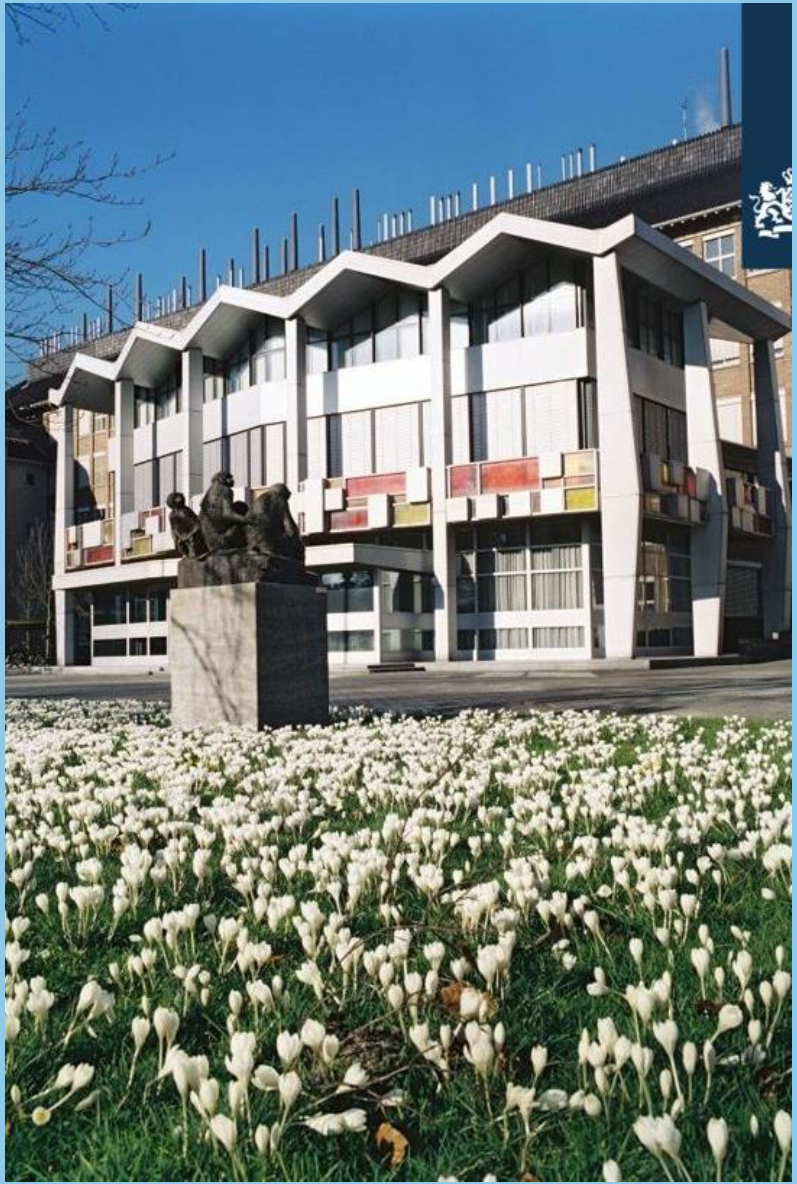


National Institute for Public Health
and the Environment
Ministry of Health, Welfare and Sport

Organisation of NGS PTs: EURL-*Salmonella*

EURL Salmonella

Angela van Hoek & Wilma Jacobs

EURL CPS
European Union Reference Laboratory
Coagulase Positive Staphylococci
http://eurl-staphylococci.anses.fr

European Union Reference Laboratory
Foodborne Viruses

European Union Reference Laboratory
Antimicrobial Resistance

European Union Reference Laboratory for Parasites

EURL Salmonella

EURL Lm
European Union Reference Laboratory for
Listeria monocytogenes
http://eurl-listeria.anses.fr

EURL Campylobacter

EU-RL VTEC

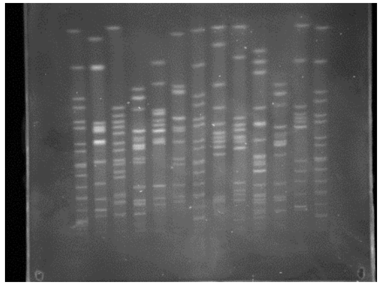
Inter-EURLs Working Group on NGS

Organisation of NGS PTs

- Simulation of an outbreak-related request from the EURL-*Salmonella* (EFSA/ECDC) to the NRL-network
- *Defined* cluster analysis on 6 wet and 6 dry *Salmonella* strains, with a NGS reference sequence provided
    - Wet lab/dry lab
    - Collection of results
    - Results analysis and report
    - Performance assessment
    - Lessons learnt

History EURL-*Salmonella* PTs Typing

- **Serotyping (obligatory for EU MSs)**
  - Started in 1995, yearly 1 set of 20 strains, 27th PT in 2022, **ongoing**
- AMR (optional)
  - 2003, 2004, 2005; in cooperation with WUR-Lelystad
- Phage typing (optional)
  - 1999-2014, yearly 10x SE plus 10x ST; in cooperation with PHE
- PFGE typing (optional)
  - 2013-2018, yearly 1 set of 10 or 11 strains; in cooperation with SSI
- **Cluster Analysis (optional, [PFGE]/[MLVA]/WGS)**
  - Started in 2019, yearly 1 set of 10-12 strains (wet/dry), 5th PT in 2023, **ongoing**



PTs Cluster Analysis: methods

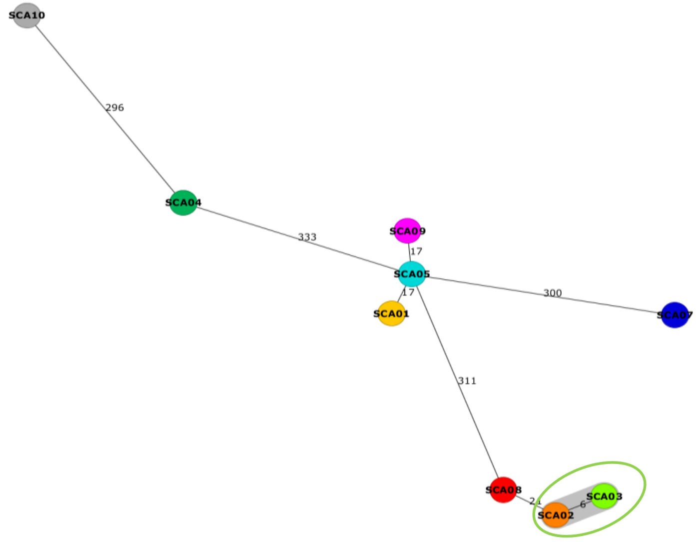
- Cluster analysis using (PFGE and/or MLVA and/or) **WGS**
  - Participants' own routine method(s) of choice
  - WGS PT 2023 scheduled for November

	2019	**2020**	**2021**	**2022**	**2023**
PFGE	6	2	na	na	na
MLVA	8	6	5	5	na
WGS	14	21	19	20	?
WGS: cgMLST	9	15	13	17	?
WGS: SNP	5	8	10	10	?
Partipants overall*	18	21	19	20	?

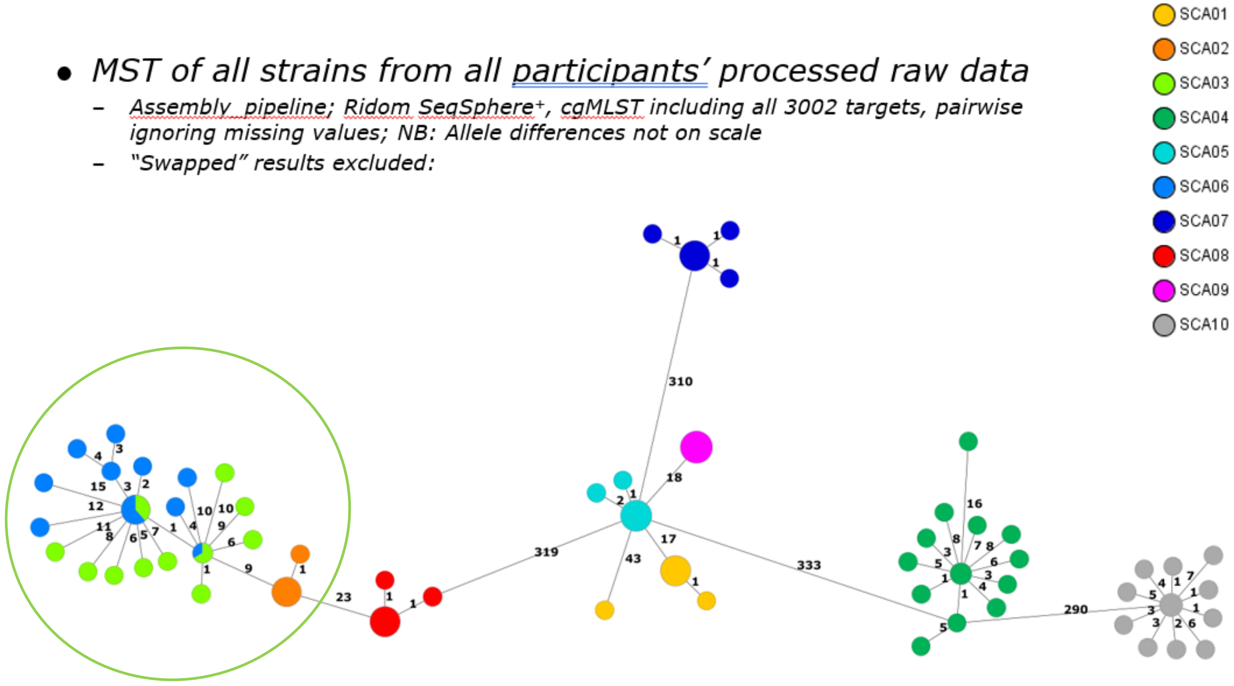
*excluding multiple data submissions per participant

Wet lab/dry lab: strain selection

- 2019: 3x STMmono plus 7x STM (all wet)
  - No specific cluster definition
  - 3 strains in expected cluster
  - 2 technical duplicates



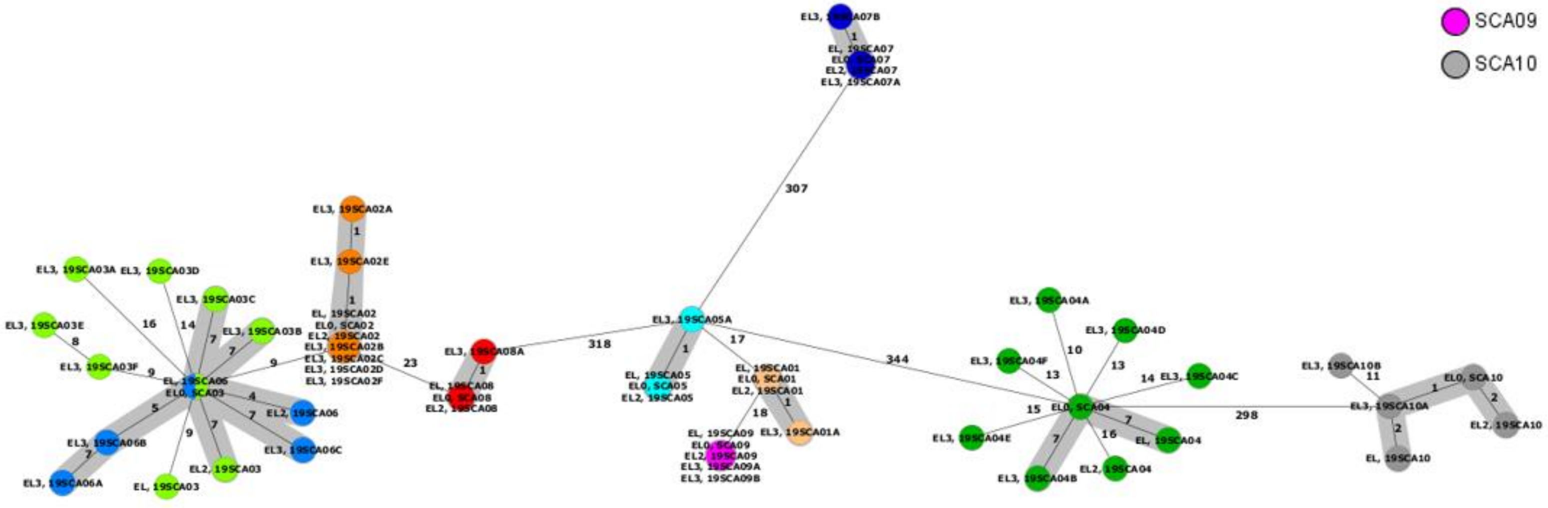
- *MST of all strains from all participants' processed raw data*
  - *Assembly pipeline; Ridom SeqSphere+, cgMLST including all 3002 targets, pairwise ignoring missing values; NB: Allele differences not on scale*
  - *"Swapped" results excluded:*



Wet lab/dry lab: strain selection

- 2019 first lesson learnt: extended pre-testing!

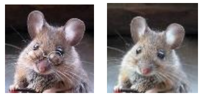
- MST of all strains, EURL-*Salmonella*
  - Assembly pipeline; Ridom SeqSphere+, cgMLST including all 3002 targets, pairwise ignoring missing values; NB: Allele differences not on scale
  - EL0: September 2019
  - EL: November 2019 (PT2019)
  - EL2: February 2020
  - EL3: November 2020, re-cultured from -70°C (November 2019); multiple colonies per plate



● SCA01
● SCA02
● SCA03
● SCA04
● SCA05
● SCA06
● SCA07
● SCA08
● SCA09
● SCA10

Wet lab/dry lab: strain selection

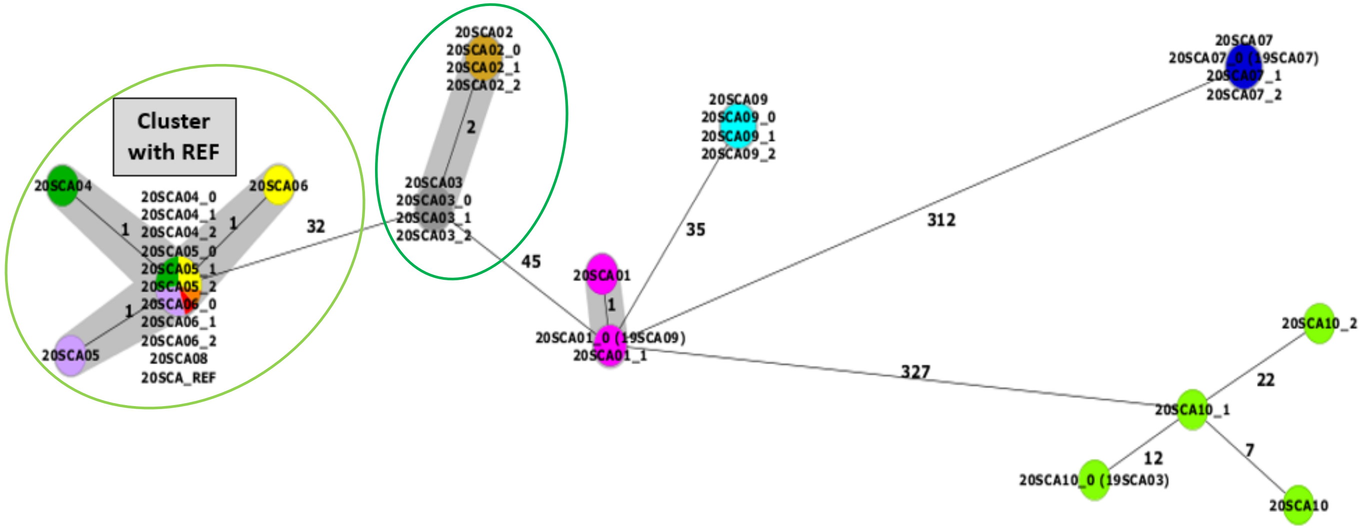
- 2020: 8x STMmono plus 2x STM (all wet)
  - Extended pre-testing
  - Specific (cgMLST-based) cluster definition for this PT (6 AD)
  - Reference strain dry data provided
  - 4 strains clustering with Reference strain, including 2 technical duplicates
  - Second cluster of 2 strains (not evaluated)



● 20SCA01
● 20SCA02
● 20SCA03
● 20SCA04
● 20SCA05
● 20SCA06
● 20SCA07
● 20SCA08
● 20SCA09
● 20SCA10
● 20REF

Strains *Salmonella* Cluster Analysis PT 2020

- 20SCA_0: Original data 2020-isolated strains and some PT 2019 strains
- 20SCA_1: WGS data for initial testing (September 2020)
- 20SCA_2: WGS data after 10 times sub-culturing (October 2020)
- 20SCA: The PT 2020 data (November 2020)

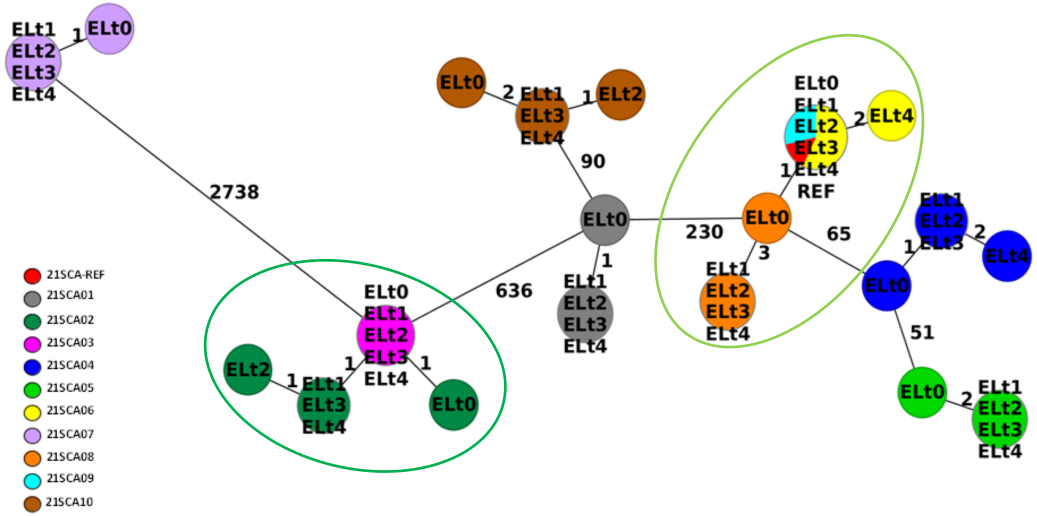


Wet lab/dry lab: strain selection

- 2021: 10x SE (all wet)

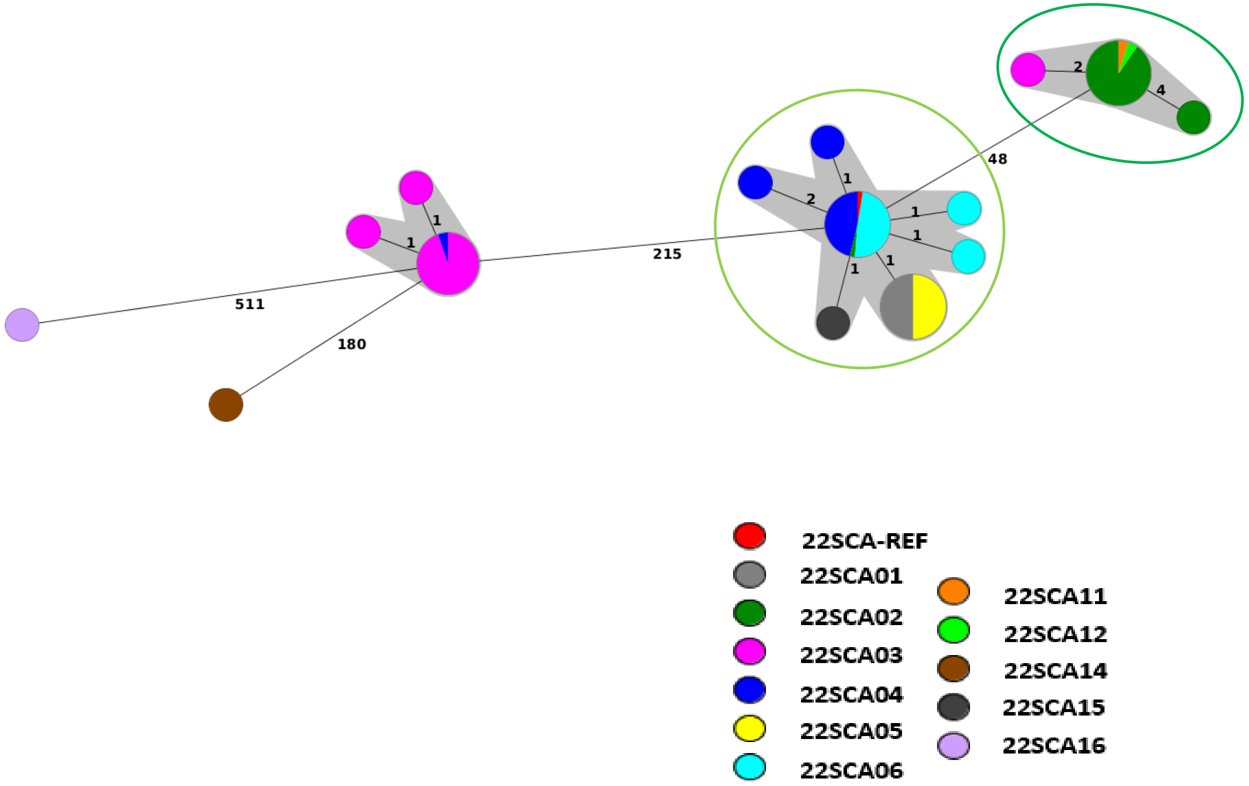
  - Extended pre-testing
  - Specific (cgMLST-based) cluster definition for this PT (7 AD)
  - Reference strain dry data provided
  - 3 strains clustering with Reference strain, including 2 technical duplicates
  - Second cluster of 2 strains (not evaluated)

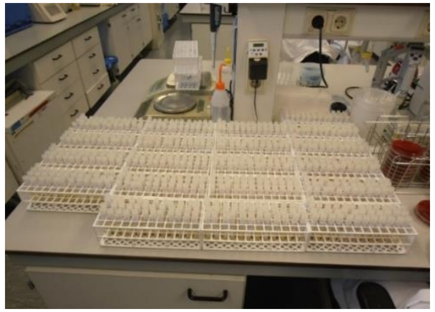
- ELt0: original WGS data from the human surveillance strains in 2019
- ELt1: WGS data initial pre-testing
- ELt2: WGS data after 10 times sub-culturing
- ELt3: WGS data at the start of the PT (November 2021)
- ELt4: WGS data at the end of the data submission period (February 2022)



Wet lab/dry lab: strain selection

- 2022: 6x SE (wet) plus 6x SE (dry)
  - Extended pre-testing
  - Specific (cgMLST-based) cluster definition for this PT (6 AD)
  - Reference strain dry data provided
  - 5 strains (1 dry) clustering with Reference strain, including 2x 2 technical duplicates (wet)
  - Second cluster of 1 wet and 2 dry strains (not evaluated, but standard question)
  - 1 dry strain “contaminated” and intended not to be included in the cluster analysis (distance matrix)





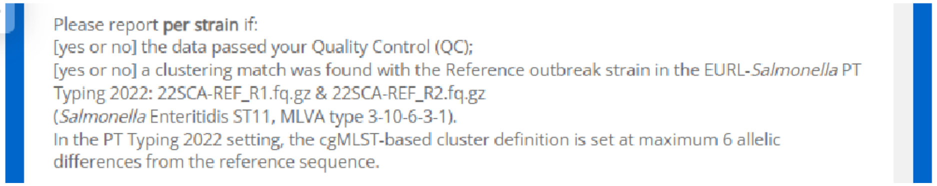
Wet lab/dry lab: strain (data) transfer

- Preparation of HI-agar tubes for shipment of strains (wet analysis)
- Shipment of strains under UN3373 (DHL)

- 2020-2022: fastq.gz data files for the Reference strain made available through a secure ftp server
- 2022: fastq.gz data files for 6 PT strains (dry analysis) and md5check data made available through a secure ftp server

Collection of results

- Online Registration for PT participation (September)
  - Timetable (generally: early November – end of January)
- PT Typing Protocol*
- Strain (data) transfer
- Online Result form*:
  - wet-lab/dry-lab protocols, including participants' quality criteria used
  - **cluster analysis result per strain** (yes/no)
- Uploading:
  - raw reads (compressed fastq.gz files)
- Emailing:
  - distance matrix (.xls or .csv format)

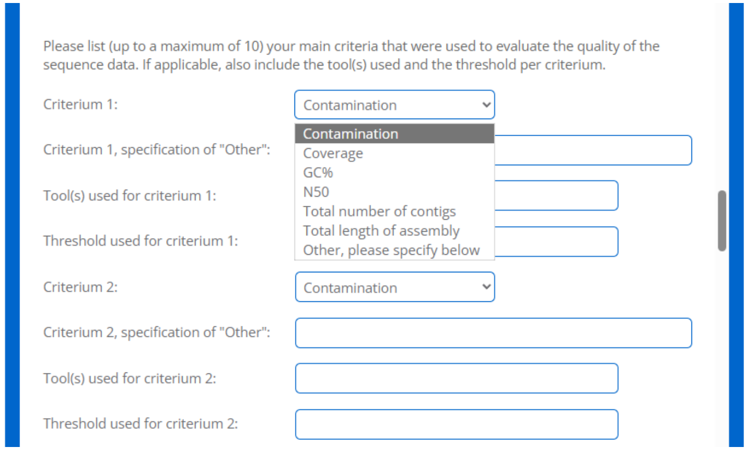


Please report **per strain** if:
[yes or no] the data passed your Quality Control (QC);
[yes or no] a clustering match was found with the Reference outbreak strain in the EURL-*Salmonella* PT Typing 2022: 22SCA-REF_R1.fq.gz & 22SCA-REF_R2.fq.gz
(*Salmonella* Enteritidis ST11, MLVA type 3-10-6-3-1).
In the PT Typing 2022 setting, the cgMLST-based cluster definition is set at maximum 6 allelic differences from the reference sequence.

*available on www.eurlsalmonella.eu

Collection of results

- Data quality criteria used by the participants
  - 2021 comparable to the PT Typing 2019 and 2020
  - (Large) variety in naming, as well as in thresholds
  - 2022: drop-down menu options in the result form



Please list (up to a maximum of 10) your main criteria that were used to evaluate the quality of the sequence data. If applicable, also include the tool(s) used and the threshold per criterium.

Criterium 1: Contamination

Criterium 1, specification of "Other":

Tool(s) used for criterium 1:

Threshold used for criterium 1:

Criterium 2: Contamination

Criterium 2, specification of "Other":

Tool(s) used for criterium 2:

Threshold used for criterium 2:

- Md5 checksum
  - to ensure the data integrity of fastq files during and after transfer

Md5 checksum
PT 2020: 11/21
PT 2021: 14/19
PT 2022: 16/20

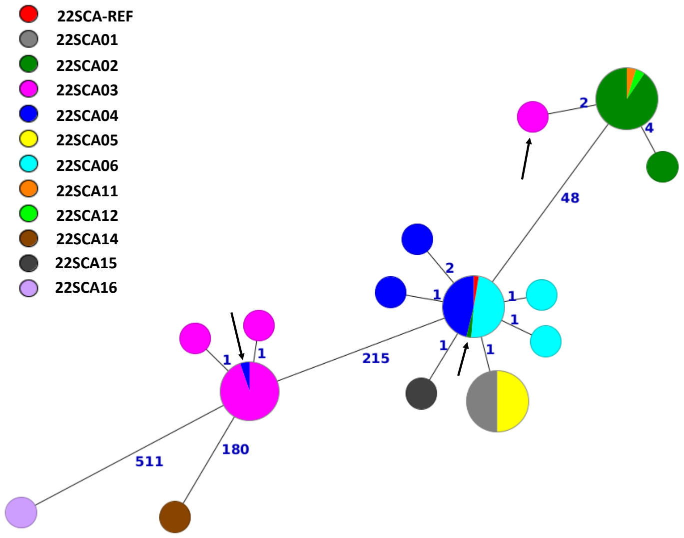
Results analysis and report

- Analysis by the EURL-*Salmonella*

- ✓ DNA extraction, library preparation, sequencing performed in-house; WGS platform: Illumina NextSeq.
- ✓ **Raw data processing**: in-house developed Juno-assembly pipeline which includes SPAdes 3.15.3.
- ✓ Cluster analysis: Ridom SeqSphere+, using the cgMLST Enterobase v2.0 scheme.

- Individual report per participant

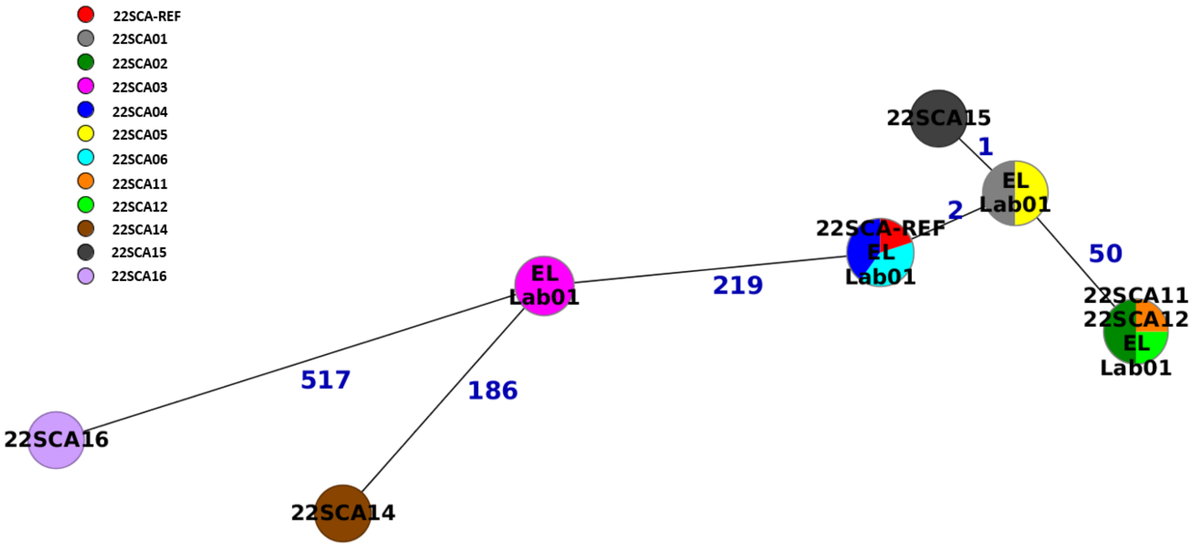
- Interim summary report
  - Published on website www.eurlsalmonella.eu
- Final report
  - Published on website www.eurlsalmonella.eu

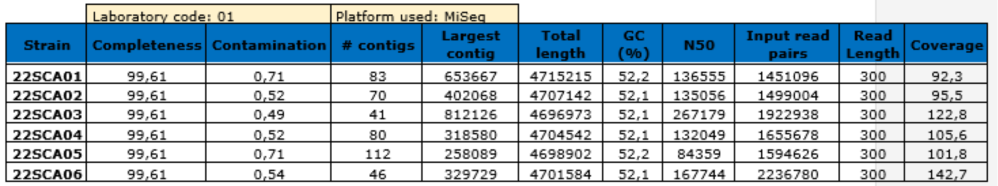


- Oral presentation/discussion during annual Workshop
  - Published on website www.eurlsalmonella.eu

Results analysis

- ✓ Evaluation participants’ raw reads by EURL-*Salmonella*
  - ✓ MST of EURL-Salmonella and participant
  - ✓ Main QC parameters per participant
- ✓ Listing wet lab/dry lab info
- ✓ Evaluation distance matrices
- ✓ Evaluation cluster analysis results
  - ✓ Per strain
  - ✓ Pre-defined cluster definition per PT

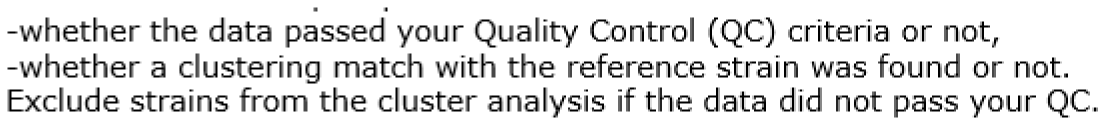




	Laboratory code: 01		Platform used: MiSeq							
Strain	**Completeness**	**Contamination**	**# contigs**	**Largest contig**	**Total length**	**GC (%)**	**N50**	**Input read pairs**	**Read Length**	**Coverage**
22SCA01	99,61	0,71	83	653667	4715215	52,2	136555	1451096	300	92,3
22SCA02	99,61	0,52	70	402068	4707142	52,1	135056	1499004	300	95,5
22SCA03	99,61	0,49	41	812126	4696973	52,1	267179	1922938	300	122,8
22SCA04	99,61	0,52	80	318580	4704542	52,1	132049	1655678	300	105,6
22SCA05	99,61	0,71	112	258089	4698902	52,2	84359	1594626	300	101,8
22SCA06	99,61	0,54	46	329729	4701584	52,1	167744	2236780	300	142,7

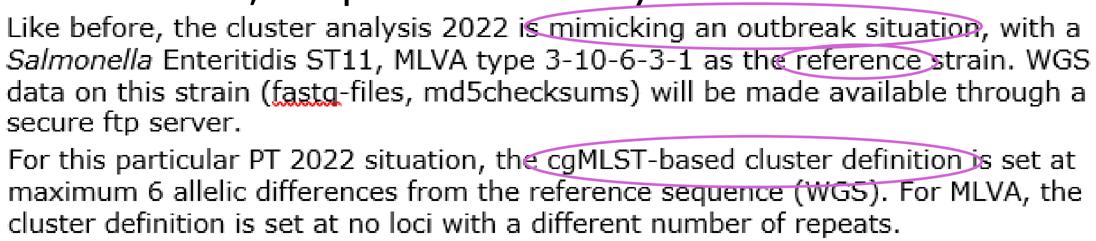
Performance assessment

- Participants were asked to report per strain:



-whether the data passed your Quality Control (QC) criteria or not,
-whether a clustering match with the reference strain was found or not.
Exclude strains from the cluster analysis if the data did not pass your QC.

- Comparing the participants' results to the expected results, as pre-defined by the EURL-*Salmonella* (Protocol PT Typing)



Like before, the cluster analysis 2022 is mimicking an outbreak situation, with a *Salmonella* Enteritidis ST11, MLVA type 3-10-6-3-1 as the reference strain. WGS data on this strain (fastq-files, md5checksums) will be made available through a secure ftp server.

For this particular PT 2022 situation, the cgMLST-based cluster definition is set at maximum 6 allelic differences from the reference sequence (WGS). For MLVA, the cluster definition is set at no loci with a different number of repeats.

- No specific performance *criteria* were set for the NGS PTs on cluster analysis
- As a minimum, it was expected to have any technical duplicate strains be reported as (part of) one cluster
- Deviations from the expected results are indicated in blue
- **Performance not expressed in terms of 'good'/'moderate'/'poor'**

Performance assessment

- Example PT 2022
  - 19/26 submissions reported completely as expected
  - Technical duplicates within one cluster:
    - 22SCA01/22SCA05: all 26
    - 22SCA04/22SCA06: 25/26

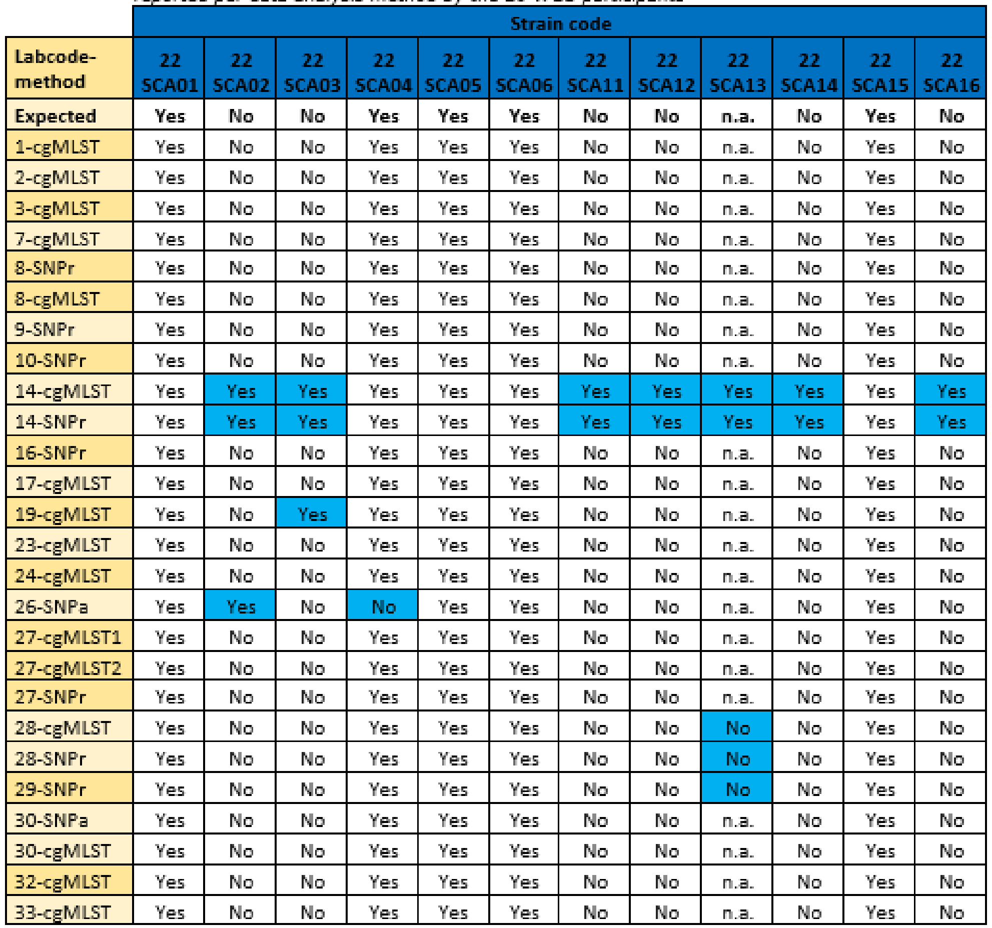


Table 4.8 Expected cluster analysis results and the cluster analysis results reported per data analysis method by the 20 WGS participants

	Strain code											
Labcode-method	**22 SCA01**	**22 SCA02**	**22 SCA03**	**22 SCA04**	**22 SCA05**	**22 SCA06**	**22 SCA11**	**22 SCA12**	**22 SCA13**	**22 SCA14**	**22 SCA15**	**22 SCA16**
Expected	**Yes**	**No**	**No**	**Yes**	**Yes**	**Yes**	**No**	**No**	**n.a.**	**No**	**Yes**	**No**
1-cgMLST	Yes	No	No	Yes	Yes	Yes	No	No	n.a.	No	Yes	No
2-cgMLST	Yes	No	No	Yes	Yes	Yes	No	No	n.a.	No	Yes	No
3-cgMLST	Yes	No	No	Yes	Yes	Yes	No	No	n.a.	No	Yes	No
7-cgMLST	Yes	No	No	Yes	Yes	Yes	No	No	n.a.	No	Yes	No
8-SNPr	Yes	No	No	Yes	Yes	Yes	No	No	n.a.	No	Yes	No
8-cgMLST	Yes	No	No	Yes	Yes	Yes	No	No	n.a.	No	Yes	No
9-SNPr	Yes	No	No	Yes	Yes	Yes	No	No	n.a.	No	Yes	No
10-SNPr	Yes	No	No	Yes	Yes	Yes	No	No	n.a.	No	Yes	No
14-cgMLST	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
14-SNPr	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
16-SNPr	Yes	No	No	Yes	Yes	Yes	No	No	n.a.	No	Yes	No
17-cgMLST	Yes	No	No	Yes	Yes	Yes	No	No	n.a.	No	Yes	No
19-cgMLST	Yes	No	Yes	Yes	Yes	Yes	No	No	n.a.	No	Yes	No
23-cgMLST	Yes	No	No	Yes	Yes	Yes	No	No	n.a.	No	Yes	No
24-cgMLST	Yes	No	No	Yes	Yes	Yes	No	No	n.a.	No	Yes	No
26-SNPa	Yes	Yes	No	No	Yes	Yes	No	No	n.a.	No	Yes	No
27-cgMLST1	Yes	No	No	Yes	Yes	Yes	No	No	n.a.	No	Yes	No
27-cgMLST2	Yes	No	No	Yes	Yes	Yes	No	No	n.a.	No	Yes	No
27-SNPr	Yes	No	No	Yes	Yes	Yes	No	No	n.a.	No	Yes	No
28-cgMLST	Yes	No	No	Yes	Yes	Yes	No	No	No	No	Yes	No
28-SNPr	Yes	No	No	Yes	Yes	Yes	No	No	No	No	Yes	No
29-SNPr	Yes	No	No	Yes	Yes	Yes	No	No	No	No	Yes	No
30-SNPa	Yes	No	No	Yes	Yes	Yes	No	No	n.a.	No	Yes	No
30-cgMLST	Yes	No	No	Yes	Yes	Yes	No	No	n.a.	No	Yes	No
32-cgMLST	Yes	No	No	Yes	Yes	Yes	No	No	n.a.	No	Yes	No
33-cgMLST	Yes	No	No	Yes	Yes	Yes	No	No	n.a.	No	Yes	No

n.a.: Not applicable (QC not passed).
In blue: Deviation from the expected result.

Lessons learnt/topics for consideration

- Pre-testing/strain selection
- Pre-defined “Cluster definitions”, (Reference strain, expected results)
  - cgMLST/SNP
- Md5 checks
- Quality control
  - Check for contamination, etc.; ISO/OneHealth/PT Criteria?
- Exchange of NGS data
  - In practice: e.g. secure ftp server, wetransfer, etc.
  - Consequences: Further use of strains/data, MTA, Public uploads, etc.
- Performance criteria?

Thank you for your attention !

Contact: wilma.jacobs@rivm.nl

Website: www.eurlsalmonella.eu

Co-funded by the
European Union

This publication was co-funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the granting authority European Health and Digital Executive Agency (HaDEA). Neither the European Union nor the granting authority can be held responsible for them.