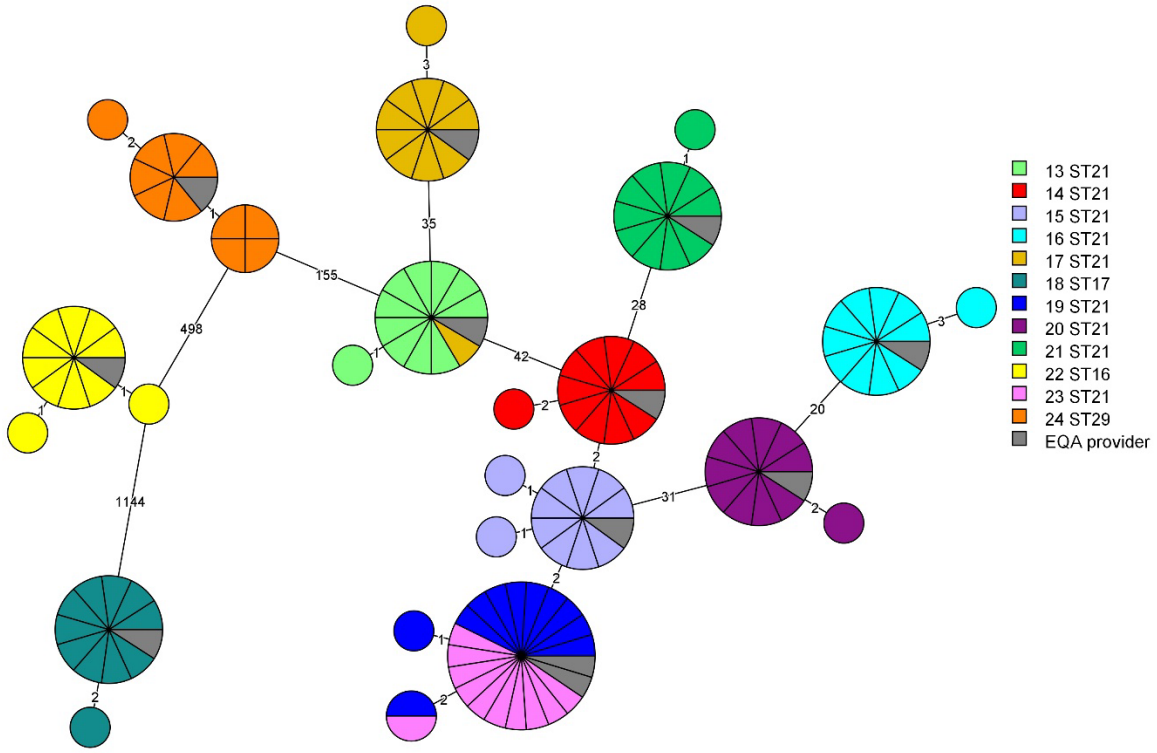
The SNP distances reported depended on the isolate selected as cluster representative, but four laboratories selected REF19 (technical duplicate) and one used REF15 (Figure 9).

3.4.2.2 Analysis of raw reads uploaded by participants

In addition to the reported cluster identification, participants submitted their FASTQ files to be evaluated by the EQA provider. The FASTQ files were uploaded to an Applied Maths calculation engine for allele calling (Enterobase) [12] and evaluated by the EQA provider's in-house quality control (QC) pipeline [15].

The overall cgMLST analysis, shown in the minimum spanning tree (MST) based on submitted raw reads (FASTQ files) from 11 laboratories reveals clear clustering of the results for each test isolate (Figure 10). Laboratory 34 appears to have sequenced the same isolate twice (REF13), as it reported a second cluster of isolates corresponding to REF13 and REF17, which also can be seen in the analysis of raw reads from laboratory 34 (Figure 10). No isolates were discarded from the analyses in this report due to low quality.

Figure 10. Minimum spanning tree of core genome multilocus sequence typing participant FASTQ files

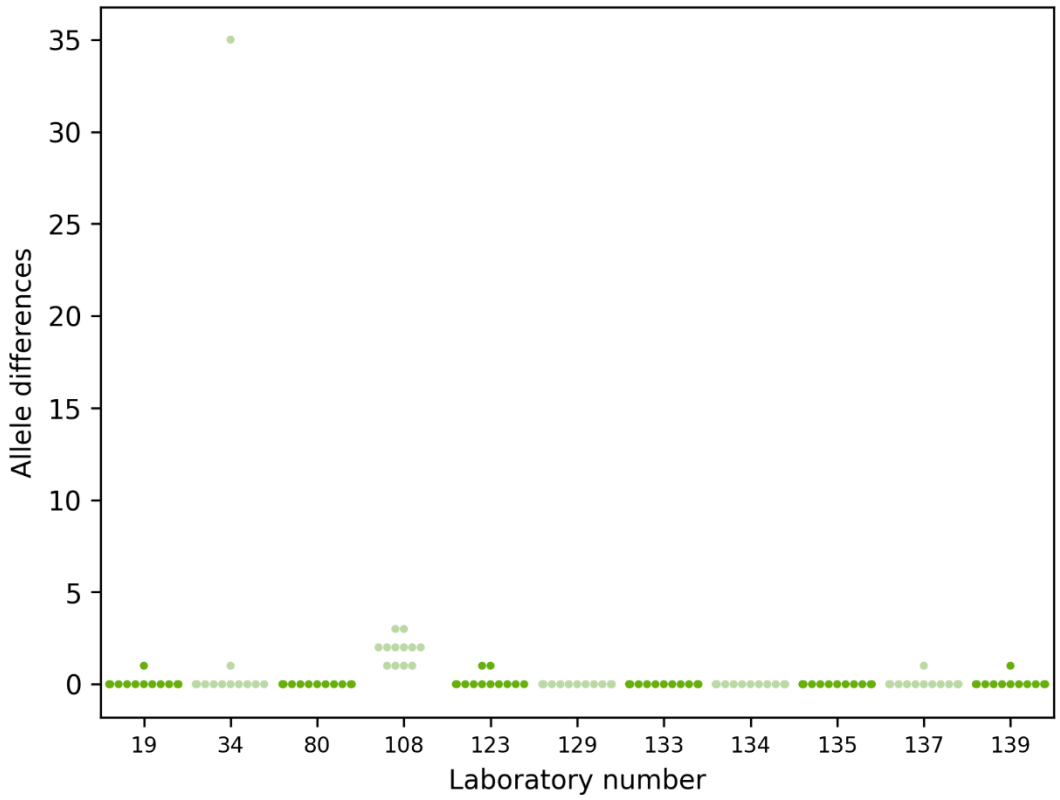


Minimum spanning tree (MST) in log scale of core genome multilocus sequence typing (cgMLST) [12] based on submitted raw reads (FASTQ files).
Each of the REF1–12 test isolates have a different colour. REF results from the EQA-provider are in grey.
An update and error in allele calling in the BioNumerics resulted in a spurious single allele difference in locus ECOLI0001 for ST17, 21 and 29 isolates being reported in the individual reports. The sequences of the reference isolates have been reanalysed for this report and the error does not impact the displayed figures. However, participant laboratories should note that this results in a slight discrepancy between the data in their individual reports and the aggregated figures shown.

The allele differences in Figure 10 do not exactly match those illustrated in the individual reports and consequently those in Figure 11, where the same data are used. This discrepancy is caused by loci being dropped if they did not pass QC for all isolates in the analysis. Joint analysis accordingly contains fewer loci.

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

Figure 11. Participant allele difference from reference result (EQA-provider) for each test isolate



Allele difference from corresponding REF isolates (EQA provider) based on submitted raw reads (FASTQ files) and analysed by EQA provider.

For 113 of 132 results (86%), no allele difference was identified. For 10 results (8%) a difference of 1 allele from the REF isolate was calculated and for eight results (6%), a difference of 2–3 alleles was observed, all by laboratory 108. Laboratory 34 identified a difference of 35 alleles, probably caused by the same isolate (REF13) wrongly being sequenced twice.

Separately, the laboratories listed quantitative and qualitative QC parameters used to evaluate their data. As seen in Table 9, coverage was the most widely used QC parameter, with acceptance thresholds ranging from 20–70X coverage. CgMLST locus coverage was widely used, as was a genus/species confirmation or contamination check and a check of the number of contigs as a measure of assembly quality. Refer to the full list of QC parameters reported by the participants in Annex 12.

Table 9. Summary of quantitative and qualitative parameters reported by participants

Parameters	Number of laboratories
Coverage	8
cgMLST coverage	6
Taxonomic confirmation	5
# contigs	4
Assembly length	3
Contamination check	3
Per-site minimum depth	3

For each laboratory, the submitted raw reads (FASTQ files) were evaluated by the EQA provider's in-house quality control pipeline [15]. For the full QC evaluation of all isolates, see Annex 13.

According to the QC parameters, sequencing quality was uniformly good. Coverage was high overall. Laboratories 134 and 137 sequenced to a lower depth than the other laboratories. There are no instances of severe contamination and no isolates were discarded from the analyses in this report due to low quality.

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

		Laboratory number										
Parameters	Ranges*	19	34	80	108	123	129	133	134	135	137	139
Number of genera detected	{1}	1	1	1	1	1	1	1	1	1	1	1
Detected species	{*Ec*}	*Ec*	*Ec*	*Ec*	*Ec*	*Ec*	*Ec*	*Ec*	*Ec*	*Ec*	*Ec*	*Ec*
Unclassified reads (%)	{<100}	2.55–8.21	1.27–8.01	1.4–4.87	1.18–2.6	1.31–7.11	0.62–3.42	1.05–5.76	1.11–2.25	1.68–7.05	1.2–12.07	0.51–1.57
Length at 25 x minimum coverage (Mbp)	{>45 ∧ <53}	5–5.4	5.2–5.6	5.2–5.6	5.1–5.5	5.2–5.5	5–5.4	5.2–5.5	2.9–5.5	5.1–5.4	0.6–5.3	4.9–5.3
Length [0–25] x minimum coverage (kbp)	{<250}	1.1–62.6	0–39.5	0–43.1	0–166.2	0–123.9	1.6–117.2	0–111.3	0–2415.3	0–95.3	0–4860.8	41.2–313
Number of contigs at 25 x minimum coverage	{>0}	200–527	166–298	207–628	229–435	183–313	183–275	152–244	146–234	150–226	90–212	201–293
Number of contigs [0–25] x minimum coverage	{<1000}	1–33	0–20	0–15	0–27	0–48	1–21	0–29	0–114	0–16	0–413	24–84
Average coverage	{>50}	51–157.4	37.3–108.2	86–146.9	37–290.5	49–73.5	78.2–169.3	53.5–163.7	28–79.7	63–247.5	22.9–84	58.3–110.9
Number of reads (x1000)		1166.8–3211.6	539.9–1486.1	1124.2–2025.2	571.8–5313.5	679.8–1010.1	1510.9–3280.7	653.5–2371.5	524.7–1482.2	1520.4–6241	644.6–2349.9	1184.9–2274.8
Number of trimmed reads (x1000)		1136.6–3179	507.2–1422.8	1049.8–1964.3	535.8–4887.6	651.3–985.9	1482.4–3235.5	636.8–2337.5	520.1–1471.1	1489.5–6058.4	644.6–2349.9	1174.6–2254.8
Maximum read length		151–151	301–301	301–301	302–307	301–301	151–151	301–301	151–151	126–126	101–101	151–151
Mean read length		123.4–139.8	174.6–228.4	200.7–225.2	212.2–224.9	204–219.8	144.6–146.8	195.6–230.6	146.2–147.2	119.3–121.5	98.9–99.7	147.2–148.2
Read insert size		208.6–327	239–350.3	247.9–323.2	-	266.2–305.1	330.7–479.5	261.4–340.2	359.3–427.2	323.3–371.9	334.3–434.2	342–414.1
Insert size StdDev		86.6–148.3	93.6–177	78.4–126.1	-	100.4–116.3	142.9–173.2	101.8–164.8	125.9–136.2	188.3–213.6	186.2–209.7	83.5–99.4
N50 (kbp)		16.8–73.3	52.9–114.3	14.6–114.2	28.3–76.2	45.4–92.7	47–95.5	92.2–126.8	45.4–124.9	76.9–104	34.4–115.5	33.2–60
N75 (kbp)		8.1–33.5	22.6–57.8	7.4–47.7	13.5–39.4	19.9–47.1	24.2–41	41.4–57.8	27.6–47.5	36.2–45.5	9.2–50.5	14.7–29.8

**: indicative QC range*
Ec: E. coli.

4 Discussion

4.1 Serotyping

Twenty-three (92%) laboratories participated in the serotyping part of the EQA-8, of which 13 participants (57%) provided phenotypic serotyping results and 10 (43%) provided molecular serotyping results (four by PCR and six by WGS). Fifteen participants performed both O group and H typing. Performance was high for both, with 80% and 92% correct O group identification and H typing respectively.

4.1.1 O group

Participation in O group typing was almost unchanged from EQA-4 through EQA-8, with 93% of participants performing O grouping (26/28; 26/29; 26/29; 27/30). A slight reduction in H typing participation from 18 in EQA-4 to 15 laboratories in EQA-8 was detected.

The performance of O grouping was low this year, mainly due to isolates O187 (REF3) and O154 (REF8) and the reporting of ND and NT as results. Eight participants (35%) reported the correct O group for all 12 test isolates. O group O187 is not so common in Europe and was reported as O74 by three laboratories and O103, O145 and O104 by three other laboratories. O187 has strong cross-reactions with both O74, O103 and O175. Furthermore, *E. coli* with the O184 group may react in an O187 antiserum. The incorrect O grouping of O187 as O74 or O103 can therefore be explained for four and possibly five of the participating laboratories (three O74 and one O103), but not for the O145 results. The O104 result from the fifth laboratory is suspected of representing the swapping of the O grouping results for isolates REF3 and REF8, where it reported O104 and O74 respectively. Four of these laboratories reported using phenotypic methodology and would have had to remove the false reactions against O74 and/or O103 by absorption. The fifth laboratory reported the use of PCR and it is of note that the published PCR method for O grouping will not detect O187 [16]). Reports of STEC O74 and O103 would therefore need further testing by reference laboratories capable of distinguishing between these O groups and reporting of STEC O187 from at least five PH NRLs may be underestimated. O154 cross-reacts with other O groups, but not with the incorrectly reported results (O26, O44, O55 and O109) in EQA-8.

Some of the more common O groups, also included in the minimum requirements of ECDC, generated the highest performances (O157: 100%, O111: 87% and O121: 87%). The average score was slightly higher in the current EQA (79%) compared to EQA-7 (69%).

4.1.2 H type

The general performance for H typing was higher than O grouping, with the majority of participants (73%, 11/15) correctly H typing all 10 test isolates (Figure 5). Compared to the previous EQA, the average score of 92% correct results was an improvement to previous EQA (81% in EQA-7).

4.1.2 OH serotyping

The O:H serotyping results ranged from 100% for isolates O157:H7 to only 53% (8/15) of the participants reporting correct serotype for O187:H28. The average percentage correct O:H serotyping score in this EQA was higher (86%) compared to EQA-7 (71%) and EQA-6 (78%).

In general, the less common European serotypes generated the lowest scores and vice versa. The performance of serotyping (O group/H type) is highly affected by the range of available antisera. Laboratories using a limited panel of antisera were encouraged to report serotype results as NT for isolates they were unable to type. The majority of incorrect serotype results (both O group and H type results) were reported as neither NT or ND and no systematic typing error was observed. Only one mistake of swapped isolates were identified by laboratory 222.

In addition to O grouping, H typing is crucial for outbreak detection, epidemiological surveillance, taxonomic differentiation of *E. coli* and detection of pathogenic serotypes. As such, it remains a main challenge to enable more PH NRLs to perform complete and reliable O:H serotyping, particular H typing.

4.2 Virulence profile

Twenty-five laboratories participated in the detection of the virulence profile with the participation rate and performance varying substantially between the different tests. As in previous EQAs, the participation rate was highest for the genotypic detection of the *stx* genes and detection of *eae* (100%) and lowest for the detection of *aaiC/aggR* and subtyping of stx genes (80%).

4.2.1 Detection of *aaiC* and *aggR*

The performance of detection of the two EAEC genes was high, with 85% and 90% of the participants respectively detecting *aaiC* and/or *aggR* correctly in the two EAEC isolates included in the EQA. One laboratory was responsible for four of the seven errors. Observing all the results submitted by this laboratory showed the results from isolate corresponding to REF3 and REF8 were swapped. The performance of detecting *aaiC* /*aggR* in EAEC isolates has been high through the four EQAs including an EAEC isolate (EQAs-4–8). In the present EQA, two EAEC isolates were included, which 85% (17/20) correctly identified.

4.2.2 Detection of *eae*

Genotyping of *eae* had a high participation rate (100%) and performance; 23 (92%) laboratories obtained a 100% score, giving an average score of 99%. The average correct score has been fairly unchanged through the EQAs (EQA-4, 96%; EQA-5, 98%; EQA-6, 97%; EQA-7, 98%).

4.2.3 Detection of *stx1* and *stx2*

Both the participation (100%) and detection rates were high for genotyping of *stx1* (99%) and *stx2* genes (98%), similar to previous EQAs. Notably, the majority of the incorrect results (5/11) originated from one laboratory.

4.2.4 Subtyping of *stx1* and *stx2*

The average scores of correct subtyping of *stx1* and *stx2* were 84% and 87% respectively, which is a slightly lower performance compared to the previous EQAs (EQA-4, 90%; EQA-5, 92%; EQA-6, 91%; EQA-7, 90%). Unlike the previous EQAs, the performance for *stx1* subtyping in EQA-8 was lower than for *stx2* (EQA-4: 94% vs. 93%; EQA-5: 98% vs. 92%; EQA-6: 100% vs. 91%; EQA-7: 99% vs. 90%). The obvious reason is the unexpected reporting of not done results, as some laboratories wrongly omitted performing the subtyping test on isolates already found negative in the initial screening (detection for *stx1* and *stx2*). This may be the routine procedure, but in this ECDC EQA, each test method is tested individually and irrespective of results obtained in the screening and detection or any other test.

When a participant signs up for a test and subsequently participates, all isolates must be tested using this test. The inconsistency in the number of performed tests per isolate and laboratory have been a recurrent problem throughout the EQAs so far.

In the current EQA, the true errors (not done results excluded) were three incorrect *stx1* subtyping results. Two errors were swapped results of REF3 and REF8. The last error was reporting of *stx1c* and *stx1d* in REF10 (*stx1d*). No isolates have ever been described with two copies of the *stx1* genes.

Of the true errors of incorrect *stx2* (n=13), six errors were reported for REF7 (*stx2a* and *stx2d* or *stx2e*) as either only *stx2a* (n=5) or only *stx2d* (n=1). At the time when the primers were designed for these three subtypes (*stx2a, stx2d* and *stx2e*), it was noted that the specific variant of *stx2e* designated *stx2e*-O8-FHI-1106-1092 had identical sequences with *stx2a* and differed only by the last nucleotide in the reverse primer sequence for *stx2d,* but also that the variant *stx2e*-O8-FHI-1106-1092 had not been described to be clinically relevant. Furthermore, it had been observed that the *stx2e* gene in the reference strain for stx_{2e}-O8-FHI-1106-1092 was often lost during the passing of cultures. It was therefore assumed that this variant was quite rare. However, the variant has since been reported from cases of diarrhoea in Norway, where it was originally identified, as well as five cases of diarrhoea in Denmark (unpublished), all in the same serotype O8:H9. The reference strain for stx_{2e}-O8-FHI-1106-1092 was O8:H2, which has not been isolated from patients with diarrhoea. It was therefore decided to use an O8:H9 strain in EQA-8. A provisional designation has been proposed [17] for a new Stx2 subtype, stx_{2h} (GenBank AM904726), but the sequence of stx_{2h} (AM904726) is identical to the variant stx_{2e}-O8-FHI-1106-1092 [18]. Accordingly giving this particular variant a new designation has to be considered and to design primers for its detection. This would require additional validation and sequence analyses beyond the scope of this report, but is under way.

All participants reported *stx2f* correctly in the present EQA, which was an improvement compared with EQA-7, when two laboratories incorrectly subtyped *stx2f*. The importance of awareness of *stx2f* has been described by Friesema et al. in 2014 [19] and Grande et al. in 2016 [20]. Routine detection of *stx2f* should therefore be included in the diagnostic repertoire of STEC in Europe. The fact that two additional participants performed only detection of *stx2f* was encouraging.

4.2.5 Additional virulence genes

An increasing number of STEC isolates have been reported to encode virulence genes from other pathotypes, recently reviewed by FAO and WHO in 2018 [21]. Most common are genes encoded by enterotoxigenic *E. coli* (ETEC). EQA-8 also included a isolate (REF3) representing such a hybrid between STEC and ETEC: ST200, O187:H28, *stx2g* and *estA*p, the porcine variant of the heat stable enterotoxin STp. Other virulence genes found in

this pathotype include long polar fimbriae (*lfpA*), the plasmid encoded catalase peroxidase *katP* and the enteroaggregative heat stable enterotoxin EAST1 encoded by *astA*. Up until September 2018, 17 Danish patients were found positive for this particular sero- and virulence type, primarily adult women aged 22–90 years, but also in four boys under 3 years of age. None of them developed HUS and only one patient reported bloody diarrhoea (unpublished). In view of the increasing number of laboratories performing WGS allowing for better identification of a whole set of virulence genes in pathogenic *E. coli*, it may be pertinent to consider including more virulence genes, such as genes found in ETEC, in future EQAs.

4.3 Molecular typing-based cluster analysis

The EQA scheme no longer covers PFGE as an independent part, but by adding cluster identification using either PFGE and/or WGS-derived data, this EQA is contemporary with the development of surveillance methods used by PH NRLs in Europe. This adjustment of the EQA appears to be well accepted by the Member States, as 18 of the 25 laboratories (72%) participated. Eleven participated in cluster identification using WGS-derived data, nine participated using PFGE-derived data and only two of the 18 laboratories participated in cluster identification using both methods. However, five laboratories participating in PFGE in EQA-7 did not participate in cluster identification in EQA-8 (two did not participate in EQA-8). This decrease in the number of participants could be caused by adding WGS or by removing PFGE as an independent part and no longer giving the laboratories an external quality assessment of their PFGE performance.

The present cluster designed by the EQA provider allowed the participants to detect the same number of closely related isolates by both PFGE and WGS.

4.3.1 PFGE-derived data

Of the 25 laboratories, nine (36%) performed cluster analysis using PFGE-derived data, but only six participants (66%) correctly identified the cluster of closely related isolates.

Compared to PFGE analysis of *Salmonella* or *Listeria*, the PFGE profile of *E. coli* contains a large number of bands within the region of 200–350 kb, which makes the cluster analysis based on PFGE harder to interpret. The PFGE gel needs to be of a very good quality in order to correctly assign all bands in this region. Two of the three laboratories, which did not identify the correct cluster, had accepted two or more band differences. One laboratory only identified the two technical isolates (REF19 and REF23) as the cluster.

This highlights the challenge of using PFGE on *E. coli* for inter-laboratory comparisons and shows PFGE can be a problematic method for cluster analysis of *E. coli*.

4.3.2 WGS-derived data

Eleven of 25 laboratories (44%) performed cluster analysis using WGS-derived data. Only one reported the use of external assistance for sequencing and the majority (10/11) reported using an Illumina platform. All reported using commercial kits for preparing the library.

Performance was very high, with 10 (91%) correctly identifying the cluster of closely related isolates. Out of 11 laboratories, eight (73%) reported using an allele-based method as the main analysis and three (27%) reported using SNP analysis. The one laboratory that did not identify the correct cluster had used SNP analysis. If only evaluating the main analysis of the laboratories reporting the correct cluster, the distances reported using SNP-based analyses were 0–5 inside the cluster and the number of allele differences using cgMLST were 0–3 inside the cluster.

The two approaches to analyse WGS-derived data (allele- and SNP-based analysis) showed comparable results. One exception was the results from the laboratory not identifying the correct cluster, because they misunderstood the question and reported all isolates analysed by the SNP analysis as the cluster of closely related isolates. The laboratory submitted raw data of excellent quality, but the reported SNP distances within the correct cluster were 0–42 and much higher than the expected 0–4 SNPs.

This is problematic in terms of inter-laboratory comparability and cluster definitions and makes the use of SNP distances obtained from non-standardised SNP analyses less suitable for communicating about genetic clusters when investigating international outbreaks. The reported high SNP distances seem unlikely to represent real biological divergence and are more likely to be artefacts of SNP calling.

Furthermore, laboratory 34 submitted results of all isolates in a joint SNP analysis without excluding remote isolates (as an additional analysis). This resulted in lower reported distances than the other participants for isolates outside the cluster, but had no discernible influence on the distances within the cluster.

High similarity was seen for the reported cgMLST results based on Enterobase (0–2 allele differences within the cluster). Other schemes used for allele-based analysis (both wgMLST and in house cgMLST schemes) showed

similar results, but up to seven allele differences within the cluster. Particularly outside the cluster the choice of scheme can have an influence on reported allele differences. For the ST21 isolates outside the cluster, the reported allelic differences were 17–56 based on cgMLST and up to 120 using wgMLST. This highlights the potential of cgMLST for standardisation and improved inter-laboratory comparability for cluster definitions.

The reported QC parameters (quantitative and qualitative) were used by the participants as QC of their data before analysis and submission. The main reported QC parameters, coverage, cgMLST allele calls and species confirmation are all essential for the end use of the data.

In order to compare the quality of the raw data, the EQA provider analysed the submitted raw reads to obtain selected QC parameters. All laboratories submitted sequences of fine quality. Two laboratories sequenced at a significantly lower depth than the remaining laboratories, but with no discernible negative effects on cluster analysis.

The EQA provider's analysis of the submitted raw data showed that when using a standardised cgMLST analysis, it is not unlikely to observe a random variation of one allele, even with high coverage (Figure 10). However, one participant (108) deviated systematically. This is likely due to a combination of sequencing technology and allele-calling software.

5 Conclusions

Twenty-five laboratories participated in the EQA-8 scheme, with 23 (92%) performing the serotyping part, 25 (100%) the virulence profile part and 18 (72%) cluster identification. In the EQA, a change was made from including quality assessment of PFGE in EQA-7 to including molecular typing-based cluster analysis using either PFGE and/or WGS-derived data in EQA-8. This adjustment of the EQA seemed to be well accepted by most Member States, but a decrease in the number of participants was seen compared with previous years for both serotyping and virulence profile. Furthermore, not all laboratories performing PFGE (EQA-7) signed up for molecular typing-based cluster analysis.

In the present EQA, 57% of the laboratories reporting serotyping results still used phenotypic serotyping. It will be interesting to follow the development of substituting conventional serotyping with molecular.

The O:H serotyping was only performed by 60% (15/25) of the participants, with an average score of 86%. As in previous EQAs, participation in the O grouping was higher than in H typing. Notably, not all laboratories demonstrated the capacity to determine all O groups and H types. Evidently, the majority of incorrect results were reported as ND or NT. In general, the more common European serotypes generated the highest scores. Serotype O187:H28 generated the lowest scores, correctly reported by eight laboratories.

The performance of detecting *aaiC/aggR* has been high through the EQAs that included EAEC isolates. The present EQA demonstrated a high performance for *aaiC* and *aggR,* with 85% and 90% average scores respectively.

Detection of *eae* had high participation rates and average scores through the EQAs has been above 96% (EQA-4: 96%; EQA-5: 98%; EQA-6: 97%; EQA-7: 98%; EQA-8: 96%).

Similar to the previous EQAs, the participation and average score for *stx1* and *stx2* gene detection were high, with a 100% average score for stx1 and 98% for *stx2.* Subtyping of *stx1* and *stx2* is highly valuable since specific subtypes have been associated with HUS. The fairly high participation 88% is therefore encouraging. Thus, the average score for subtyping of both *stx1* and *stx2* has been affected this year by the participants' unexpected reporting of not done results, i.e. not performing the subtyping test on isolates that are negative in the initial screening (detection for *stx1* and *stx2*). In the current EQA, the 'true' incorrect *stx2* results were mainly due to reporting only one *stx2* subtype for a specific isolate harbouring two types.

Incorporating a molecular typing-based cluster analysis in this EQA is up to date with the development of surveillance methods used by PH NRLs in Europe. Eleven laboratories performed cluster analysis using WGS-derived data. Performance was high, with 10 (91%) of participants correctly identifying the cluster of four closely related isolates, but one laboratory reported a much bigger cluster of 10 isolates with larger internal SNP distances in the correct cluster. A major part (7/18, 39%) of the participants in the cluster analysis used only PFGE and three did not identify the correct cluster using PFGE. The higher performance by WGS compared to PFGE emphasises the advantage of using WGS instead of PFGE for cluster analysis of STEC.

An allele-based method was preferred by most laboratories, as 73% (8/11) used cgMLST or wgMLST compared to 27% (3/11) using SNP as the main reported cluster analysis. The use of a standard cgMLST scheme (e.g. Enterobase) gives a very high degree of homogeneity in the results, but allele-based methods seem to be useful for inter-laboratory comparability and communication about cluster definitions. More challenges can appear if a non-standardised SNP analysis is used for comparison and communication between laboratories. For example, one laboratory reported significantly different SNP distances compared with the others for the predefined cluster.

Another difficulty is illustrated by data from one laboratory where the inclusion of remote isolates in the SNP analysis resulted in a reduced core genome and consequently lower SNP distances. However, this had little impact on cluster isolates.

The current EQA scheme for typing STEC is the eighth organised for laboratories in FWD-Net. The molecular surveillance system implemented as part of TESSy relies on the capacity of FWD-Net laboratories to produce analysable and comparable typing results in a central database. WGS-based typing for surveillance is increasingly used in EU. It was planned to allow WGS variables for *L. monocytogenes* to be submitted to the TESSy database in 2018. It is anticipated that Member States will also be able to upload WGS variables for STEC to the TESSy database in the near future.

6 Recommendations

6.1 Laboratories

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

At least two laboratories are suspected to have mislabelled isolates or swapped the results. Laboratories are encouraged to check correct labelling by checksum or otherwise before submission if internal labelling is used.

Laboratories are expected to employ each method as an individual test irrespective of results obtained in the screening and detection or any other test. Therefore, when a participant signs up for a test and subsequently participates, all isolates must be tested using this test, e.g. subtyping of *stx*.

6.2 ECDC and FWD-Net

ECDC will encourage and assist new participants potentially through training or workshops. ECDC plans to standardise the TESSy system for use of MLST nomenclature and cgMLST.

6.3 EQA provider

This year, the EQA provider changed the invitation letter to contain the recommended methods and a short description of the molecular typing-based cluster analysis. Requirements for submission and evaluation criteria were also listed. The submission protocol was short and precise. In the next round, participants who do not comply with the requested naming of FASTQ files will be asked to rename their files.

The link to the online submission was personal to the email listed during registration, so participants needed to circulate the email within their institute. The participants will be made aware of this issue in the next round.

In the next round of EQAs, laboratories will be asked to report the number of loci in the used allelic scheme and the name of the used SNP pipeline if publicly available. The EQA provider will try to give the participants more time to test and report results.

7 References

1. European Parliament and European Council. Regulation (EC) No 851/2004 of the European Parliament and of the Council of 21 April 2004 establishing a European centre for disease prevention and control – Article 5.3. Strasbourg: European Parliament and European Council; 2004. Available from: http://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:32004R0851.
2. European Parliament and European Council. Decision No 1082/2013/EU of the European Parliament and of the Council of 22 October 2013 on serious cross-border threats to health and repealing Decision No 2119/98/EC (Text with EEA relevance). Strasbourg: European Parliament and European Council; 2013. Available from: http://publications.europa.eu/en/publication-detail/-/publication/8d817a1f-45fa-11e3-ae03-01aa75ed71a1.
3. European Food Safety Authority and European Centre for Disease Prevention and Control. The European Union summary report on trends and sources of zoonoses, zoonotic agents and food-borne outbreaks in 2016. EFSA Journal. 2017;15(12):5077. Available from: http://ecdc.europa.eu/publications-data/european-union-summary-report-trends-and-sources-zoonoses-zoonotic-agents-and-9.
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. 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.
8. International Organization for Standardization. ISO/IEC 17043:2010 – Conformity assessment -- General requirements for proficiency testing. Vernier: ISO; 2010. Available from: http://www.iso.org/iso/catalogue_detail.htm?csnumber=29366.
9. 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 from: http://www.ssi.dk/English/HealthdataandICT/National%20Reference%20Laboratories/Bacteria/~/media/49802860CB5E44D6A373E6116ABBDC0D.ashx.
10. 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.
11. PulseNet International. Standard operating procedure for PulseNet PFGE of *Escherichia coli* O157:H7, *Salmonella* serotypes, *Shigella sonnei*, and *Shigella flexneri*. Atlanta: PulseNet International; 2013. Available at: http://www.pulsenetinternational.org/assets/PulseNet/uploads/pfge/PNL05_Ec-Sal-ShigPFGEprotocol.pdf.
12. Warwick Medical School. EnteroBase [Internet]. Coventry: University of Warwick; 2018 [cited 21 August 2018]. Available from: http://enterobase.warwick.ac.uk.
13. 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.
14. European Centre for Disease Prevention and Control. 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 from: http://ecdc.europa.eu/publications-data/survey-national-reference-laboratory-capacity-six-fwd-eueea-countries.
15. Statens Serum Institut. SerumQC [Internet; software package]. Copenhagen: Statens Serum Institut; 2018 [cited 21 August 2018]. Available from: https://www.github.com/ssi-dk/SerumQC.
16. Iguchi A, Iyoda S, Seto K, Morita-Ishihara T, Scheutz F, Ohnishi M, et al. *Escherichia coli* O-Genotyping PCR; a Comprehensive and Practical Platform for Molecular O Serogrouping. J.Clin.Microbiol. 2015 Aug;53(8):2427-32.

17. Lacher, DW, Gangiredla J, Patel I, Elkins CA, Feng PC. Use of the *Escherichia coli* Identification Microarray for Characterizing the Health Risks of Shiga Toxin-Producing *Escherichia coli* Isolated from Foods. J Food Prot. 2016 Oct;79:1656-1662.

18 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.

19. Friesema I, van der Zwaluw K, Schuurman T, Kooistra-Smid M, Franz E, van Duynhovn Y, et al. Emergence of *Escherichia coli* encoding Shiga toxin 2f in human Shiga toxin-producing *E. coli* (STEC) infections in the Netherlands, January 2008 to December 2011. Euro Surveill. 2014 May 1:19(17):26-32. Available from: http://www.eurosurveillance.org/images/dynamic/EE/V19N17/art20787.pdf.

20. Grande L, Michelacci V, Bondi R, Gigliucci F, Franz E, Badouei MA, et al. Whole-Genome Characterization and Strain Comparison of VT2f-Producing *Escherichia coli* Causing Hemolytic Uremic Syndrome. Emerg Infect Dis. 2016 Dec;22:2078-2086.

21. Food and Agriculture Organization of the United Nations and World Health Organization. 2018. Shiga toxin-producing *Escherichia coli* (STEC) and food: attribution, characterization, and monitoring: report. Rome: FAO and WHO; 2018. http://www.who.int/iris/handle/10665/272871.

Annex 1. List of participants

Country	Laboratory	National institute
Austria	Nationale Referenzzentrale für Escherichia coli einschließlich Verotoxin bildender E. coli	Institut für Klinische Mikrobiologie und Hygiene (AGES)
Belgium	NRC STEC	Universitair Ziekenhuis Brussel
Czech Republic	NRL for E.coli and Shigella	National Institute of Public Health
Denmark	Foodborne Infections	Statens Serum Institut
Estonia	Laboratory of Communicable Diseases	Health Board
Finland	Expert Microbiology Unit	National Institute for Health and Welfare
France	CNR associé E.coli - Service de Microbiologie	CHU Robert Debré
Germany	NRC for Salmonella	Robert Koch Institute
Greece	National Reference Centre for Salmonella, Shigella, VTEC	National School of Public Health
Hungary	National Reference Laboratory of Enteral pathogen bacteria	National Public Health Institute
Iceland	Department of Clinical Microbiology	Landspítali University Hospital
Ireland	VTEC Reference Laboratory	Public Health Laboratory – Health Service Executive
Italy	Microbiological Food Safety and Foodborne Disease Unit	Istituto Superiore di Sanità
Latvia	Infectology Centre of Latvia, National Microbiology Reference Laboratory	Riga East University Hospital
Lithuania	National Public Health Surveillance Laboratory	Budget organisation
Luxembourg	Epidemiology and Microbial Genomics	Laboratoire National de Sante
Norway	National Reference Laboratory for Enteropathogenic Bacteria	Norwegian Institute of Public Health
Poland	Department of Bacteriology	National Institute of Public Health – National Institute of Hygiene
Portugal	LNR Infeções Gastrintestinais	Instituto Nacional de Saúde Doutor Ricardo Jorge
Romania	Molecular Epidemiology Laboratory	Cantacuzino National Institute of Research
Slovenia	National Laboratory of Health, Environment and Food	Centre for Medical Microbiology
Spain	Laboratorio Nacional de Referencia e Investigación en Enfermedades Transmitidas por Agus y Alimentos	Instituto de Salud Carlos III
Sweden	Microbiology	Folkhälsomyndigheten
The Netherlands	Department of Bacterial Surveillance and Response	National Institute for Public Health and the Environment
United Kingdom	Gastrointestinal Bacteria Reference Unit	Public Health England

Annex 2. Participation overview EQA-7/EQA-8

	2016-2017 (EQA-7)					2017-2018 (EQA-8)				
	All	PFGE	Serotyping	Virulence	Phenotyping	All	Serotyping	Virulence	PFGE	WGS
Laboratory number									Cluster	
19	x	x	x	x	x	x	x	x	x	x
34	x	x	x	x	x	x	x	x		x
80	x		x	x	x	x	x	x		x
88	x		x	x	x	x		x		
90	x	x		x		x		x	x	
94	x		x	x	x					
100	x	x	x	x	x	x	x	x	x	
108	x		x	x	x	x	x	x		X
114	x	x	x	x	x					
123	x	x	x	x	x	x	x	x	x	X
124	x	x	x	x	x	x	x	x	x	
125	x		x	x	x					
126	x				x					
127	x	x	x	x	x	x	x	x	x	
128	x		x	x	x	x	x	x		
129	x		x	x	x	x	x	x		X
130	x	x	x	x	x	x	x	x		
131	x		x	x	x	x	x	x		
132	x	x	x	x	x	x	x	x	x	
133	x	x	x	x	x	x	x	x		X
134	x	x	x	x	x	x	x	x		X
135	x	x	x	x	x	x	x	x		X
136	x	x	x	x	x	x	x	x	x	
137	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		X
145	x	x	x	x	x	x	x	x		
153	x		x	x	x	x	x	x		
180	x	x								
222	x	x	x	x	x	x	x	x	x	
Number of participants	**30**	**19**	**27**	**28**	**28**	**25**	**23**	**25**	**9**	**11**

Annex 3. Serotyping result scores

O group

		Laboratory number																					
Isolate	**EQA**	**19**	**34**	**80**	**100**	**108**	**123**	**124**	**127**	**128**	**129**	**130**	**131**	**132**	**133**	**134**	**135**	**136**	**137**	**138**	**139**	**145**	**153**	**222**
REF1	O126	126	126	126	126	126	126	NT	126	ND	126	ND	126	ND	126	126	126	126	126	NT	126	126	126	126
REF2	O121	121	121	121	121	121	121	121	121	121	NT	ND	121	ND	121	121	121	121	121	121	121	121	121	121
REF3	O187	187	74	NT	74	187	187	NT	74	ND	NT	ND	103	ND	187	187	187	ND	187	145	187	NT	ND	104
REF4	O157	157	157	157	157	157	157	157	157	157	157	157	157	157	157	157	157	157	157	157	157	157	157	157
REF5	O63	63	63	NT	63	63	63	63	63	63	63	ND	63	NT	63	63	63	63	63	146	63	63	ND	NT
REF6	O146	146	146	146	ND	146	146	146	NT	146	NT	146	146	ND	146	146	146	146	146	146	146	146	ND	146
REF7	O8	8	8	8	8	8	8	8	8	ND	8	ND	8	ND	8	8	8	8	8	113	8	NT	ND	8
REF8	O104	104	104	104	104	104	104	104	104	104	104	104	104	ND	104	104	104	104	104	NT	104	104	104	74
REF9	O91	91	91	91	91	91	91	91	91	91	91	91	91	ND	91	91	91	91	91	91	91	NT	ND	91
REF10	O154	154	154	NT	154	154	154	154	NT	ND	154	26	45	ND	154	154	154	154	154	55	154	109	ND	NT
REF11	O111	111	111	111	111	111	111	111	NT	111	111	ND	111	ND	111	111	111	111	111	111	111	111	111	111
REF12	O157	157	157	157	157	157	157	157	157	157	157	157	157	157	157	157	157	157	157	157	157	157	157	157
Method	**A**	**A**	**B**	**B**	**A**	**C**	**A**	**B**	**A**	**A**	**C**	**A**	**A**	**A**	**C**	**C**	**A**	**B**	**C**	**A**	**C**	**A**	**A**	**A**

n=23 participants
Purple shading: incorrect result
A: phenotypic serotyping, B: PCR-based serotyping, C: WGS-based serotyping
NT: non-typable
ND: not done.

H type

		Laboratory number																					
Isolate	**EQA**	**19**	**34**	**80**	**100**	**108**	**123**	**124**	**127**	**128**	**129**	**130**	**131**	**132**	**133**	**134**	**135**	**136**	**137**	**138**	**139**	**145**	**153**	**222**
REF1	H27/H$^-$	H$^-$	27	NT		27	27	H$^-$	27		27		H-		27	27	27	27	27		27			
REF2	H19	19	19	19		19	19	NT	19		19		19		19	19	19	19	19		19			
REF3	H28	28	28	28		28	28	NT	28		28		2		28	28	28	28	28		28			
REF4	H7	7	7	7		7	7	7	7		7		7		7	7	7	7	7		7			
REF5	H6/H$^-$	6	6	NT		6	6	NT	6		6		6		6	6	6	6	6		6			
REF6	H28	H$^-$	28	28		28	28	H-	28		28		H-		28	28	28	28	28		28			
REF7	H9	9	9	NT		9	9	NT	9		9		9		9	9	9	9	9		9			
REF8	H4	4	4	4		4	4	NT	4		4		4		4	4	4	4	4		4			
REF9	H14	14	14	NT		14	14	H-	14		14		14		14	14	14	14	14		14			
REF10	H31	31	36	NT		31	31	NT	31		31		31		31	31	31	31	31		31			
REF11	H8/H$^-$	H$^-$	H$^-$	8		8	8	H$^-$	8		8		H$^-$		8	8	8	8	8		8			
REF12	H7	7	7	7		7	7	7	7		7		7		7	7	7	7	7		7			
Method	**A**	**A**	**A**	**B**		**C**	**A**	**B**	**A**		**C**		**A**		**C**	**C**	**A**	**B**	**C**		**C**			

n=15 participants
Purple shading: incorrect results
A: phenotypic serotyping, B: PCR-based serotyping, C: WGS-based serotyping
NT: non-typable.

Annex 4. Virulence profiles result scores

Detection of *aaiC*

		Laboratory number																			
Isolate	EQA	19	34	80	88	90	100	108	123	124	127	129	131	133	134	136	137	139	145	153	222
REF1	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	-	+	+
REF2	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
REF3	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	+
REF4	-	-	-	-	-	-	-	-	-	-	-	ND	-	-	-	-	-	-	-	-	-
REF5	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
REF6	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
REF7	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
REF8	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	-
REF9	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
REF10	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
REF11	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
REF12	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-

n=20 participants
Purple shading: incorrect results
ND: not done.

Detection of *aggR*

		Laboratory number																			
Isolate	EQA	19	34	80	88	90	100	108	123	124	127	129	131	133	134	136	137	139	145	153	222
REF1	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
REF2	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
REF3	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	+
REF4	-	-	-	-	-	-	-	-	-	-	-	ND	-	-	-	-	-	-	-	-	-
REF5	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
REF6	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
REF7	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
REF8	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	-
REF9	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
REF10	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
REF11	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
REF12	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-

n=20 participants
Purple shading: incorrect results
ND: not done.

Detection of *eae*

		Laboratory number																							
Isolate	EQA	19	34	80	88	90	100	108	123	124	127	128	129	130	131	132	133	134	135	136	137	138	139	145	153	222
REF1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
REF2	+	+	+	+	+	+	+	+	+	+	+	+	+	-	+	+	+	+	+	+	+	+	+	+	+	+
REF3	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
REF4	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
REF5	+	+	+	+	+	+	+	+	+	+	+	+	+	-	+	+	+	+	+	+	+	+	+	+	+	+
REF6	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
REF7	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
REF8	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
REF9	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
REF10	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	+	-
REF11	+	+	+	+	+	+	+	+	+	+	+	+	+	-	+	+	+	+	+	+	+	+	+	+	+	+
REF12	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+

n=25 participants
Purple shading: incorrect results.

Detection of *stx1*

		Laboratory number																								
Isolate	EQA	19	34	80	88	90	100	108	123	124	127	128	129	130	131	132	133	134	135	136	137	138	139	145	153	222
REF1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
REF2	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
REF3	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
REF4	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
REF5	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
REF6	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
REF7	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
REF8	-	-	-	-	+	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
REF9	+	+	+	+	-	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
REF10	+	+	+	+	+	+	+	+	+	+	+	+	+	-	+	+	+	+	+	+	+	+	+	+	+	+
REF11	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
REF12	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+

n=25 participants
Purple shading: incorrect results.

Detection of *stx2*

		Laboratory number																								
Isolate	EQA	19	34	80	88	90	100	108	123	124	127	128	129	130	131	132	133	134	135	136	137	138	139	145	153	222
REF1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
REF2	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
REF3	+	+	+	+	+	+	+	+	+	+	+	+	+	-	+	+	+	+	+	+	+	+	+	+	+	-
REF4	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	-	+	+	+	+	+	+	+	+
REF5	+	+	+	+	+	+	+	+	+	+	+	+	+	-	+	-	+	+	+	+	+	+	+	+	+	+
REF6	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
REF7	+	+	+	+	+	+	+	+	+	+	+	+	+	-	+	+	+	+	+	+	+	+	+	+	+	+
REF8	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	+
REF9	+	+	+	+	+	+	+	+	+	+	+	+	+	-	+	+	+	+	+	+	+	+	+	+	+	+
REF10	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
REF11	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
REF12	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+

n=25 participants.

stx subtyping

stx1

		Laboratory number																						
Isolate	EQA	19	34	80	88	90	100	108	123	124	127	128	129	131	133	134	135	136	137	138	139	145	222
REF1	-	-	-	-	ND	ND	-	-	-	-	-	ND	-	ND	-	-	ND	-	-	-	-	ND	ND
REF2	-	-	-	-	ND	ND	-	-	-	-	-	ND	-	ND	-	-	ND	-	-	-	-	ND	ND
REF3	-	-	-	-	ND	ND	-	-	-	-	-	ND	-	ND	-	-	ND	-	-	-	-	ND	ND
REF4	-	-	-	-	ND	ND	-	-	-	-	-	ND	ND	ND	-	-	ND	-	-	-	-	ND	ND
REF5	-	-	-	-	ND	ND	-	-	-	-	-	ND	-	ND	-	-	ND	-	-	-	-	ND	ND
REF6	1a	1a	1a	1a	1a	1a	1a	1a	1a	1a	1a	ND	1a	1a	1a	1a	ND	1a	1a	1a	1a	1a	1a
REF7	-	-	-	-	ND	ND	-	-	-	-	-	ND	-	ND	-	-	ND	-	-	-	-	ND	ND
REF8	-	-	-	-	1a	ND	-	-	-	-	-	ND	-	ND	-	-	ND	-	-	-	-	ND	ND
REF9	1a	1a	1a	1a	ND	1a	1a	1a	1a	1a	1a	ND	1a	1a	1a	1a	ND	1a	1a	1a	1a	1a	1a
REF10	1d	1d	1d	1d	1d	1d	1d	1d	1d	1d	1d	ND	1d	1c; 1d	1d	1d	ND	1d	1d	1d	1d	1d	1d
REF11	1a	1a	1a	1a	1a	1a	1a	1a	1a	1a	1a	ND	1a	1a	1a	1a	ND	1a	1a	1a	1a	1a	1a
REF12	1a	1a	1a	1a	1a	1a	1a	1a	1a	1a	1a	ND	1a	1a	1a	1a	ND	1a	1a	1a	1a	1a	1a

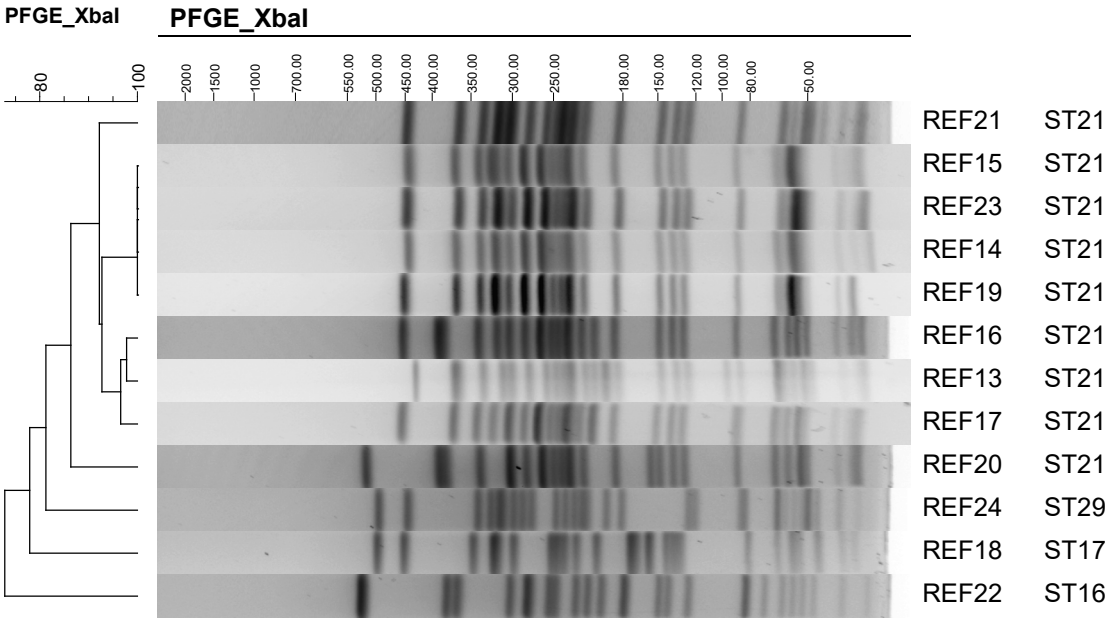
n=22 participants
Purple shading: incorrect result
ND: not done.

stx2

		Laboratory number																				
Isolate	**EQA**	**19**	**34**	**80**	**88**	**90**	**100**	**108**	**123**	**124**	**127**	**128**	**129**	**131**	**133**	**134**	**135**	**136**	**137**	**138**	**139**	**145**	**222**
REF1	-	-	-	-	ND	ND	-	-	-	-	-	ND	-	ND	-	-	ND	-	-	-	-	ND	ND
REF2	2a	2a	2a	2a	2a	2a	2a	2a	2a	2a	2a	ND	2a	2a 2c	2a	2a	ND	2a	2a	2a	2a	2a	2a
REF3	2g	2g	2g	2g	2g	2g	2g	2g	2g	2g	2g	ND	ND	2g	2g	2g	ND	2g	2g	2g	2g	2g	ND
REF4	2a 2c	2c	2a 2c	2a 2c	2a 2c	2a 2c	2a 2c	2a 2c	2a 2c	2a 2c	2a 2c	ND	ND	2a 2c 2d	2a 2c	-	ND	2a 2c	2a 2c	2a 2c	2a 2c	2a 2c	2a 2c
REF5	2f	2f	2f	2f	2f	2f	2f	2f	2f	2f	2f	2f	2f	2f	2f	2f	2f	2f	2f	2f	2f	2f	2f
REF6	2b	2b	2b	2b	2b	2b	2b	2b	2b	2b	2b	ND	2b	2b	2b	2b	ND	2b	2b	2b	2b	2b	2b
REF7	2a 2d/2e	2d	2a 2d	2a 2d	2a 2d	2a	2a 2d	2e	2e	2a	2a	ND	2e	2a 2d	2a	2e	ND	2a 2d	2a	2a 2d	2a 2d	2a 2d	2a 2d
REF8	-	-	-	-	-	ND	-	-	-	-	-	ND	-	ND	-	-	ND	-	-	-	-	ND	2g
REF9	2b	2b	2b	2b	2b	2b	2b	2b	2b	2b	2b	ND	2b	2b	2b	2b	ND	2b	2b	2b	2b	2b	2b
REF10	-	-	-	-	ND	ND	-	-	-	-	-	ND	-	ND	-	-	ND	-	-	-	-	ND	ND
REF11	-	-	-	-	-	ND	-	-	-	-	-	ND	-	ND	-	-	ND	-	-	-	-	ND	ND
REF12	2a	2a	2a	2a	2a	2a	2a	2a	2a	2a	2a	ND	2a	2a 2c	2a	2a	ND	2a	2a	2a	2a	2a	2a

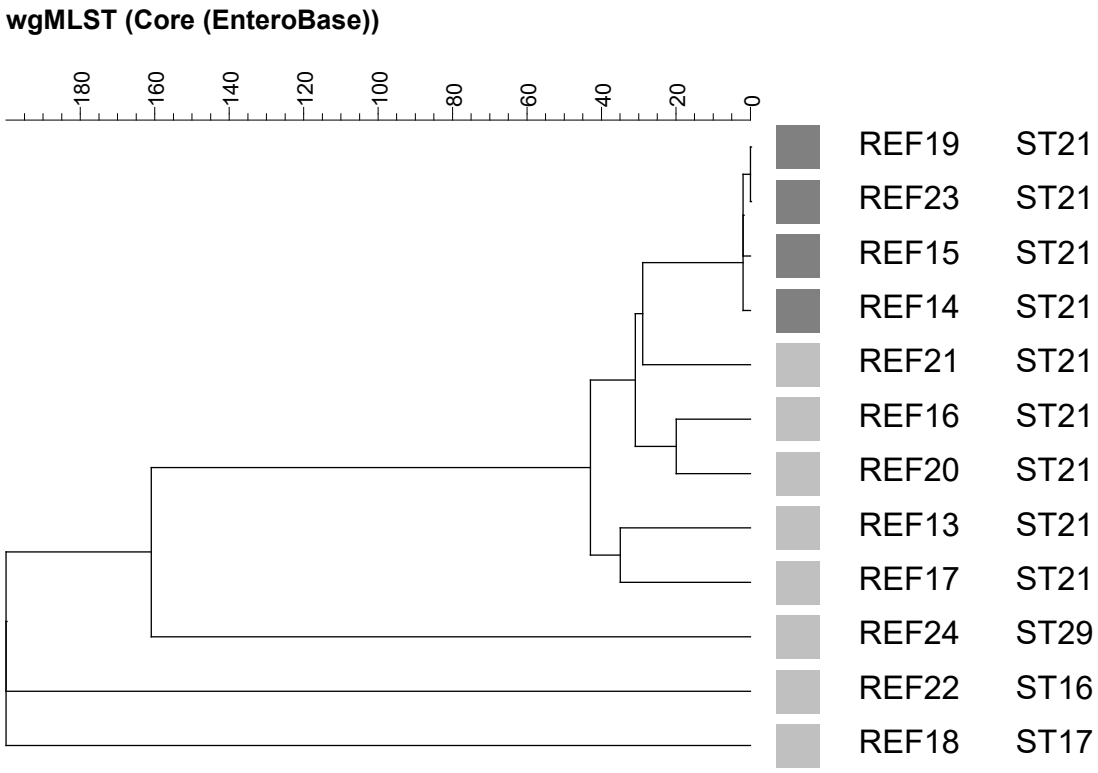
n=22 participants
Purple shading: incorrect results
ND: not done.

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



Cluster of closely related isolates: O26:H11, stx1a, REF14, REF15, REF19 and REF23.
REF19 and REF23 are technical duplicates.

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



Single linked dendrogram of core genome multilocus sequence typing (cgMLST) profiles of STEC EQA-8 isolates (cgMLST, EnteroBase, http://enterobase.warwick.ac.uk).
Analysed in BioNumerics: maximum distance of 200 exceeded, results clipped.
Cluster isolates: dark grey, outside cluster isolates: light grey.
REF19 and REF23 are technical duplicates.

Annex 7. Reported cluster of closely related isolates based on PFGE-derived data

Lab	Reported cluster	Corresponding REF isolates	Correct
		REF14, REF15, REF19, REF23 (19 and 23 technical duplicates)	
19	9191, 9211, 9523, 9649	REF15, REF14, REF19, REF23	Yes
90	9009, 9190, 9647, 9691	REF15, REF23, REF14, REF19	Yes
100	9136, 9623	REF19, REF 23	No
123	9479, 9886, 9067, 9159	REF23, REF14, REF19, REF15	Yes
124	9038, 9052, 9934, 9995, 9929	REF14, REF19, REF15, REF23, REF21	No
127	9152, 9209, 9704, 9793	REF15, REF23, REF19, REF14	Yes
132	9095, 9383, 9388, 9637, 9754, 9828, 9881, 9899	REF14, REF19, REF23, REF15, REF17, REF21, REF16, REF13	No
136	9507, 9797, 9826, 9864	REF15, REF14, REF19, REF23	Yes
222	9149, 9060, 9330, 9360	REF14, REF15, REF23, REF19	Yes

Annex 8. Reported band differences

			Laboratory number								
Isolate ID	**ST**	**Expected *XbaI* bands**	**19**	**90**	**100**	**123**	**124**	**127**	**132**	**136**	**222**
REF13	21	24	24	23	23	23	24	23	21	24	19
REF14‡	21	19	19	17	21	19	19	19	22	19	17
REF15‡	21	19	19	17	21	20	19	19	20	19	17
REF16	21	23	23	20	23	21	22	22	20	23	21
REF17	21	21	21	18	22	21	21	21	21	21	18
REF18	17	Clearly unrelated	22	16	19	19	18	19	17	20	16
REF19#‡	21	19	19	17	21	19	19	19	19	19	17
REF20	21	23	23	19	22	21	20	20	19	20	17
REF21	21	22	22	20	22	20	21	20	22	21	19
REF22	16	Clearly unrelated	21	20	21	20	21	17	20	21	20
REF23#‡	21	19	19	17	21	19	19	19	19	19	17
REF24	29	Clearly unrelated	23	20	20	22	22	20	20	21	17

			Laboratory number								
Isolate ID	**ST**	**Bands with shared *XbaI***	**19**	**90**	**100**	**123**	**124**	**127**	**132**	**136**	**222**
REF13	21	Approximately 18	18	17	15	9999	17	17	18	19	16
REF14‡	21	19	19	17	21	19	19	19	16	19	17
REF15‡	21	19	19	17	20	19	19	19	17	19	17
REF16	21	Approximately 18	16	17	15	13	18	18	14	18	17
REF17	21	Approximately 18	16	16	14	17	18	18	16	18	16
REF18	17	Clearly unrelated	9999	12	4	9999	5	5	9	10	13
REF19#‡	21	19	19	17	21	19	19	19	19	19	17
REF20	21	Approximately 14	14	15	13	9999	14	16	10	16	15
REF21	21	Approximately 17	17	17	16	16	19	17	19	19	17
REF22	16	Clearly unrelated	9	14	5	9999	10	10	5	15	14
REF23#‡	21	19	19	17	21	19	19	19	19	19	17
REF24	29	Clearly unrelated	13	14	8	9999	11	10	10	14	15

ST: sequence type
‡: cluster identification of closely related isolates (based on PFGE-derived data)
#: technical duplicate
9999: not reported by laboratory.

Annex 9. Reported sequencing details

Sequencing performed	Protocol (library prep)	Commercial kit	Sequencing platform
In own laboratory	Commercial kits	Illumina Nextera XT	MiSeq
In own laboratory	Commercial kits	Illumina Nextera XT	MiSeq
Externally	Commercial kits	Nextera XT	HiSeq 2500
In own laboratory	Commercial kits	Ion Xpress TM Plus Fragment Library Kit	Ion Torrent S5XL
In own laboratory	Commercial kits	Nextera XT	NextSeq
In own laboratory	Commercial kits	NexteraXT*	NextSeq
In own laboratory	Commercial kits	Illumina Nextera XT*	MiSeq
In own laboratory	Commercial kits	Nextera XT, Illumina*	MiniSeq, Illumina
In own laboratory	Commercial kits	Nextera XT, Illumina*	MiSeq
In own laboratory	Commercial kits	KAPA HyperPlus (Kapa Biosystems)	MiSeq
In own laboratory	Commercial kits	Nextera	HiSeq 2500

**: half volume for all reagents.*

Annex 10. Reported cluster of closely related isolates based on WGS-derived data

Lab	Reported cluster	Corresponding to REF isolates	Correct
		REF14, REF15, REF19, REF23 (19 and 23 technical duplicates)	
19	9523, 9191, 9211, 9649	REF19, REF15, REF14, REF23	Yes
34	9171, 9581, 9901, 9965	REF19, REF15, REF14, REF23	Yes
80	9096, 9922, 9985, 9717	REF23, REF19, REF15, REF14	Yes
108	9135, 9198, 9587, 9362	REF15, REF19, REF23, REF14	Yes
123	9479, 9067, 9886, 9159	REF23, REF19, REF14, REF15	Yes
129	9324, 9526, 9559, 9816	REF15, REF14, REF19, REF23	Yes
133	9347, 9486, 9999, 9844	REF19, REF23, REF14, REF15	Yes
134	9298, 9369, 9681, 9719	REF23, REF14, REF15, REF19	Yes
135	9505, 9917, 9972, 9973	REF19, REF23, REF14, REF15	Yes
137	9879, 9098, 9432, 9283	REF15, REF14, REF23, REF19	Yes
139	9051, 9134, 9170, 9197, 9375, 9402, 9543, 9570, 9589, 9708	REF20, REF16, REF21, REF24, REF15, REF14, REF19, REF21, REF13, REF17	No

Annex 11. Reported SNP distance and allelic differences

SNP distances

			Laboratory number				
Isolate ID	ST	Provider	19*	34*	108	137	139
REF13	21	444	374	36	397	548	397
REF14‡	21	4	4	4	5	5	28
REF15‡	21	4	4	4	5	4	0¤
REF16	21	256	238	71	321	359	332
REF17	21	285	280	35	328	294	375
REF18	17	9999	9999	1194	9999	9999	9999
REF19#‡	21	0	0¤	0¤	0¤	0¤	35
REF20	21	338	319	64	394	303	375
REF21	21	372	330	51	390	451	504
REF22	16	9999	9999	1119	9999	9999	9999
REF23#‡	21	0	0	0	0	0	42
REF24	29	9999	9999	241	9999	9999	988

Allelic differences

			Laboratory number							
Isolate ID	ST	Provider	19	34	80	123	129	133	134	135
REF13	21	44	43	46	45	45	26	108	30	64
REF14‡	21	2	2	0¤	2	2	3	0¤	2	3
REF15‡	21	2	2	2	2	2	0¤	5	2	3
REF16	21	36	33	53	37	9999	27	96	40	77
REF17	21	54	52	45	56	9999	34	120	55	115
REF18	17	1745	9999	1744	1744	9999	1141	2878	1403	64
REF19#‡	21	0	0¤	2	0	0	1	7	0¤	0¤
REF20	21	32	31	32	32	32	20	89	23	50
REF21	21	29	23	27	28	28	17	90	22	40
REF22	16	835	9999	928	816	9999	558	1557	780	1300
REF23#‡	21	0¤	0	2	0¤	0¤	1	6	1	0
REF24	29	163	163	161	163	162	112	417	134	246

ST: sequence type.
**: additional analysis*
‡: closely related isolates
#: technical duplicate isolate
¤: isolate used as cluster representative by participant
9999: isolates not included in analysis by participant

Annex 12. Reported QC parameters

	QC parameters	Thresholds
A	Coverage	> 50 fold average coverage
C	average coverage (assembled)	30
D	Average coverage	30x
E	Coverage	>40
F	Coverage	>50
G	Coverage (unassembled)	>70
H	Coverage	20 x
I	Average depth coverage (SNP-typing)	≥ 30x
A	% cgMLST targets	> 95% good targets
C	Percentage of good targets	90
D	Percentage good cgMLST targets	95%
E	core percent	>97
F	cgMLST alleles found	>95%
G	% good targets cgMLST	>95%
A	Confirmation of genus	-
B	Confirmation of genus	Main genus match in kraken must match supplied genus
G	Kraken	-
H	Confirmation of genus/species	-
I	Confirmation of genus, species and subspecies	N/A
A	Number of contigs	< 600
B	Number of contigs	<1 000
E	Number of contigs	>200
G	Contig count (assembled)	<1 000
B	Assembly length	> 4 500 000 & <5 300 000
F	Length of contig assembly	< reference genome + 10%
H	Genome size	+/- 20%
B	Contamination check	Only one genus >5% on mini kraken
G	Kraken	-
I	Contamination check	Contamination if ≥ 15% match to unexpected genome
B	Minimum per site coverage of assembly	25
D	VarScan min coverage/read depth at a position to match a cell	10
I	Minimum depth coverage (SNP-typing)	≥ 10x
G	MLST	-
I	Average coverage of all alleles (Achtman 7 gene MLST)	100 %
I	Maximum percentage non-consensus bases (Achtman 7 gene MLST)	≥ 15%
I	Minimum consensus depth (Achtman 7 gene MLST)	> 0
A	Average contig length	> 5.000
G	Contig length (assembled)	>5000
D	Average read length	180bp
G	Average read length (unassembled)	>195
D	Required identity to reference locus	90%
D	Required percentage aligned to reference locus	100%
A	N50	> 50 000
E	N50	90 000
D	VarScan min supporting reads at a position to cell variants	8
I	Variant ratio (SNP-typing)	≥ 0.9
B	Difference of number of contigs with avg coverage >0 and >25	1 000
B	Difference of sum of lengths of contigs with avg coverage >0 and >25	250 000
I	Minimum metric yield (sequence quality)	≥ 150 Mb

Annex 13. Calculated qualitative/ quantitative parameters

		Laboratory 19											
Parameters	Ranges*	9191	9155	9287	9211	9470	9649	9487	9523	9668	9855	9759	9994
Number of genera detected	{1}	1	1	1	1	1	1	1	1	1	1	1	1
Detected species	{Ec}	Ec	Ec	Ec	Ec	Ec	Ec	Ec	Ec	Ec	Ec	Ec	Ec
Unclassified reads (%)	{<100}	8.21	3.22	3.01	3.26	3.24	3.33	2.87	3.57	3.09	3.03	2.55	3.63
Length at 25 x min. coverage (Mbp)	{>45 ∧ <53}	5	5.1	5.3	5.3	5.3	5.3	5.2	5.2	5.1	5.4	5.2	5.2
Length [0–25] x min. coverage (kbp)	{<250}	62.6	14.8	27.3	12.8	18	14.4	9.8	24	7.2	1.1	8.4	5.7
Number of contigs at 25 x min. coverage	{>0}	527	249	329	267	283	277	266	276	200	279	269	234
Number of contigs [0–25] x min. coverage	{<1000}	33	9	18	7	9	10	7	12	5	1	6	3
Average coverage	{>50}	51.0	154.9	128.4	134.1	126.9	140.4	155.8	148.4	157.4	130.6	142.5	146.8
Number of reads (x1000)		1 166.8	3 093.7	2 714.8	2 751.7	2 658.5	2 879.2	3 211.6	3 077	3 123.7	2 758.2	2 893	3 014.3
Number of trimmed reads (x1000)		1 136.6	3 062.3	2 688.5	2 722.9	2 630.8	2 850.9	3 179	3 046.6	3 091.2	2 729.9	2 866.8	2 984.9
Maximum read length		151	151	151	151	151	151	151	151	151	151	151	151
Mean read length		123.4	139.8	137.9	139.2	137.6	139.3	137.7	137.8	138.8	138.4	139	137.7
Read insert size		208.6	316.6	313	327	318	317.2	299.9	304	308.8	314.3	301.8	301.9
Insert size StdDev		86.6	137.1	145.3	146.2	148.3	143.8	135.3	141.4	139.4	143.2	132.4	139.4
N50 (kbp)		16.8	47.7	41.9	50.9	50.3	47.1	53.9	48.3	73.3	48.4	58.1	56.8
N75 (kbp)		8.1	25	19.9	25.1	25.5	22.7	28.3	20.9	33.5	26.3	26.3	31.5

		Laboratory 34											
Parameters	**Ranges***	**9171**	**9468**	**9450**	**9173**	**9476**	**9581**	**9900**	**9803**	**9746**	**9901**	**9965**	**9632**
Number of genera detected	{1}	1	1	1	1	1	1	1	1	1	1	1	1
Detected species	{Ec}	Ec	Ec	Ec	Ec	Ec	Ec	Ec	Ec	Ec	Ec	Ec	Ec
Unclassified reads (%)	{<100}	8.01	5.14	7.66	1.27	2.06	5.46	1.48	3.34	6.46	6.49	5.89	1.87
Length at 25 x min. coverage (Mbp)	{>45 ∧ <53}	5.4	5.4	5.3	5.4	5.2	5.4	5.4	5.6	5.3	5.4	5.4	5.4
Length [0–25] x min. coverage (kbp)	{<250}	0	23.9	39.5	0	0	1.1	0	0	0	0	0	0
Number of contigs at 25 x min. coverage	{>0}	217	251	298	261	166	239	217	205	184	214	206	230
Number of contigs [0–25] x min. coverage	{<1000}	0	15	20	0	0	1	0	0	0	0	0	0
Average coverage	{>50}	51.6	37.3	49.3	56.9	46.7	53.5	42.1	80.1	108.2	57.5	78.0	86.5
Number of reads (x1000)		778.8	621.6	838.2	763.7	624.7	776.3	539.9	1 189.3	1 486.1	778.8	1 071.7	1 208.8
Number of trimmed reads (x1000)		744.9	618.1	809.5	734	592.4	747.1	507.2	1 143.3	1 422.8	739.9	1 055.7	1 150.5
Maximum read length		301	301	301	301	301	301	301	301	301	301	301	301
Mean read length		199.2	174.6	205.6	217.7	210.1	200.8	228.4	200.3	215.4	221.7	207.9	209.3
Read insert size		293.9	239	297.5	303.3	326.6	297.2	350.3	311.8	328.8	347.6	342	318.1
Insert size StdDev		141.1	93.6	139.3	133.3	154.6	140.8	154.1	150.1	155.2	161.3	177	149.9
N50 (kbp)		107.9	78.1	52.9	65.8	108.6	89.9	95.6	114.3	96.5	96.1	114.3	83.2
N75 (kbp)		46	31.8	22.6	31.1	57.8	41.1	42.6	46.5	42.2	41.5	43.4	43.5

		Laboratory 80											
Parameters	**Ranges***	**9096**	**9590**	**9200**	**9717**	**9457**	**9201**	**9782**	**9730**	**9720**	**9922**	**9961**	**9985**
Number of genera detected	{1}	1	1	1	1	1	1	1	1	1	1	1	1
Detected species	{Ec}	Ec	Ec	Ec	Ec	Ec	Ec	Ec	Ec	Ec	Ec	Ec	Ec
Unclassified reads (%)	{<100}	3.51	1.83	2.44	4.54	2.14	2.91	1.52	2.83	2.52	4.87	1.4	4.49
Length at 25 x min. coverage (Mbp)	{>45 ∧ <53}	5.4	5.2	5.4	5.4	5.6	5.3	5.3	5.2	5.5	5.4	5.4	5.4
Length [0–25] x min. coverage (kbp)	{<250}	43.1	35.5	0	0	0	33.5	26.9	21.1	28.9	0	16.9	0
Number of contigs at 25 x min. coverage	{>0}	628	328	225	208	210	207	227	230	287	218	386	222
Number of contigs [0–25] x min. coverage	{<1000}	6	2	0	0	0	4	7	15	14	0	1	0
Average coverage	{>50}	86.0	86.0	111.5	97.5	108.1	127.1	95.1	146.9	125.8	103.5	118.4	104.2
Number of reads (x1000)		1 178.2	1 124.2	1 530.7	1 300.6	1 461.2	1 625.5	1 230.3	2 025.2	1 773.3	1 368.1	1 570	1 385.9
Number of trimmed reads (x1000)		1 123.4	1 049.8	1 426.3	1 234.5	1 365.9	1 565.5	1 175.9	1 964.3	1 716.4	1 302.3	1 488.4	1 319.5
Maximum read length		301	301	301	301	301	301	301	301	301	301	301	301
Mean read length		221.1	222.4	220.7	218.8	225.2	220.9	222.7	200.7	208.5	220.4	223.8	218.8
Read insert size		317.9	316.5	322.5	309.4	323.2	301	320.7	247.9	270.6	317.8	320	314
Insert size StdDev		126.1	113.4	118	123.5	118.7	105.8	109.9	78.4	91.1	117.2	109.7	117.4
N50 (kbp)		14.6	31	105.9	107.9	114.2	108	104.2	76.2	92	108	32.9	107.3
N75 (kbp)		7.4	16.5	46	43.5	46.1	47.7	43.5	36.5	32.9	46.6	15.9	46

		Laboratory 108											
Parameters	Ranges*	9138	9033	9198	9362	9474	9135	9497	9996	9629	9925	9587	9653
Number of genera detected	{1}	1	1	1	1	1	1	1	1	1	1	1	1
Detected species	{Ec}	Ec	Ec	Ec	Ec	Ec	Ec	Ec	Ec	Ec	Ec	Ec	Ec
Unclassified reads (%)	{<100}	2.36	1.18	1.91	1.81	1.18	2.43	1.39	2.15	1.94	2.6	1.78	1.85
Length at 25 x min. coverage (Mbp)	{>45 ∧ <53}	5.2	5.3	5.3	5.3	5.3	5.3	5.3	5.1	5.4	5.5	5.3	5.2
Length [0–25] x min. coverage (kbp)	{<250}	7.4	0	1.1	0	0	0	3.4	166.2	1.1	1.1	0	2.6
Number of contigs at 25 x min. coverage	{>0}	229	407	391	365	397	420	408	248	406	393	435	299
Number of contigs [0–25] x min. coverage	{<1000}	6	0	1	0	0	0	3	27	1	1	0	1
Average coverage	{>50}	38.2	49.7	51.1	52.8	55.6	61.3	55.2	37.0	49.6	46.4	59.7	290.5
Number of reads (x1000)		571.8	788.5	800.6	846.4	871.4	971	897.9	577	812.5	758.3	930.6	5313.5
Number of trimmed reads (x1000)		535.8	737.5	750.7	783.4	821.7	915.5	849.6	542.9	761.8	710.7	872.4	4887.6
Maximum read length		307	306	305	307	303	305	304	302	306	306	307	304
Mean read length		219.4	218.6	222.1	219.1	219.7	224.1	217.2	212.2	217.6	216.8	224.9	217.5
Read insert size		NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
Insert size StdDev		NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
N50 (kbp)		76.2	30.6	31.6	39.9	35.5	31.1	33	66.4	36	33	28.3	42.7
N75 (kbp)		34.9	14.6	17.6	20.1	15	13.8	15	39.4	15.1	15.3	13.5	22.1

		Laboratory 123											
Parameters	Ranges*	9479	9622	9067	9423	9275	9159	9185	9120	9491	9770	9804	9886
Number of genera detected	{1}	1	1	1	1	1	1	1	1	1	1	1	1
Detected species	{Ec}	Ec	Ec	Ec	Ec	Ec	Ec	Ec	Ec	Ec	Ec	Ec	Ec
Unclassified reads (%)	{<100}	2.4	1.88	4.86	7.11	1.43	3.91	2.15	1.64	2.15	1.31	1.82	3.17
Length at 25 x min. coverage (Mbp)	{>45 ∧ <53}	5.2	5.2	5.4	5.2	5.3	5.4	5.3	5.4	5.5	5.3	5.5	5.4
Length [0–25] x min. coverage (kbp)	{<250}	123.9	25.3	4.3	20	0	2.1	0	1.7	0	25.5	77.1	4.6
Number of contigs at 25 x min. coverage	{>0}	313	183	230	198	243	245	237	243	293	280	243	242
Number of contigs [0–25] x min. coverage	{<1000}	48	9	3	4	0	1	0	1	0	6	22	3
Average coverage	{>50}	49.0	63.4	54.0	73.4	73.5	53.0	72.7	73.5	66.7	64.8	53.2	65.9
Number of reads (x1000)		679.8	858.1	765.4	995.2	972.1	709.5	937.8	1 010.1	936.4	862.6	709.9	880.4
Number of trimmed reads (x1000)		651.3	836.5	747.1	974.9	938.2	679	918.9	985.9	907.6	836.7	697.5	852.9
Maximum read length		301	301	301	301	301	301	301	301	301	301	301	301
Mean read length		208.1	204	206.3	212	218.4	218.9	219.8	214.6	212	216.9	218.5	216.6
Read insert size		282.8	267.6	266.2	271.9	290.9	305.1	291.4	278.1	286.1	293.6	288.9	292.2
Insert size StdDev		100.8	105.8	100.4	102.6	112	110.2	114.3	104.2	116	116.3	109.4	113.6
N50 (kbp)		45.4	92.4	92.7	91	83.6	87.8	79.5	91	79.5	58.5	71.9	90.8
N75 (kbp)		19.9	47.1	41.1	39.9	33	37.9	33.6	41.5	30.9	27.5	32.4	37.6

		Laboratory 129											
Parameters	**Ranges***	**9034**	**9324**	**9321**	**9422**	**9526**	**9696**	**9816**	**9517**	**9538**	**9810**	**9559**	**9699**
Number of genera detected	{1}	1	1	1	1	1	1	1	1	1	1	1	1
Detected species	{Ec}	Ec	Ec	Ec	Ec	Ec	Ec	Ec	Ec	Ec	Ec	Ec	Ec
Unclassified reads (%)	{<100}	0.84	1.7	1.52	1.0	3.42	2.29	2.32	0.62	0.93	0.76	2.87	1.01
Length at 25 x min. coverage (Mbp)	{>45 ∧ <53}	5.3	5.2	5	5.1	5.3	5.2	5.3	5.3	5.4	5.2	5.3	5.3
Length [0–25] x min. coverage (kbp)	{<250}	26.7	48	117.2	18	4.7	15.3	4.6	96.9	9.8	9.6	2.1	1.6
Number of contigs at 25 x min. coverage	{>0}	226	264	230	183	196	217	202	227	275	216	208	192
Number of contigs [0–25] x min. coverage	{<1000}	9	21	16	6	1	7	2	9	6	4	2	1
Average coverage	{>50}	87.3	78.2	86.4	90.3	104.9	87.7	78.4	102.8	116.7	114.1	139.1	169.3
Number of reads (x1000)		1 743.6	1 510.9	1 627.9	1 665.8	2 085.1	1 710.5	1 531.4	2 007.5	2 367.7	2 261.6	2 778.3	3 280.7
Number of trimmed reads (x1000)		1 715.1	1 482.4	1 598	1 642.1	2 049.8	1 672.4	1 498.4	1 986.8	2 334.1	2 186.7	2 741.3	3 235.5
Maximum read length		151	151	151	151	151	151	151	151	151	151	151	151
Mean read length		146.4	145.6	145.6	146.2	146.4	145.8	145.5	146.8	145.1	144.6	145.7	145.9
Read insert size		479.5	452.9	421.3	401.6	426.2	416	382.4	415.5	330.7	410.3	359.7	367.4
Insert size StdDev		173.2	161.1	152.3	151.4	170.2	159.7	158.1	156.1	145.3	154	142.9	166.2
N50 (kbp)		59.9	47	49.5	69.7	90.7	81.9	95.5	72.4	52.6	68.2	91.9	92.1
N75 (kbp)		31.1	24.2	26.2	37.4	37.6	32.8	37.5	31.2	30.6	32.3	33.9	41

		Laboratory 133											
Parameters	**Ranges***	**9395**	**9347**	**9431**	**9207**	**9694**	**9789**	**9896**	**9486**	**9563**	**9999**	**9844**	**9991**
Number of genera detected	{1}	1	1	1	1	1	1	1	1	1	1	1	1
Detected species	{Ec}	Ec	Ec	Ec	Ec	Ec	Ec	Ec	Ec	Ec	Ec	Ec	Ec
Unclassified reads (%)	{<100}	2.62	5.76	1.05	2.67	1.76	1.73	1.24	4.51	3.58	5.3	4.85	3.46
Length at 25 x min. coverage (Mbp)	{>45 ∧ <53}	5.2	5.4	5.4	5.4	5.2	5.4	5.3	5.4	5.3	5.4	5.3	5.5
Length [0–25] x min. coverage (kbp)	{<250}	15.2	0	4.2	0	0	111.3	19.3	0	1.2	0	0	0
Number of contigs at 25 x min. coverage	{>0}	181	222	222	244	152	225	213	230	228	218	244	230
Number of contigs [0–25] x min. coverage	{<1000}	7	0	2	0	0	29	8	0	1	0	0	0
Average coverage	{>50}	53.5	62.2	83.3	114.9	100.8	71.7	90.3	108.6	163.7	124.9	129.2	144.9
Number of reads (x1000)		653.5	832.7	1 174.6	1 742.2	1 204.8	915.9	1 152	1 490.2	2 371.5	1 765.6	1 889.2	2 079
Number of trimmed reads (x1000)		636.8	793.1	1 111.4	1 702.7	1 188.2	904.5	1 134	1 429	2 337.5	1 736.9	1 842.7	2 010.6
Maximum read length		301	301	301	301	301	301	301	301	301	301	301	301
Mean read length		228.1	223.6	209.3	195.6	230.6	230.3	223.2	218	200.1	208.2	206.5	208.7
Read insert size		340.2	334	307.8	261.4	327.9	320.3	302.1	309.2	267.8	286.4	282.8	280
Insert size StdDev		135.4	119.4	124.7	101.8	164.8	153.1	137.8	119.5	115	121.3	113.7	106.8
N50 (kbp)		97.1	106.3	95.6	104.1	126.8	104.2	92.2	107.9	99.4	107.9	106.4	115.1
N75 (kbp)		42.2	43.8	43.4	41.4	57.8	41.4	43.4	43.5	46.8	46	43.5	43.1

		Laboratory 134											
Parameters	Ranges*	9012	9094	9298	9369	9719	9681	9706	9749	9729	9818	9914	9990
Number of genera detected	{1}	1	1	1	1	1	1	1	1	1	1	1	1
Detected species	{Ec}	Ec	Ec	Ec	Ec	Ec	Ec	Ec	Ec	Ec	Ec	Ec	Ec
Unclassified reads (%)	{<100}	1.54	1.11	1.72	2.07	1.76	2.25	2.12	1.82	1.39	1.22	1.72	2.15
Length at 25 x min. coverage (Mbp)	{>45 ∧ <53}	5.1	5.3	5.3	5.3	2.9	5.3	5.4	5.5	5.3	5.3	5.2	5.2
Length [0–25] x min. coverage (kbp)	{<250}	6.7	0	0	0	2415.3	0	1.3	0	2.4	0	1.1	0
Number of contigs at 25 x min. coverage	{>0}	155	195	196	195	146	196	234	208	215	197	190	166
Number of contigs [0–25] x min. coverage	{<1000}	2	0	0	0	114	0	1	0	1	0	1	0
Average coverage	{>50}	58.6	62.8	74.3	70.7	28.0	75.5	68.3	57.9	72.8	64.4	71.3	79.7
Number of reads (x1000)		1 067	1 180.4	1 406.9	1 342.7	524.7	1 434.6	1 332	1 125	1 388.3	1 208.1	1 335.1	1 482.2
Number of trimmed reads (x1000)		1 059.5	1 172.8	1 397	1 332.1	520.1	1 424.1	1 322	1 117	1 378.1	1 200.2	1 324.9	1 471.1
Maximum read length		151	151	151	151	151	151	151	151	151	151	151	151
Mean read length		146.9	146.9	146.9	146.8	147.2	146.3	146.5	146.9	146.2	146.9	146.5	146.3
Read insert size		388.2	374.3	373.7	364	427.2	375.4	388.9	397.4	385.7	390.8	387.5	359.3
Insert size StdDev		132.2	129.6	129.9	125.9	136.1	136.2	134.7	134.8	136	133.3	134.5	127
N50 (kbp)		124.9	92.3	91.9	97.9	45.4	91.9	81.1	89.2	95.4	84.1	97.9	99.6
N75 (kbp)		47.5	41	39.9	39.8	27.6	41.2	34.1	40.8	35.6	38.3	41.9	40.7

		Laboratory 135											
Parameters	Ranges*	9617	9410	9600	9851	9973	9917	9904	9972	9158	9505	9105	9093
Number of genera detected	{1}	1	1	1	1	1	1	1	1	1	1	1	1
Detected species	{Ec}	Ec	Ec	Ec	Ec	Ec	Ec	Ec	Ec	Ec	Ec	Ec	Ec
Unclassified reads (%)	{<100}	1.9	2.06	1.73	1.68	2.57	3.38	2.85	3.04	1.94	2.44	2.03	7.05
Length at 25 x min. coverage (Mbp)	{>45 ∧ <53}	5.2	5.2	5.3	5.1	5.2	5.2	5.2	5.2	5.2	5.2	5.4	5.2
Length [0–25] x min. coverage (kbp)	{<250}	54.7	58.8	95.3	0	51.4	53.2	4.7	52	20.1	14.8	0	4.6
Number of contigs at 25 x min. coverage	{>0}	197	192	189	150	196	193	189	187	202	193	226	157
Number of contigs [0–25] x min. coverage	{<1000}	9	9	9	0	16	13	1	16	3	2	0	2
Average coverage	{>50}	68.6	65.9	63.0	98.9	70.8	68.4	119.2	93.6	80.9	153.2	112.8	247.5
Number of reads (x1000)		1 606.5	1 540.8	1 520.4	2 238.1	1 659.8	1 627.6	2 810.2	2 213.2	1 951	3 641.3	2 788.6	6241
Number of trimmed reads (x1000)		1 573.4	1 510.4	1 489.5	2 191.5	1 628.9	1 596.7	2 747	2 165.7	1 908.5	3 572.5	2 731.5	6 058.4
Maximum read length		126	126	126	126	126	126	126	126	126	126	126	126
Mean read length		120.5	120.3	120.3	121.5	120.6	119.3	120.6	120.7	120.2	121.3	119.5	121.5
Read insert size		340.1	327.4	338.1	331.7	323.3	327.6	335	357	345.4	328.8	330.7	371.9
Insert size StdDev		202.4	197	205.1	188.3	193.8	205.1	200.3	211.8	210.9	190.8	205.1	213.6
N50 (kbp)		89.1	76.9	91.9	104	89.7	95.4	91.9	91.9	95.3	95.4	80.1	92.1
N75 (kbp)		39.8	40.9	41.1	45.5	36.2	36.4	43.1	39.8	38.6	40.5	37.5	41.2

		Laboratory 137											
Parameters	**Ranges***	**9247**	**9643**	**9879**	**9283**	**9098**	**9923**	**9819**	**9830**	**9208**	**9987**	**9466**	**9432**
Number of genera detected	{1}	1	1	1	1	1	1	1	1	1	1	1	1
Detected species	{Ec}	Ec	Ec	Ec	Ec	Ec	Ec	Ec	Ec	Ec	Ec	Ec	Ec
Unclassified reads (%)	{<100}	12.07	1.82	3.63	3.63	3.19	2.08	1.4	1.2	1.79	2.93	1.3	3.01
Length at 25 x min. coverage (Mbp)	{>45 ∧ <53}	2	0.6	2.3	5.1	5.3	5.3	5.2	5.1	5.1	1.5	5.3	5.3
Length [0–25] x min. coverage (kbp)	{<250}	2 845.4	4 860.8	2 935.5	104.9	0	26.2	5	0	137.6	3 554.4	0	0
Number of contigs at 25 x min. coverage	{>0}	101	90	109	185	191	199	202	140	212	95	199	193
Number of contigs [0–25] x min. coverage	{<1000}	413	120	91	14	0	9	1	0	6	195	0	0
Average coverage	{>50}	28.6	22.9	28.9	69.3	55.8	31.8	53.7	46.6	80.7	23.9	71.2	84.0
Number of reads (x1000)		789.1	645.2	799.5	1 945	1 554.9	888.6	1 484.3	1 241.1	2 269.3	644.6	2 025.4	2 349.9
Number of trimmed reads (x1000)		789.1	645.2	799.5	1 945	1 554.9	888.6	1 484.3	1 241.1	2 269.3	644.6	2 025.4	2 349.9
Maximum read length		101	101	101	101	101	101	101	101	101	101	101	101
Mean read length		99.4	99.5	99.2	99.3	99.7	99.3	98.9	99.7	99.3	99.1	99.6	99.6
Read insert size		347.4	366.9	350.1	350.1	415.6	355.9	349.2	434.2	347.6	334.3	382.9	370.8
Insert size StdDev		186.2	192.8	186.2	194.9	209.7	196.8	201	209.1	192	188.4	200.9	191.3
N50 (kbp)		34.4	79.1	79.7	79.4	90	79.7	81.9	115.5	78.9	54.4	79.4	91.9
N75 (kbp)		9.2	40.8	37.5	36.4	39.2	39.2	39.2	50.5	39.1	26.1	37.5	38.8

		Laboratory 139											
Parameters	**Ranges***	**9134**	**9051**	**9570**	**9589**	**9197**	**9170**	**9300**	**9543**	**9708**	**9402**	**9375**	**9821**
Number of genera detected	{1}	1	1	1	1	1	1	1	1	1	1	1	1
Detected species	{Ec}	Ec	Ec	Ec	Ec	Ec	Ec	Ec	Ec	Ec	Ec	Ec	Ec
Unclassified reads (%)	{<100}	0.51	0.54	0.76	0.8	1.57	1.5	1.11	0.69	0.74	0.66	0.71	1.33
Length at 25 x min. coverage (Mbp)	{>45 ∧ <53}	5.1	5.1	4.9	4.9	5.3	5.2	4.9	5.1	5	5.2	5.2	5
Length [0–25] x min. coverage (kbp)	{<250}	120.5	73.7	260.9	313	68.1	120.4	210.3	142	157.1	68.4	52.2	41.2
Number of contigs at 25 x min. coverage	{>0}	236	289	293	256	242	281	201	252	258	257	250	278
Number of contigs [0–25] x min. coverage	{<1000}	42	37	79	84	30	43	52	49	51	32	24	25
Average coverage	{>50}	76.7	110.9	60.2	58.3	72.5	74.8	79.8	69.5	79.6	91.8	93.2	85.3
Number of reads (x1000)		1 532.6	2 274.8	1 232.5	1 184.9	1 491.8	1 553.3	1 557	1 378.8	1 650.4	1 876.7	1 883.2	1 719.6
Number of trimmed reads (x1000)		1 518.3	2 254.8	1 221.8	1 174.6	1 478.3	1 538.2	1 542.4	1 366.2	1 636.2	1 860.3	1 865.8	1 703.2
Maximum read length		151	151	151	151	151	151	151	151	151	151	151	151
Mean read length		148.1	148.2	147.8	147.5	148.1	148	148.1	147.8	147.4	148.1	148	147.2
Read insert size		479.5	452.9	421.3	401.6	426.2	416	382.4	415.5	330.7	410.3	359.7	367.4
Insert size StdDev		173.2	161.1	152.3	151.4	170.2	159.7	158.1	156.1	145.3	154	142.9	166.2
N50 (kbp)		59.9	47	49.5	69.7	90.7	81.9	95.5	72.4	52.6	68.2	91.9	92.1
N75 (kbp)		31.1	24.2	26.2	37.4	37.6	32.8	37.5	31.2	30.6	32.3	33.9	41

*Quality assessment made by the EQA-provider in-house quality control pipeline. *: indicative QC ranges; Ec: E. coli.*

Annex 14. 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").

1. STEC EQA-8 2017

Dear Participant

Welcome to the eight External Quality Assessment (EQA-8) scheme for typing of STEC in 2017-2018.

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

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

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

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

Available options in this submission form include:

- Click "Options" and "Pause" to save your results and finish at a later time (using the same link)

- Click "Options" and "Print" to print your answers. This can be done at any time, but before pressing "Submit results"

- Click "Previous" to go back to the questions you have already answered

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

2. Country

(State one answer only)

- ❑ Austria
- ❑ Belgium
- ❑ Czech Republic
- ❑ Denmark
- ❑ Estonia
- ❑ Finland
- ❑ France
- ❑ Germany
- ❑ Greece
- ❑ Hungary
- ❑ Iceland
- ❑ Ireland
- ❑ Italy
- ❑ Latvia
- ❑ Lithuania
- ❑ Luxembourg
- ❑ Norway
- ❑ Portugal
- ❑ Republic of Macedonia
- ❑ Romania
- ❑ Slovenia
- ❑ Spain
- ❑ Sweden
- ❑ The Netherlands
- ❑ Turkey
- ❑ UK

3. Institute name

4. Laboratory name

5. Laboratory ID

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

6. E-mail

7. Serotyping and virulence gene determination of STEC

8. Submitting results

(State one answer only)

❑ Submit serotyping/virulence gene determination results (please fill in the strain ID´s in the next section) - Go to 9

❑ Did not participate in the serotyping nor virulence determination part(s) - Go to 20

9. Serotyping/virulence strain IDs

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

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

10. Serotyping/virulence strain IDs

(State one answer only)

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

11. Results for serotyping (O Group)

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

12. Please specify the method used:

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

Method

❑ Phenotypic
❑ PCR based
❑ WGS based

13. Results for serotyping (H Type)

Please type the number of H Type by using (1-56)
H-: 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	___

14. Please specify the method used:

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

Method

- ❑ Phenotypic
- ❑ PCR based
- ❑ WGS based

15. Submitting results - Virulence gene determination

(State only one answer per question)

- ❑ Submit virulence gene determination data (*eae, aaiC, aggR, stx1a, stx2* or subtyping - Go to 16
- ❑ Submit only virulence gene determination Subtyping data - Go to 18
- ❑ Did not participate in the virulence gene determination (*eae, aaiC, aggR, stx1a, stx2* or subtyping) - Go to 20

16. Results for virulence gene determination

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

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

17. Results for virulence gene determination

(State one answer only)

- ❑ Submit subtyping data
- ❑ Did not participate in subtyping - Go to 20

18. Results for subtyping

Subtyping of *stx1,* please select the variant (*stx1a, stx1c, stx1d*)
Not done: ND

(State only one answer per question)

	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	❑	❑	❑	❑	❑	❑	❑	❑

19. Subtyping of *stx2*

Please select the variant *(stx2a, stxb, stx2c, stx2d, stx2e, stx2f, stx2g)*
Not done: ND

(State only one answer per question)

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

20. Submitting cluster analysis results

(State one answer only)

❑ Cluster analyses based on PFGE and/or WGS - Go to 21

❑ Did not participate in the Cluster part - Go to 134

21. Cluster strain ID´s

please enter the cluster strain ID (4 digits)

We recommend to print this page out! To have the overview of strain ID's and strain No. 1-12, it will make the work easier.

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

22. Submitting cluster analysis results

(State one answer only)

- ❑ Cluster analysis based on PFGE - Go to 23
- ❑ Do not wish to submit any cluster results based on PFGE analysis - Go to 28

23. Cluster analysis based on PFGE data

24. Please list the ID for the strains included in the cluster of closely related strains detected by PFGE results:

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

25. Select a representative strain with the cluster profile detected by PFGE:

Indicate the strain ID

26. XbaI - Total number of bands (>33kb) in the selected representative cluster strain

27. Results for cluster analysis - PFGE (XbaI)

Please use 9999 for not analysed

	XbaI - Total number of bands (>33kb)	XbaI - Number of bands with same/shared position as the profile of the selected cluster strain (>33kb)
Strain 1		
Strain 2		
Strain 3		
Strain 4		
Strain 5		
Strain 6		
Strain 7		
Strain 8		
Strain 9		
Strain 10		
Strain 11		
Strain 12		

28. Submitting cluster results

(State one answer only)

- ❑ Cluster analysis based on WGS data - Go to 23
- ❑ Do not wish to submit any cluster results based on WGS data - Go to 134

29. Cluster analysis based on WGS data

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

As basis for the cluster detection, only one data analysis can be reported. If more than one analysis is performed, please report later in this submission

(State one answer only)

- ❑ SNP based - Go to 32
- ❑ Allele based - Go to 38
- ❑ Other - Go to 31

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

- Go to 44

32. Please select the approach used for the SNP analysis

(State one answer only)

- ❑ Reference based - Go to 33
- ❑ Assembly based - Go to 36

33. Reference genome used

Please indicate multilocus sequence type (e.g. ST8) and strain ID (e.g. one of the strains from the current cluster, ID of a public reference strain or an in-house strain)

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

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

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

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

38. Please select tools used for the allele analysis

(State one answer only)

- ❑ BioNumerics - Go to 40
- ❑ Allele-based - Go to 40
- ❑ Other - Go to 40
- ❑ Allele-based - Go to 39

39. If another tool is used, please enter here:

40. Please indicate allele calling method:

(State one answer only)

- ❑ Assembly based and mapping based - Go to 41
- ❑ Only assembly based - Go to 41
- ❑ Only mapping based - Go to 42

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

42. Please select scheme used for the allele analysis

(State one answer only)

- ❑ Applied Math (wgMLST) - Go to 44
- ❑ Applied Math (cgMLST/Enterobase) - Go to 44
- ❑ Enterobase (cgMLST) - Go to 44
- ❑ Other - Go to 43

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

44. Cluster detected by analysis on data derived from WGS

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

45. Please list the ID´s for the strains included in the cluster

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

46. Select a representative strain in the cluster

Indicate the strain ID

47. Results for cluster analysis (e.g. SNP or allele based)

Please use 9999 for not analysed

	ST	Distance (e.g. SNP) to the selected cluster strain
Strain 1	_____	_____
Strain 2	_____	_____
Strain 3	_____	_____
Strain 4	_____	_____
Strain 5	_____	_____
Strain 6	_____	_____
Strain 7	_____	_____
Strain 8	_____	_____
Strain 9	_____	_____
Strain 10	_____	_____
Strain 11	_____	_____
Strain 12	_____	_____

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 84

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

(State one answer only)

- ❑ SNP based - Go to 51
- ❑ Allele based - Go to 57
- ❑ Other - Go to 50

50. If another analysis is used, please describe your approach

- Go to 63

51. Please select the approach used for the SNP analysis

(State one answer only)

- ❑ Reference based - Go to 52
- ❑ Assembly based - Go to 55

52. Reference genome used

Please indicate multilocus sequence type (e.g. ST8) and strain ID (e.g. one of the strains from the current cluster, ID of a public reference strain or an in-house strain)

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

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

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

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

57. Please select tool used for the allele analysis

(State one answer only)

- ❑ BioNumerics - Go to 59
- ❑ SeqPhere - Go to 59
- ❑ Enterobase - Go to 59
- ❑ Other - Go to 58

58. If another tool is used, please list here:

59. Please indicate allele calling method

(State one answer only)

- ❑ Assembly based and mapping based - Go to 60
- ❑ Only assembly based - Go to 60
- ❑ Only mapping based - Go to 61

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

61. Please select scheme used for the allele analysis

(State one answer only)

- ❑ Applied Math (wgMLST) - Go to 63☐
- ❑ Applied Math (cgMLST/Enterobase) - Go to 63
- ❑ Enterobase (cgMLST) - Go to 63
- ❑ Other - Go to 62

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

63. Additional analysis on data derived from WGS

64. Select a representative strain in the cluster detected by the additional analysis

Indicate the strain ID

65. Results for the additional cluster analysis (e.g. SNP or allele based)

Please use 9999 for not analysed

	ST	Distance (e.g. SNP) to selected cluster strain
Strain 1	_____	_____
Strain 2	_____	_____
Strain 3	_____	_____
Strain 4	_____	_____
Strain 5	_____	_____
Strain 6	_____	_____
Strain 7	_____	_____
Strain 8	_____	_____
Strain 9	_____	_____
Strain 10	_____	_____
Strain 11	_____	_____
Strain 12	_____	_____

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

e.g. if SNP based results are submitted, you can also report allele based results or results from a second SNP analysis

(State one answer only)

- ❑ Yes - Go to 67
- ❑ No - Go to 84

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

(State one answer only)

- ❑ SNP based - Go to 69
- ❑ Allele based - Go to 75
- ❑ Other - Go to 68

68. If another analysis is used, please describe your approach:

- Go to 81

69. Please select the approach used for the SNP analysis

(State one answer only)

- ❑ Reference based - Go to 70
- ❑ Assembly based - Go to 73

70. Reference genome used

Please indicate multilocus sequence type (e.g. ST8) and strain ID (e.g. one of the strains from the current cluster, ID of a public reference strain or an in-house strain)

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

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

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

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

75. Please select tool used for the allele analysis

(State one answer only)

- ❑ BioNumerics - Go to 77
- ❑ SeqPhere - Go to 77
- ❑ Enterobase - Go to 77
- ❑ Other - Go to 76

76. If another tool is used, please enter here:

77. Please indicate allele calling method:

(State one answer only)

- ❑ Assembly based and mapping based - Go to 78
- ❑ Only assembly based - Go to 78
- ❑ Only mapping based - Go to 79

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

79. Please select scheme used for the allele analysis

(State one answer only)

- ❑ Applied Math (wgMLST) - Go to 81
- ❑ Applied Math (wgMLST/Enterobase) - Go to 81
- ❑ Enterobase (cgMLST) - Go to 81
- ❑ Other - Go to 80

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

81. Third analysis on data derived from WGS

82. Select a representative strain in the cluster detected by the third analysis

Indicate the strain ID

83. Results for the third cluster analysis (e.g. SNP or allele based)

Please use 9999 for not analysed

	ST	Distance (e.g. SNP) to selected cluster strain
Strain 1		
Strain 2		
Strain 3		
Strain 4		
Strain 5		
Strain 6		
Strain 7		
Strain 8		
Strain 9		
Strain 10		
Strain 11		
Strain 12		

84. Additional questions to the WGS part

85. Where was the sequencing performed

(State one answer only)

- ❑ In own laboratory
- ❑ Externally

86. Protocol used to prepare the library for sequencing:

(State one answer only)

- ❑ Commercial kits - Go to 87
- ❑ Non-commercial kits - Go to 89

87. Please indicate name of commercial kit:

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

- Go to 90

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

90. The sequencing platform used

(State one answer only)

- ❑ Ion Torrent PGM - Go to 92
- ❑ Ion Torrent Proton - Go to 92
- ❑ Genome Sequencer Junior System (454) - Go to 92
- ❑ Genome Sequencer FLX System (454) - Go to 92
- ❑ Genome Sequencer FLX+ System (454) - Go to 92
- ❑ PacBio RS - Go to 92
- ❑ PacBio RS II - Go to 92
- ❑ HiScanSQ - Go to 92
- ❑ HiScanSQ 1000 - Go to 92
- ❑ HiScanSQ 1500 - Go to 92
- ❑ HiScanSQ 1000 - Go to 92
- ❑ HiScanSQ 2000 - Go to 92
- ❑ HiScanSQ 2500 - Go to 92
- ❑ HiScanSQ 4000 - Go to 92
- ❑ Genome Analyzer Iix - Go to 92
- ❑ MiSeq - Go to 92
- ❑ MiSeq Dx - Go to 92
- ❑ MiSeq FGx - Go to 92
- ❑ ABI SOLiD - Go to 92
- ❑ NextSeq - Go to 92
- ❑ MinION (ONT) - Go to 92
- ❑ Other - Go to 91

91. If another platform is used, please list here:

92. Quantitative criteria used to evaluate the quality of sequence data.

Please list up to 10 different criteria (e.g. coverage, N50, number of contigs)

93. Quantitative criteria 1:

94. Threshold used for quantitative criteria 1:

95. Quantitative criteria 2:

96. Threshold used for quantitative criteria 2:

97. Quantitative criteria 3:

98. Threshold used for quantitative criteria 3:

99. Quantitative criteria 4:

100. Threshold used for quantitative criteria 4:

101. Quantitative criteria 5:

102. Threshold used for quantitative criteria 5:

103. Quantitative criteria 6:

104. Threshold used for quantitative criteria 6:

105. Quantitative criteria 7:

106. Threshold used for quantitative criteria 7:

107. Quantitative criteria 8:

108. Threshold used for quantitative criteria 8:

109. Quantitative criteria 9:

110. Threshold used for quantitative criteria 9:

111. Quantitative criteria 10:

112. Threshold used for quantitative criteria 10:

113. Qualitative criteria used to evaluate the quality of sequence data.

Please list up to 10 different criteria (e.g. contamination, confirmation of genus)

114. Qualitative criteria 1:

115. If relevant threshold used for qualitative criteria 1:

116. Qualitative criteria 2:

117. If relevant threshold used for qualitative criteria 2:

118. Qualitative criteria 3:

119. If relevant threshold used for qualitative criteria 3:

120. Qualitative criteria 4:

121. If relevant, threshold used for qualitative criteria 4:

122. Qualitative criteria 5:

123. If relevant, threshold used for qualitative criteria 5:

124. Qualitative criteria 6:

125. If relevant, threshold used for qualitative criteria 6:

126. Qualitative criteria 7:

127. If relevant, threshold used for qualitative criteria 7:

128. Qualitative criteria 8:

129. If relevant, threshold used for qualitative criteria 8

130. Qualitative criteria 9

131. If relevant, threshold used for qualitative criteria 9

132. Qualitative criteria 10:

133. If relevant, threshold used for qualitative criteria 10:

134. Comment(s):

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

135. Thank you for your participation

Thank you for filling out the Submission form for the STEC EQA-8.

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

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

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

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

European Centre for Disease Prevention and Control (ECDC)

Gustav III:s Boulevard 40, 16973 Solna, Sweden

Tel. +46 858601000
Fax +46 858601001
www.ecdc.europa.eu

An agency of the European Union
www.europa.eu

Subscribe to our publications
www.ecdc.europa.eu/en/publications

Contact us
publications@ecdc.europa.eu

Follow us on Twitter
@ECDC_EU

Like our Facebook page
www.facebook.com/ECDC.EU

ECDC is committed to ensuring the transparency and independence of its work

In accordance with the *Staff Regulations for Officials and Conditions of Employment of Other Servants of the European Union* and the *ECDC Independence Policy*, ECDC staff members shall not, in the performance of their duties, deal with matters in which they may, directly or indirectly, have a personal interest that could impair their independence. Declarations of interest must be received from any prospective contractor before a contract can be awarded.
www.ecdc.europa.eu/en/aboutus/transparency

Publications Office