External Quality Assessment of Molecular typing of Shiga toxin-producing *Escherichia coli* 2022

Susanne Schjørring and Nadia Boisen M.Sc, Ph.D

Statens Serum Institut
Sektion of foodborne infection

Background

- 2021: 6 534 confirmed cases of STEC infection were reported by 30 EU/EEA countries
- The EU/EEA notification: 2.2 cases per 100 000 population
- 37.5% increase compared with the previous year
- The five most frequently reported serogroups:
  - O157 (15.1%), O26 (14.7%), O103 (8.4%), O145 (4.6%), O146 (3.7%)
- Among HUS cases the most frequently reported serogroups:

  O26 (34%) and O157 (19.8%)
- The proportion of cases where no serotype could be retrieved was 25.9%
- Antigen H was reported for 2 496 confirmed cases (38.2%).

Molecular typing EQA

Funded:

By European Centre for Disease Prevention and Control (ECDC)

Organised:

Statens Serum Institut, Denmark, Section of Foodborne Infections

•2012-2016

•2017-2020

•**2022-2025**

Main objective of the EQAs scheme:

- Assess the general standard of performance (‘state-of–the-art’)
- Assess the effects of analytical procedures (method principle, instruments, reagents, calibration)
- Support method development
- Evaluate individual laboratory performance
- Identify problem areas
- Provide continuing education
- Identify needs for training activities

Methods and study design EQA-11

STATENS SERUM INSTITUT

Serotyping:

- Included 12 test strains

Virulence gene detection:

- Included 12 test strains

Molecular typing-based cluster analysis:

- WGS-derived
  - 12 test strain+8 provided FASTQ sequences
  - Detect cluster of closely related strains
  - QC observations and QC-status decision

Method	Serotyping		Virulence profile		Cluster analysis			
No. strains/sequences	12 strains		12 strains		12 strains / 8 sequences			
Strain ID						ST	QC-status	Cluster
Strain1	Strains for Serotyping	O26:H11	Strains for virulence profile	stx1a, stx2a, eae	Strains/sequences for cluster analysis	21	-	
Strain2		O187:H28		stx2g, esta		200	-	
Strain3#‡		O157:H-/H7		stx1a, eae		11	-	Yes
Strain4		O177:H-/H25		stx2a, stx2c*		342	-	
Strain5		O91:H14		stx1a, stx2b		33	-	
Strain6		O80:H2		stx2d, eae		301	-	
Strain7#‡		O157:H-/H7		stx1a, eae		11	-	Yes
Strain8		O157:H-/H7		stx1a, stx2c, eae		11	-	
Strain9		O128:H-/H2		stx2f, eae		20	-	
Strain10		O145:H-/H28		stx2a, eae		32	-	
Strain11		O146:H21		stx1c, stx2b		442	-	
Strain12		O104:H4		aggR		678	-	
Strain13^	-	O157:H7		stx1a, eae		11	C	NA
Strain14	-	O157:H7		stx1a, stx2c, eae		11	A	
Strain15‡	-	O157:H7		stx1a, stx2c, eae		11	A	Yes
Strain16	-	O157:H7		stx1a, stx2c, eae		11	A	
Strain17‡	-	O157:H7		stx1a, stx2c, eae		11	A	Yes
Strain18	-	O157:H7		stx1a, stx2c, eae		11	A	
Strain19^	-	-		-		-	B/C	NA
Strain20#‡	-	O157:H7		stx1a, eae		11	A	Yes

Participation

- 27 laboratories signed up
- 26 completed and submitted results
- 20/26 submitted both serotyping, virulence determination and cluster analysis

	Serotyping[1]	Virulence profile determination[2]	Cluster analysis[3]
Number of participants	25	25	20
% of participants	96*	96*	77*

Participation (continue)

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

	Serotyping		Virulence profile determination					Cluster analysis
	n=25		n=25					n=20
	O group	H type	*aggR*	*eae*	*esta*	*stx1* and *stx2*	*stx* subtyping	WGS
Number of participants	25[#]	19Δ	22	24	19	25	22	20
Percentage of participants^	100%	76%	88%	96%	76%	100%	88%	100%
Percentage of participants *	96%	73%	85%	92%	73%	96%	85%	77%

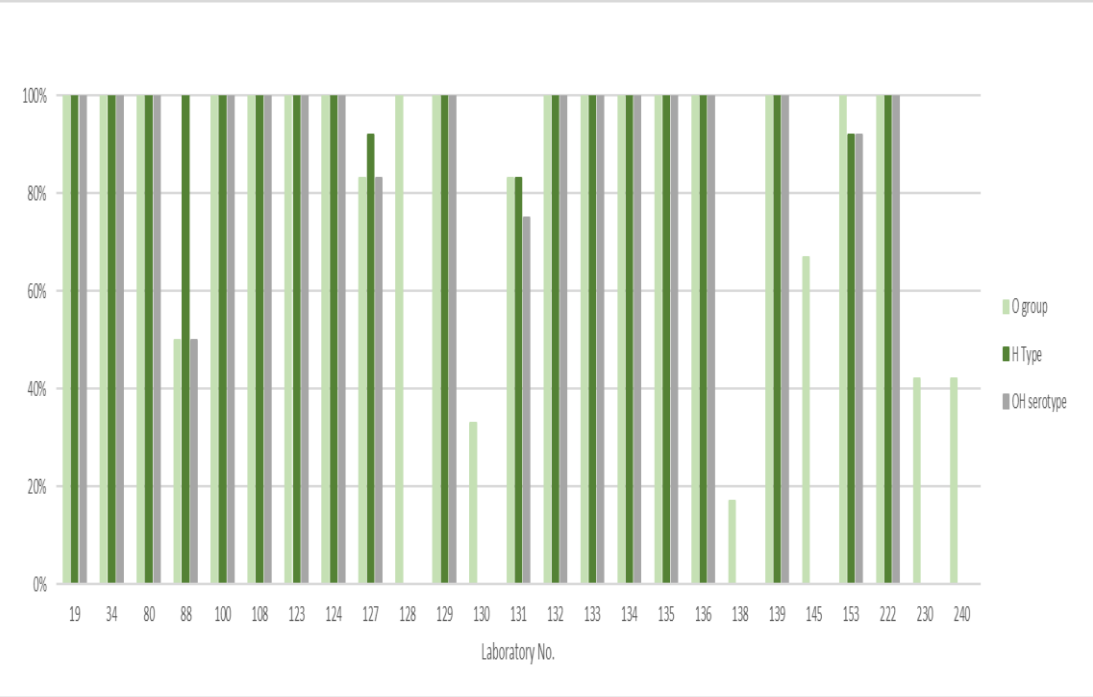
^: percentage of participants in respective part of EQA
**: percentage of total number of participating laboratories (26)*
[#]: phenotypic (n=9)/PCR-based (n=1)/WGS-based (n=15)
Δ: phenotypic (n=2)/PCR-based (n=1)/WGS-based (n=16)

Results: serotyping

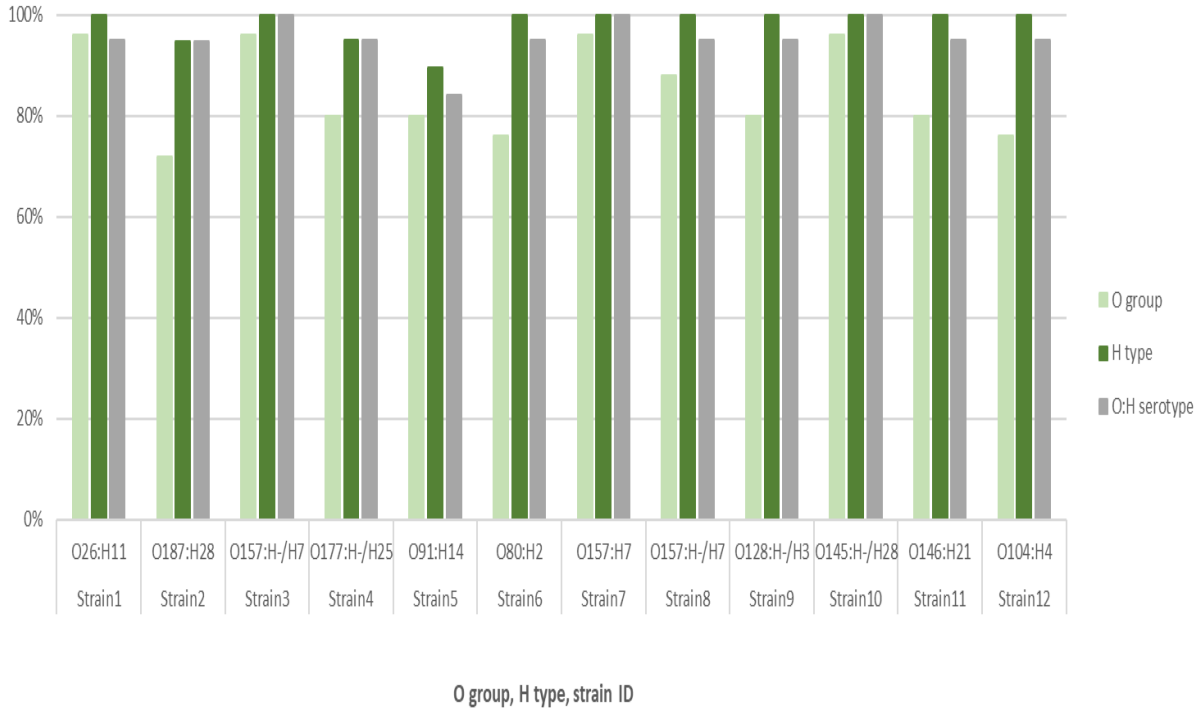
- O grouping:
  - 25 participants
  - 17/25 correctly serotyped all 12 strains
  - 85% average performance score

- H typing:
  - 19 participants
  - 16/19 correctly serotyped all 12 strains
  - 98% average performance score

Participant percentage scores for O grouping

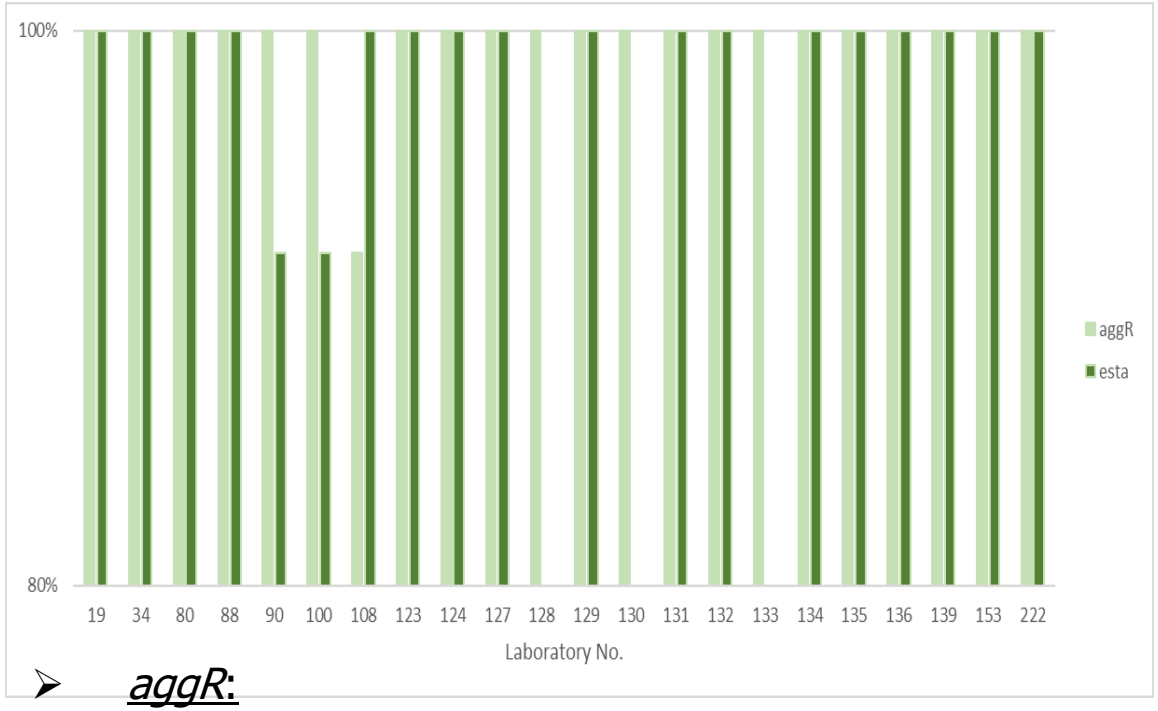


Average percentage test strain score for O and H



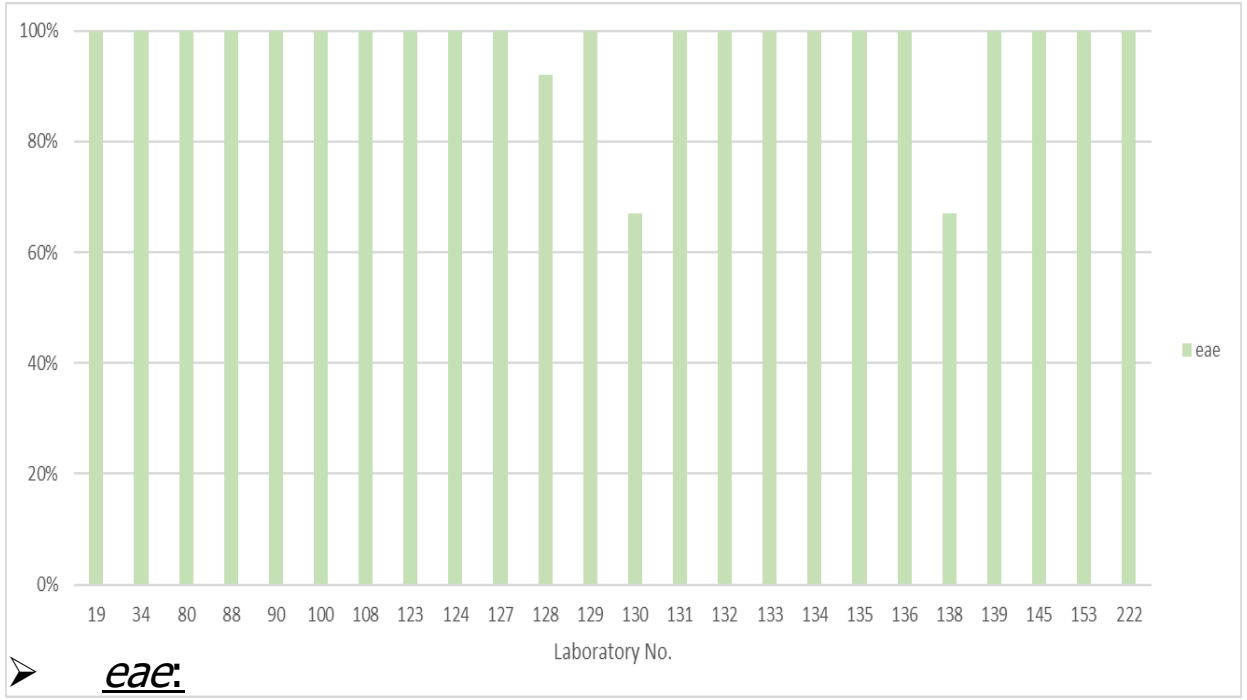
Results: virulence profile determination

Detection of the EAEC (*aggR*), ETEC (*esta*), and *eae*



- *aggR*:
  - 22 participants
  - 21/22 correctly identified all 12 strains
  - 100% average performance

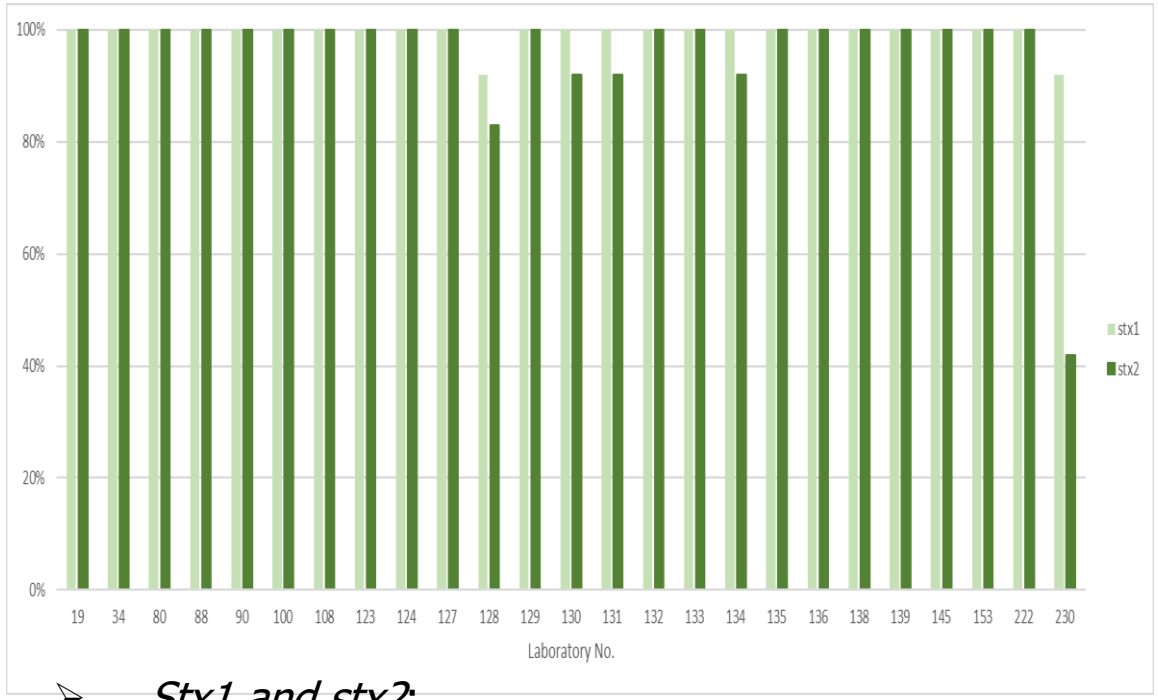
- *esta*:
  - 19 participants
  - 17/19 correctly identified all 12 strains
  - 99% average performance



- *eae*:
  - 24 participants
  - 21/24 correctly identified all 12 strains
  - 97% average performance

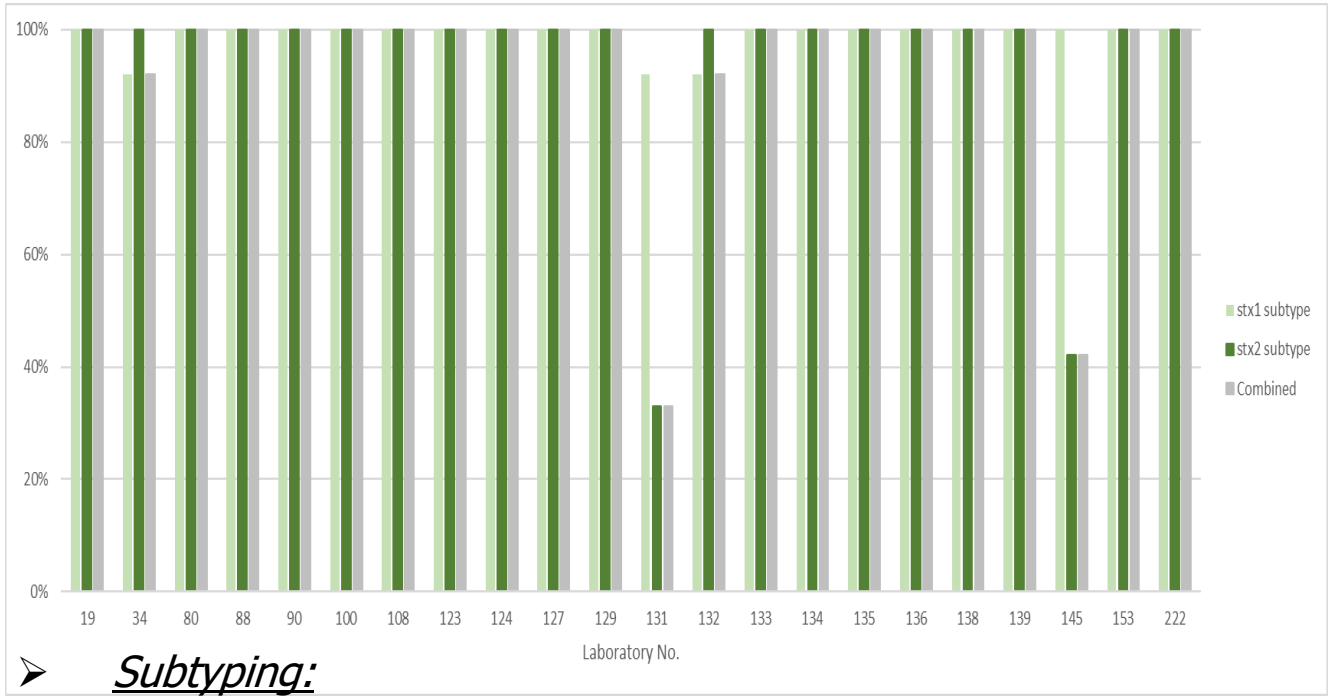
Results: virulence profile determination

Detection of virulence genes *stx1* and *stx2* and subtyping



➢ *Stx1 and stx2*:

- 25 participants
- *Stx1*: 23/25 correctly identified all 12 strains
- *Stx2*: 20/25 correctly identified all 12 strains
- Average performance
  Stx1: 99% *Stx2*: 96%



➢ *Subtyping:*

- 22 participants
- *Stx1 sub*: 19/22 correctly identified all 12 strains
- *Stx2 sub*: 20/22 correctly identified all 12 strains
- Combined: 18/22 correctly identified all 12 strains
- Average performance
  Stx1: 99%, *stx2*: 94% combined: 93%

Results: virulence profile determination

- **Subtyping of *stx2* (continue)**

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

* strain4 (*stx2a* and *stx2c*) was disregard, ND#: not done.

Results: cluster analysis

➢ WGS-derived data (no PFGE)

- 20 laboratories
- 12 test strains and 8 additional strains (genomic sequences FASTQ files

Method	Cluster analysis			
No. strains/sequences	12 strains / 8 sequences			
Strain ID		ST	QC-status	Cluster
Strain1	Strains/sequences for cluster analysis	21	-	
Strain2		200	-	
Strain3#‡		11	-	Yes
Strain4		342	-	
Strain5		33	-	
Strain6		301	-	
Strain7#‡		11	-	Yes
Strain8		11	-	
Strain9		20	-	
Strain10		32	-	
Strain11		442	-	
Strain12		678	-	
Strain13^		11	C	NA
Strain14		11	A	
Strain15‡		11	A	Yes
Strain16		11	A	
Strain17‡		11	A	Yes
Strain18		11	A	
Strain19^		-	B/C	NA
Strain20#‡		11	A	Yes

- EQA provider: found at most two allele differences or five SNPs between any two strains in the cluster.
- All downloaded sequences should be QC evaluated and included in an analysis with the own produced WGS data
- *#‡: closely related strains*
- *#: technical triplicates strains*
- *ST: sequence type*
- *^modified sequences:*
  - *strain13, a non-cluster sequence with reduced coverage and removal of genes,*
  - *strain19, a non-cluster sequence contaminated with approx. 14% E. albertii*
- *NA: Not applicable*
- *A: Acceptable quality*
- *B: Quality only acceptable for outbreak situations (less good quality)*
- *C: Not acceptable quality - strain not analysed*

Results: cluster analysis

- **WGS-derived data**
  - Sequencing details:
    - 19 labs used their own laboratory, only one used an externally sequcing facility
    - Mix of platforms: 1 MiniSeq, 8 MiSeq, 7 NextSeq, 2 Novaseq, 1 Ion GeneStudio S5 System and 1 Ion Torrent
    - Commercial kits for library preparation
    - 16/20 Illumina's Nextera kit
    - Four participant listed changes volume from the manufactory protocol

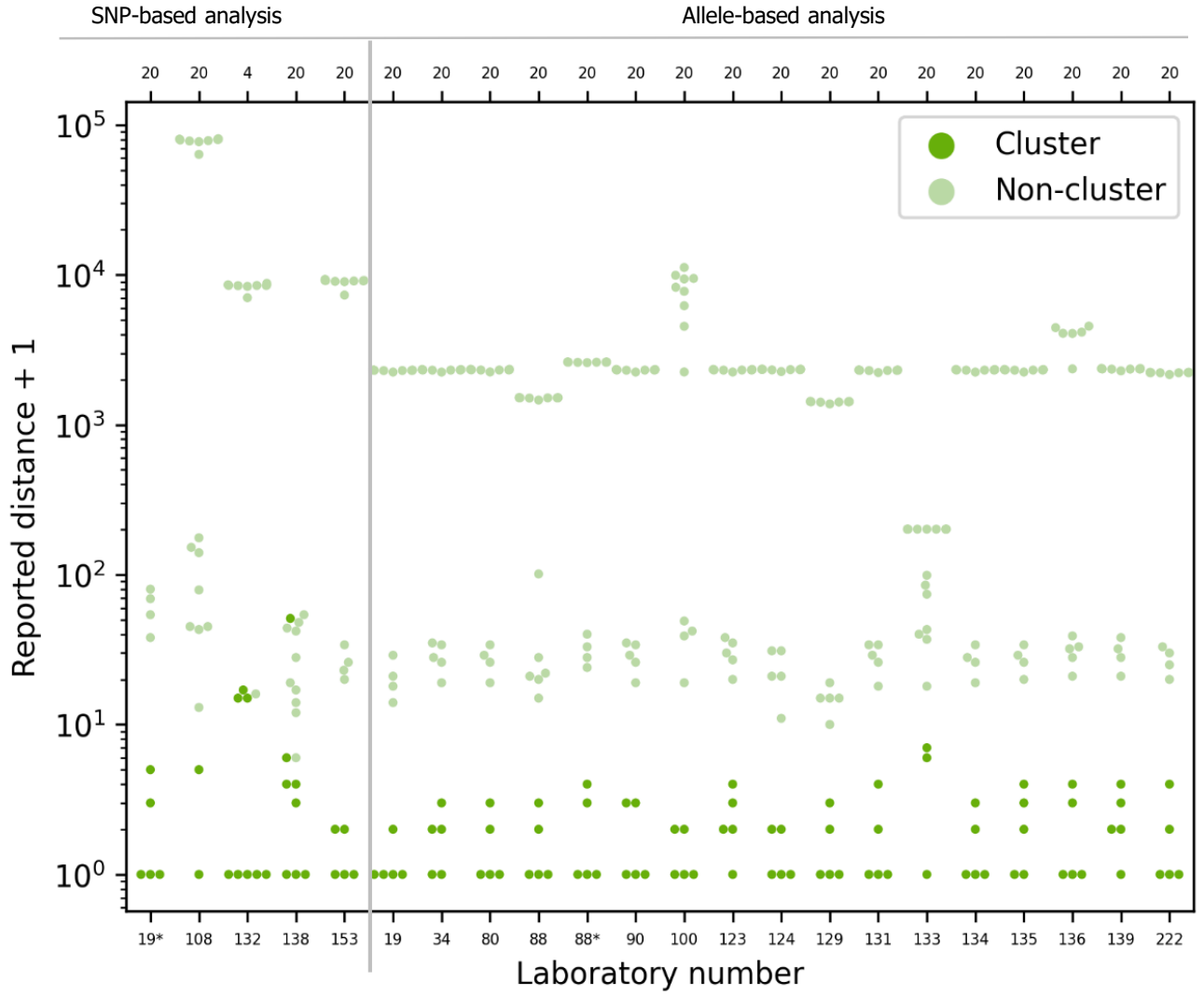
Results: cluster analysis (continue)

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

Additional analysis: [a] = SNP based, [b] = single-nucleotide variant (SNV based), [c] = Allele based

Results: cluster analysis (continue)

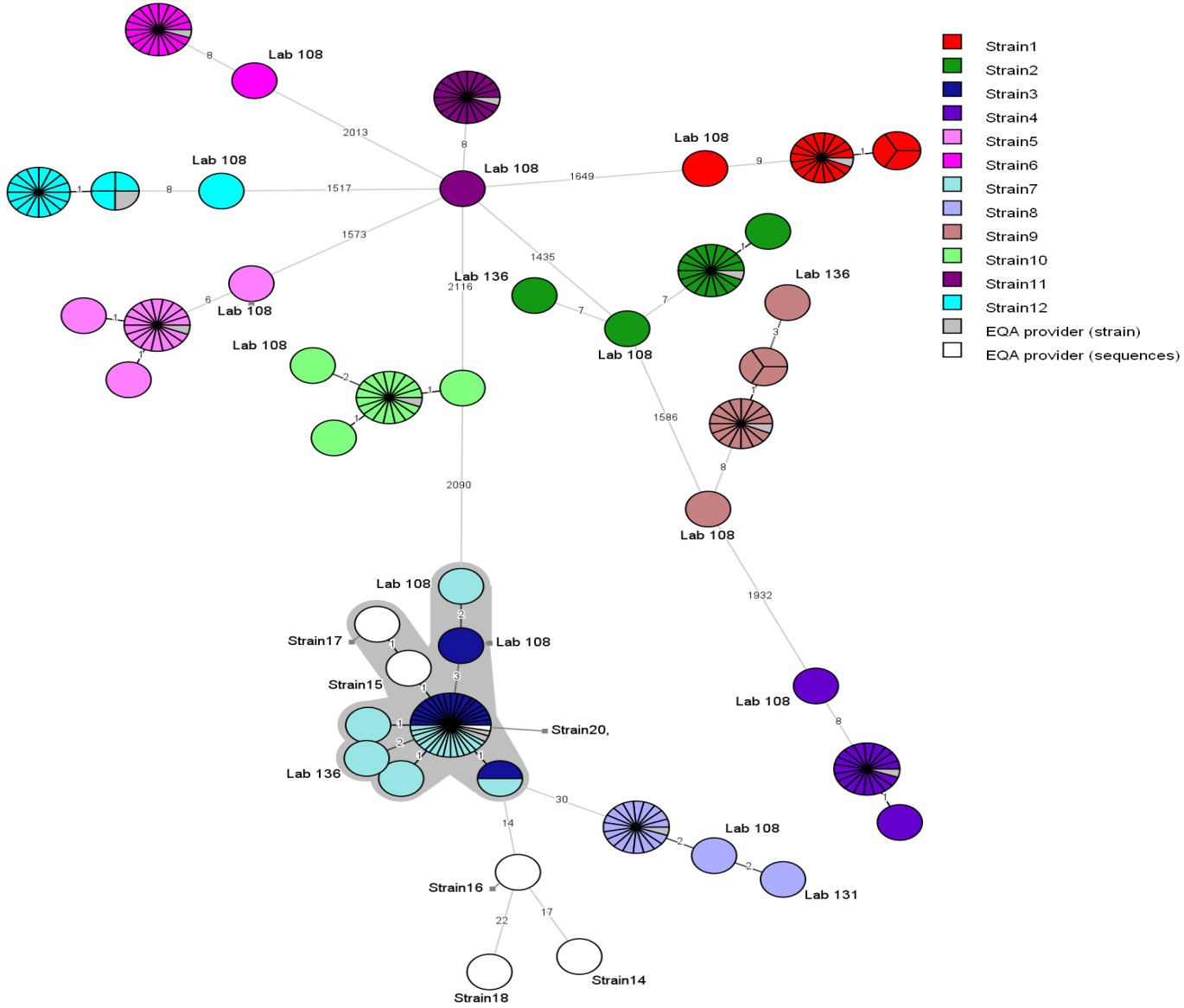
Reported SNP distances or allelic differences for each test strain to selected cluster representative strain



- Participants were instructed to select Strain20 as reference (listed as "20" on the top scale).
- Dark green: reported cluster of closely related isolates
- Light green: not reported as part of cluster
- Only one of four laboratories identified the correct cluster when using SNP analyses, reported SNP with a maximum of 0-–1 SNP distances

Results: cluster analysis (continue)

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



- Each of the strain1–12 test strains have a different colour.
- Grey: REF results from the EQA-provider
- White: provided sequences
- Technical triplicates: Strain3/strain7 and strain/sequence20

FASTQ evalutated by EQA provider QC-pipeline

Ranges*	{Ec}			{5%}	{4.5-5.8}	{<250}	{>0}	{<1000}	{>50}					
Lab No.	Detected species	Species 1 (%)	Species 2 (%)	Unclassified reads (%)	Length at >25 x min. coverage (Mbp))	Length [1-25] x min. coverage (kbp)	Number of contigs at 25 x min. coverage	No. of contigs [1-25] x min. coverage	Average coverage	Number of reads (x1000)	Average read length	Average insert size	N50 (kb)	QC status (Bifrost)
19	Ec	81.1-93.7	0.3-3.1	5.4-14.2	5.0-5.4	26.1-123.1	341.0-709.0	36.0-148.0	58.0-91.0	2308.0-3443.0				
34	Ec, Pt	72.7-88.7	5.0-11.1	5.7-13.9	5.0-5.5	0.0-0.0	173.0-451.0	0.0-0.0	143.0-319.0	6500.0				Warning
80	Ec, Pt	72.5-92.6	1.9-7.7	4.6-17.8	5.1-5.6	0.0-127.1	88.0-304.0	0.0-8.0	67.0-10					Warning
88	Ec	77.8-95.9	0.4-2.4	2.8-17.9	5.0-5.6	0.0-58.3	183.0-396.0	0.0-5						Warning
90	Ec	84.7-95.9	0.1-2.5	2.9-12.0	5.0-5.6	0.6-56.7	21						25.0-62.0	
100	Ec	85.7-98.1	0.0-2.4	1.7-10.9	5.0-5.6	0						275.0-316.0	89.0-181.0	
108#	Ec	89.6-97.9	0.4-1.6	1.4-7.1							238.0-305.0	0.0-0.0	2.0-12.0	
123	Ec	62.7-97.2	0.2-2.0							1704.0-2455.0	239.0-260.0	275.0-327.0	53.0-140.0	Warning
124	Ec	90.1-98.1							258.0-288.0	6000.0-6000.0	251.0-251.0	400.0-416.0	91.0-199.0	
129	Ec							3.0-39.0	74.0-166.0	2690.0-6347.0	144.0-149.0	296.0-387.0	24.0-106.0	
131							89.0-321.0	0.0-0.0	132.0-160.0	4758.0-5982.0	148.0-149.0	263.0-342.0	72.0-168.0	
132						0.0-2257.1	89.0-352.0	0.0-66.0	26.0-201.0	927.0-7833.0	138.0-148.0	224.0-492.0	40.0-166.0	Warning
133	E				5.1-5.6	0.0-0.0	79.0-262.0	0.0-0.0	101.0-187.0	1965.0-3820.0	279.0-289.0	337.0-402.0	86.0-181.0	Warning
134	Ec			1.7-16.4	5.1-5.6	0.0-3.4	93.0-294.0	0.0-3.0	38.0-95.0	1362.0-3459.0	144.0-148.0	294.0-325.0	72.0-180.0	Warning
135	Ec		0.1-2.1	3.1-25.1	5.1-5.6	0.0-0.0	106.0-331.0	0.0-0.0	130.0-223.0	5092.0-7873.0	149.0-149.0	300.0-319.0	72.0-160.0	Warning
136	Ec	86.3-97.1	0.2-3.2	2.3-8.7	5.1-5.6	0.0-0.0	74.0-254.0	0.0-0.0	219.0-1421.0	8638.0-51514.0	146.0-149.0	326.0-492.0	90.0-169.0	
138	Ec	73.0-98.4	0.1-2.3	1.4-23.6	5.1-5.6	0.0-0.0	91.0-344.0	0.0-0.0	540.0-906.0	19386.0-31809.0	148.0-150.0	236.0-347.0	86.0-166.0	Warning
139	Ec	88.0-96.1	0.2-3.2	3.2-6.9	5.0-5.6	0.0-13.9	216.0-480.0	0.0-13.0	132.0-356.0	4853.0-13157.0	143.0-143.0	356.0-429.0	34.0-81.0	Warning
153	Ec	63.6-97.6	0.1-2.2	1.9-33.7	0.3-5.5	0.0-5124.7	99.0-280.0	0.0-229.0	32.0-50.0	1161.0-1742.0	148.0-149.0	310.0-356.0	59.0-157.0	Warning
222	Ec	90.4-97.5	0.2-2.7	2.0-5.8	5.1-5.6	0.0-0.0	81.0-281.0	0.0-0.0	219.0-586.0	7798.0-20849.0	151.0-151.0	337.0-375.0	84.0-166.0	

Warnings:
"Average coverage" is below 50.
"% unclassified" value above threshold (20%)
"% Species 1" + "% unclassified" below 95% (conatmination with *Shigella flexneri* or *Pseudomonas tolaasii*)

Results: cluster analysis (continue)

QC observation and status for strain 13-20 (Genomic sequences):

- A = Acceptable quality
- B = Quality only acceptable for outbreak situations (less good quality)
- C = Not acceptable quality - strain not analyzed

Modified strains:

- Strain13
- Strain19

Results of the participants' QC assessment of the EQA modified provided sequences

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

Summary

Participation:

- 26 Laboratories participated
- 25 (96%) performed serotyping
- 25 (96%) performed virulence gene determination
- 20 (77%) performed cluster analysis

Serotyping

- O group: 85% average performance, H typing: 98% and O:H serotyping:95%
  - O187:H28, O80:H2, and O91:H14 hard to identify phenotypicly

Virulence profile determination

- *aggR*: 100%, always had a high performance
- *esta*: 98%, high performance eventhough it is a new gene in the EQA scheme
- *eae:* 97%, always had a high performance score above 96% since EQA-4
- Stx1*:* 99%, always had a high performance
- Stx2*:* 96%, lways had a high performance
- Stx Subtyping*:* 99% (stx1), 94%* (*stx2) and* 93%* combined (disregarding strain4)

Summary (continue)

Cluster analysis

- Only WGS analysis - no PFGE
- Some QC warnings of the submitted FASTQ files from the participant, overall acceptable quality

Modified sequences (QC issues):

- All of the 20 laboratories did not in some degree identified the massive reduced coverage
  - 16 Not acceptable quality, 4 acceptable for outbreak situations
- 3/20 did not identified the contamination with approx. 14% *E. albertii*
  - Some laboratories need to include a contamination cheek

Cluster of closely related:

- 16/20 (80%) correctly identified the cluster
  - 3 Laboratories using SNP and 1 using Allele based analysis did not identify the correct cluster (incl. 2 new user of WGS)

- 16/20 (80%) used allele based
- 4/20 (20%) used SNP based

Two main challenges:

- Difficulty in comparing SNP with cgMLST
- Variations between SNP analyses

Evaluation of the EQA scheme

Results of evaluation of the EQA scheme

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

Suggestions for improvements

- ECDC´s EQAs more evenly distributed around the year
- guideline with standardized QC criteria?
- cut-off' discussion?
- detection from a clinical specimen?
- make the cluster analysis (WGS) a bit more challenging?