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

Results EURL-*Salmonella* PT Typing 2019

- Wilma Jacobs-Reitsma
- EURL-*Salmonella*

EURL Salmonella

- With back-up from our
- RIVM typing experts

RIVM typing experts

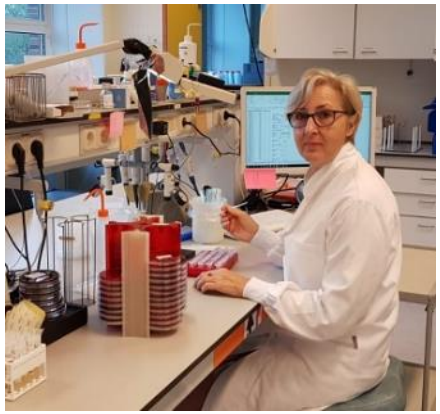
Sad message

Henny Maas passed away last July, in the age of 65 years

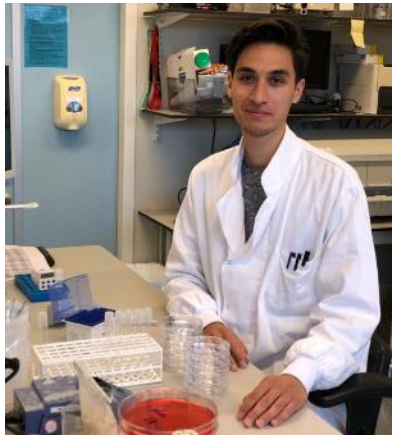
She was our outstanding serotyping expert from the early days on, till her retirement in 2017

The Salmonella serovar Bilthoven in our Quiz is dedicated to Henny

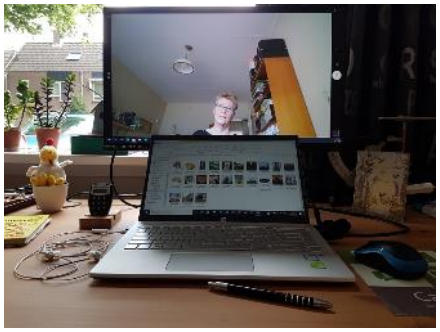
RIVM typing experts



- Serotyping:
  - Anjo Verbruggen



- Cluster Analysis:
  - Angela van Hoek
  - Robin Diddens
  - Wilma Jacobs

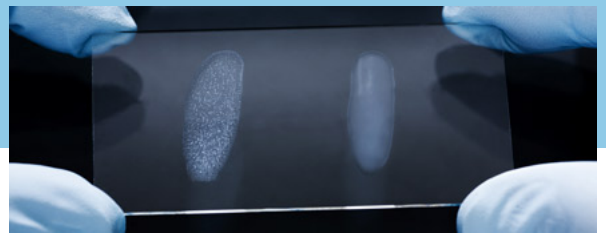


- RIVM contact typing studies:
  - Wilma Jacobs

Participants 2019

- National Reference Laboratories for *Salmonella* from all 28 EU member states
- Albania, Republic of North Macedonia and Serbia as EU candidate countries
- Iceland, Norway, Switzerland (EFTA)

- Serotyping
  - 35 participants
- Pilot on Cluster Analysis
  - 18 participants



Part on Serotyping (24th PT)

- 20 different serovars of *Salmonella enterica* subsp. *enterica*
- 1 additional strain of an uncommon type (optional)
- Serotyping and reporting in accordance with the White-Kauffmann-le Minor scheme (2007)
- Laboratories have to report only those results, on which the identification of serovar names is based.
  - Examples of preferred reporting:

O-antigens	H-antigens (phase 1)	H-antigens (phase 2)	Serovar name
9,12	g,m	-	Enteritidis
4,12	i	2	Typhimurium
4,5,12	i	-	4,5,12:i:-
6,7	-	1,5	6,7:-:1,5
42	g,t	-	42:g,t:-

Salmonella strains (1)

Strain code	O-antigens	H-antigens (phase 1)	H-antigens (phase 2)	Serovar
S7	1,4,[5],12,27	d	1,2	Stanley
S8	1,4,[5],12	f,g,s	[1,2]	Agona
S20	1,4,[5],12,27	g,s,t	[1,2]	Kingston
S9[a)]	**1,4,[5],12**	**i**	**1,2**	**Typhimurium**
S14	**1,4,[5],12**	**i**	**-**	**1,4,[5],12:i:-**
S2	1,4,[5],12	y	1,2	Coeln
S5	**1,9,12**	**g,m**	**-**	**Enteritidis**
S4	9,46	z_{38}	-	Fresno
S11	3,{10},{15}	l,z_{13}	1,5	Uganda
S6	3,{10}{15}{15,34}	e,h	l,w	Meleagridis

a) Potentially contaminated with an *E. coli* strain.
Results strain S9 were excluded from the evaluation.

Salmonella strains (2)

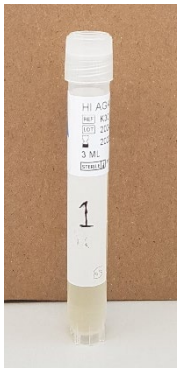
Strain code	O-antigens	H-antigens (phase 1)	H-antigens (phase 2)	Serovar
S10*	6,7	r	1,6	Nigeria
S16	**6,7,14**	**r**	**1,2**	**Virchow**
S17	**6,7,14**	**r**	**1,5**	**Infantis**
S15	6,8	d	1,2	Muenchen
S1	**6,8**	**z_{10}**	**e,n,x**	**Hadar**
S12	11	r	e,n,x	Rubislaw
S3	1,13,22	z	1,6	Poona
S19	1,13,23	i	l,w	Kedougou
S18*	16	y	1,5	Saphra
S13	{6,7,14}{54}	g,m,[p],s	[1,2,7]	Montevideo
S21	48	g,z_{51}	-	IV 48:g,z_{51}:-

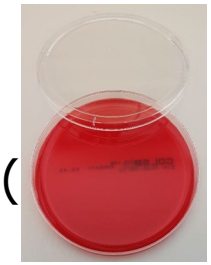
* Represented in an EURL-*Salmonella* PT Serotyping for the first time.

Salmonella strain S9[a)]

Production of the agar transport tubes in general:

or



()

a) Potentially contaminated with an *E. coli* strain.
Results strain S9 were excluded from the evaluation.

Salmonella strain S9[a)]

Strain code	O-antigens	H-antigens (phase 1)	H-antigens (phase 2)	Serovar	Lab code
S-9	1,4,[5],12	i	1,2	Typhimurium	REF
S-9	6,8	k	5	Blockley	27
S-9	-	-	-	-:-:-	9
S-9	-	-	-	Escherichia coli	26
S-9	4,12	i	2	Typhimurium var Copenhagen	32

in variants:

S-9	1,4,5,12	i	1,2	Typhimurium	31 x

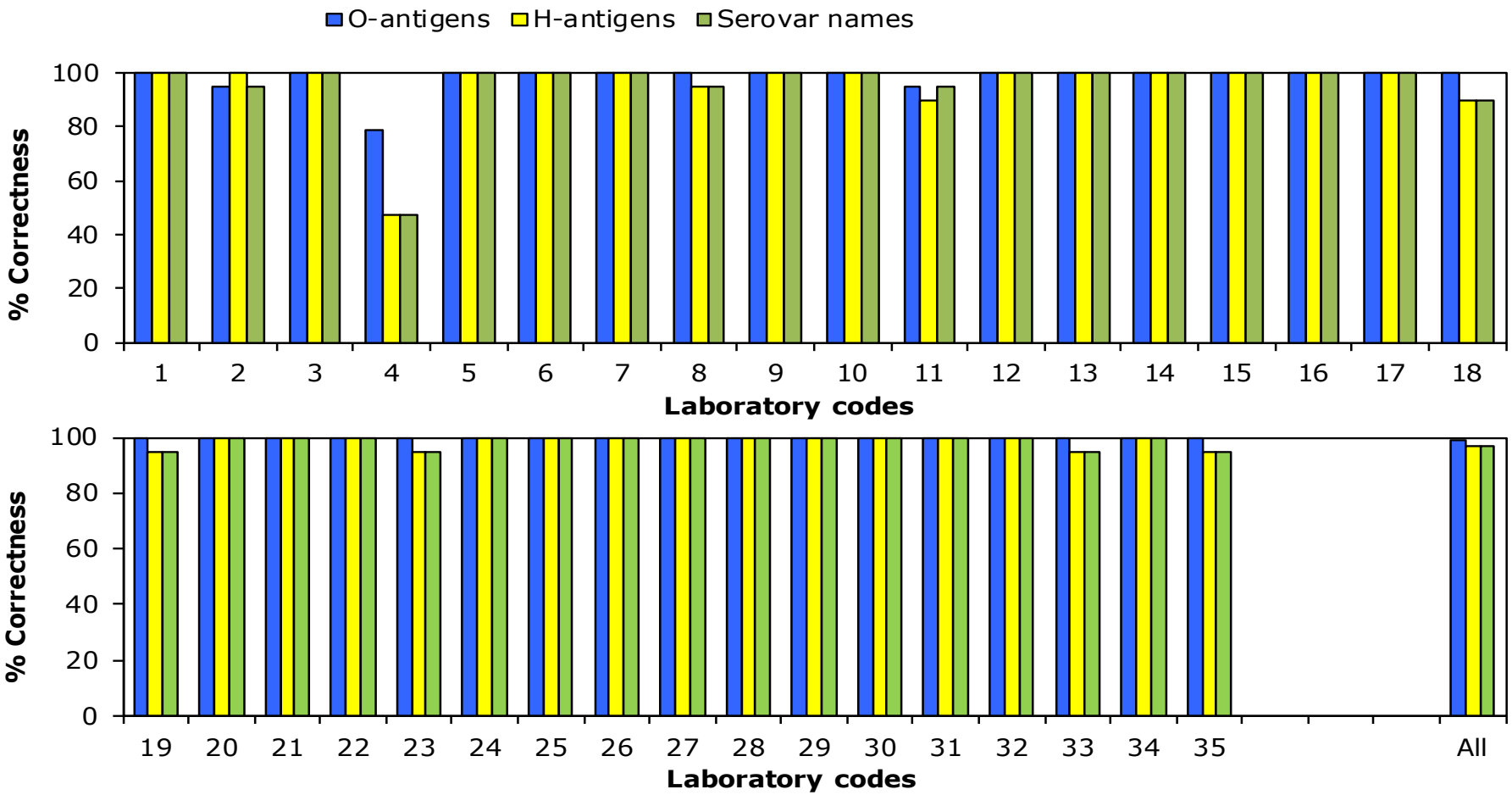
Remarks by participants:
5 additional stock colonies tested: 4x *Salmonella* , 1x *E. coli*
-
S-9 E.coli, as determined by MALDI-TOF
S9 turned out to be E. coli using MALDI-TOF
All samples, were deemed pure when plated onto media

in variants:

8 x mixed bacterial culture (contamination with *E. coli*)
23 x no comment on this topic

a) Potentially contaminated with an *E. coli* strain.
Results strain S9 were excluded from the evaluation.

Percentages of correct serotyping results, per NRL



Correct typing of strains

- 100% accurate serotype naming by 26/35 (74%) laboratories

- 7 serovars completely correct identified by all participants: Poona (S3), Enteritidis (S5), Montevideo (S13), Virchow (S16), Infantis (S17), Saphra (S18), and Kingston (S20).
- 8 serovars completely correct identified by all participants, except for *new* participant L4: Coeln (S2), Meleagridis (S6), Stanley (S7), Agona (S8), Nigeria (S10), Rubislaw (S12), 1,4,[5],12:i:- (S14), Kedougou (S19)

Details on the problems in serotyping

Strain code	O-antigens	H-antigens (phase 1)	H-antigens (phase 2)	Serovar	Lab code
S-1	6,8	z10	e,n,x	Hadar	REF
S-1	6,8	i	e,n,x	Bonariensis	35
S-3	1,13,22	z	1,6	Poona	REF
S-3	13,22	z	1,2	Poona	11
S-4	9,46	z38	-	Fresno	REF
S-4	9	z38	-	Elomrane	2
S-4	9,46	?	-	9,46 : ? :-	8
S-4	9	-	-	9:-:-	11
S-4	9,46	?	?	9,46:?	19
S-4	9,46	-	-	9,46:-:-	23
S-11	3,{10}{15}	l,z13	1,5	Uganda	REF
S-11	3,10	l,z13	1,5	Ouganda	5
S-11	3,10	l,v	1,6	London	18
S-11	3,10	l,v	1,5	Sinstorf	33
S-15	6,8	d	1,2	Muenchen	REF
S-15	6,8	d	1,5	Manhattan	18

Reference strain
remark (e.g. spelling error)
not typable (e.g. antisera not available, rough strain)
partly correct; in the naming: no penalty points
incorrect; in the naming: 1 penalty point
incorrect; in the naming: 4 penalty points

NB: Results of new participant Laboratory 4 are not included

Additional strain S21

- Human origin
- Examined by 30/35 labs
- *S. enterica* subsp. *houtenae* (IV)

Strain code	O-antigens	H-antigens	H-antigens	Serovar	Lab code
S-21	**48**	**g,z51**	**-**	**IV 48:g,z51:-**	**REF**
S-21	9,46	g,m,s	-	Macclesfield	4
S-21	48	g,z51	-	Salmonella enterica subspecies arizonae	7
S-21	48	g	z51	48 : g : z51 (IV)	21
S-21	48	g	-	Houtenae	10
S-21	48	g	-	Subspec IV, Antigenicformula=48:g:-***	20
S-21	-	g	-	-:g:-	33
S-21	-	-	-	-:-:-	23

in variations:

S-21	48	g,z51	-	S. enterica subsp. houtenae 48:g,z51:- (IV)	23x

Evaluation of Performance

- **"Good Performance" (**Workshop Bilthoven, 2007)

- **4 penalty points:** Incorrect typing of *S.* Enteritidis, *S.* Typhimurium (*including the monophasic variant*), *S.* Hadar, *S.* Infantis or *S.* Virchow **or** assigning the name of one of these 5 serotypes to another strain.
- **1 penalty point:** Incorrect typing of all other *Salmonella* serotypes.

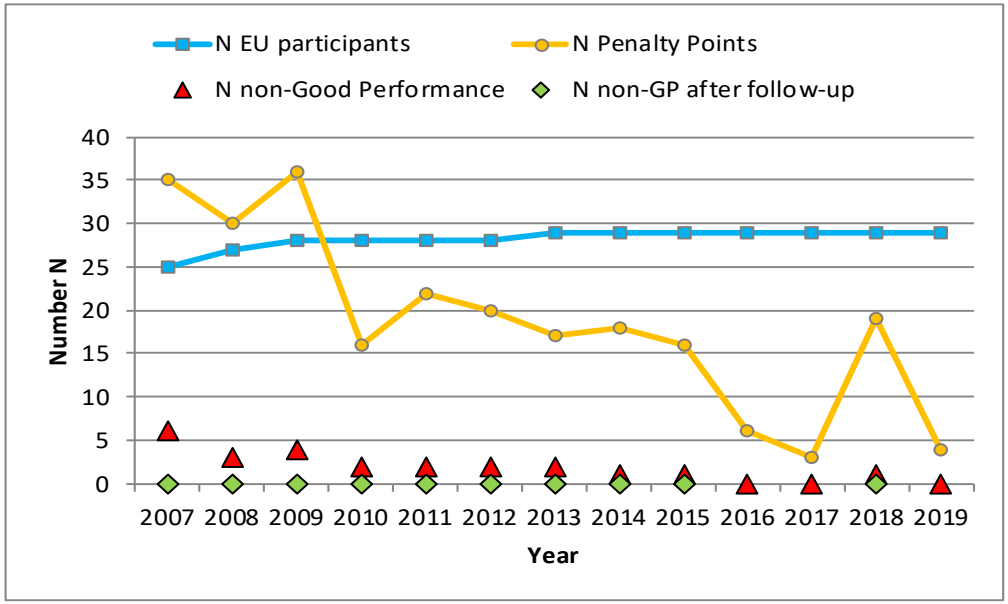
- For each NRL-*Salmonella* the total amount of penalty points is determined.
- "Good Performance" is: less than 4 penalty points.
- A follow-up is obligatory for NRLs with 4 penalty points or more.

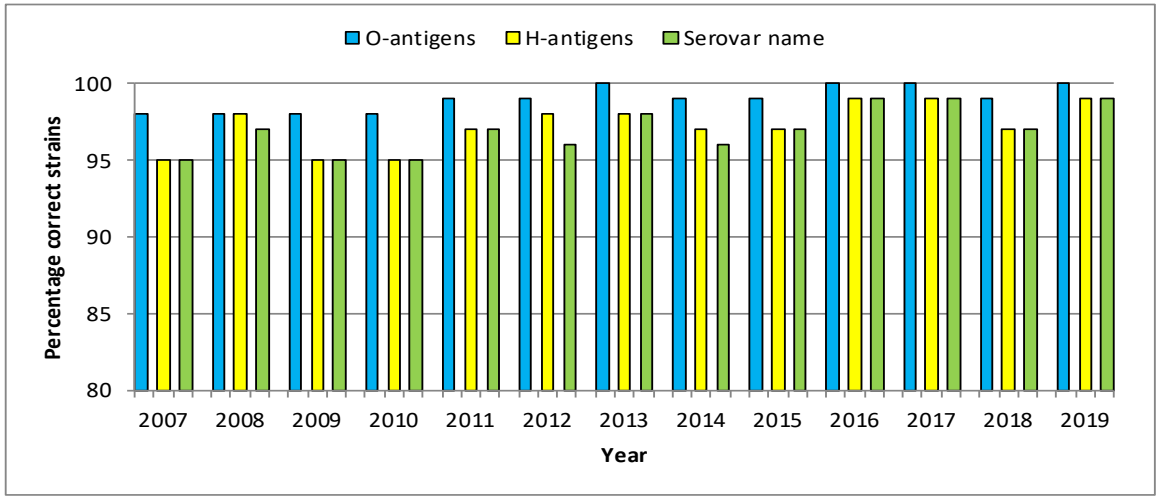
Evaluation of Performance

Lab code	Penalty points	Good performance
1	0	yes
2	1	yes
3	0	yes
4	18	NO
5	0	yes
6	0	yes
7	0	yes
8	0	yes
9	0	yes
10	0	yes
11	0	yes
12	0	yes
13	0	yes
14	0	yes
15	0	yes
16	0	yes
17	0	yes
18	2	yes

Lab code	Penalty points	Good performance
19	0	yes
20	0	yes
21	0	yes
22	0	yes
23	0	yes
24	0	yes
25	0	yes
26	0	yes
27	0	yes
28	0	yes
29	0	yes
30	0	yes
31	0	yes
32	0	yes
33	1	yes
34	0	yes
35	4	NO

Results in time (EU NRLs only)





Conclusions 24th serotyping study (all participants)

- O-antigens: 99% of strains typed correctly
- H-antigens: 97% of strains typed correctly
- Serovar names: 97% of strains typed correctly

- 100% accurate serotyping by 26/35 (74%) laboratories
  - 4 participants lacking the appropriate but less common antisera (HME, H:z38, H:z41) for strain S4 (Fresno)
  - 5 participants with only 1 or 2 mistakes
  - 1 *new* participant may benefit from on-site training

- 33 laboratories (all EU-MS) "Good Performance"
- 2 non-EU-MS laboratories *no* "Good Performance"
  - No Follow-up needed/training pending

Part on Cluster Analysis, first pilot

- Cluster analysis using PFGE and/or MLVA and/or WGS
- Participants' own routine method(s) of choice

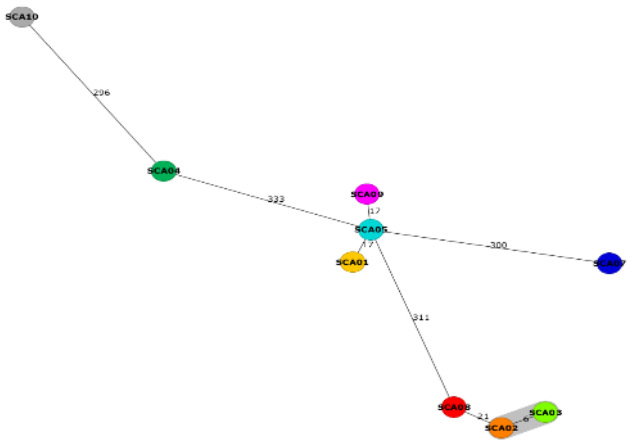
- PFGE
  - Result form: protocol used, position of lanes, total number of bands per profile, **cluster identification**
  - Emailing: PFGE gel image, zip file of the analysis in BioNumerics
- MLVA
  - Result form: scheme used, allelic profile, **cluster identification**
- WGS
  - Result form: wet-lab/dry-lab protocols, **cluster identification**
  - Uploading: raw reads (fastq files)
  - Emailing: distance matrix

Strains *Salmonella* Cluster Analysis

- 15 potentially suitable strains (<2015, PFGE/MLVA) re-cultured from storage (HI tubes, room temperature)
- WGS analysis
- 9 strains selected for inclusion in the cluster analysis
- Technical duplicates: SCA03 and SCA06
  - All those shipment tubes prepared from identical blood agar plate culture

Strain code	Serovar	ST	MLVA-profile
SCA01 [a)]	4,5,12:i:-	34	3-13-9-NA-211
SCA02	Typhimurium	19	3-16-17-18-311
SCA03 [b)]	Typhimurium	19	3-16-7-17-311
SCA04	Typhimurium	19	2-20-8-11-212
SCA05 [a)]	4,5,12:i:-	34	3-11-9-NA-211
SCA06 [b)]	Typhimurium	19	3-16-7-17-311
SCA07	Typhimurium	19	5-9-14-9-211
SCA08	Typhimurium	19	3-14-17-25-311
SCA09 [a)]	4,5,12:i:-	34	3-13-9-NA-211
SCA10	Typhimurium	19	2-12-7-9-212

[a)] Typhimurium, monophasic variant as determined by PCR.
[b)] Technical duplicates (in bold).



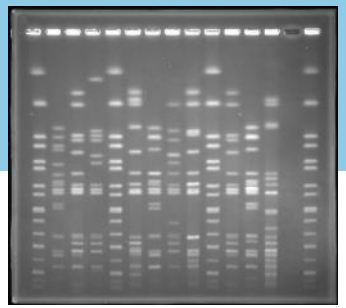
Evaluation (Comparison) Cluster Analysis

- Evaluation per methodology (PFGE/MLVA/WGS):
  - The ability to correctly identify cluster(s) of genetically closely related isolates, as pre-defined by the EURL-*Salmonella* ("expected results")
- *However, cluster definitions may vary depending on the situation or the specific research question, e.g. in outbreak investigations or surveillance.*
- The participants of this pilot PT were free to use their own interpretation of "cluster(s) of closely related isolates"
- **No performance criteria were set** for this pilot on cluster analysis
  - As a minimum, it was expected that the participants would report the **technical duplicate strains** SCA03 and SCA06 to be (part of) one cluster.

Participation to the pilot on cluster analysis

- 18 participants
  - 6 on PFGE analysis
  - 8 on MLVA analysis
  - 14 on WGS analysis

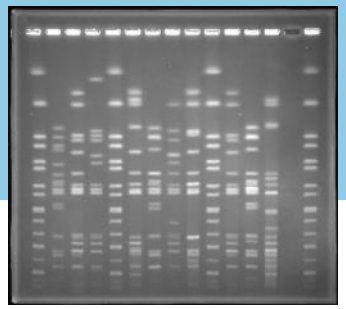
Participating in:			Number of participants	Laboratory codes
PFGE			2	8, 25
	MLVA		1	13
		WGS	8	9, 15, 16, 17, 27, 28, 29, 32
PFGE	MLVA		1	21
	MLVA	WGS	3	20, 22, 26
PFGE	MLVA	WGS	3	11, 14, 34
Total PFGE:	Total MLVA:	Total WGS:	Total overall:	
6	8	14	18	



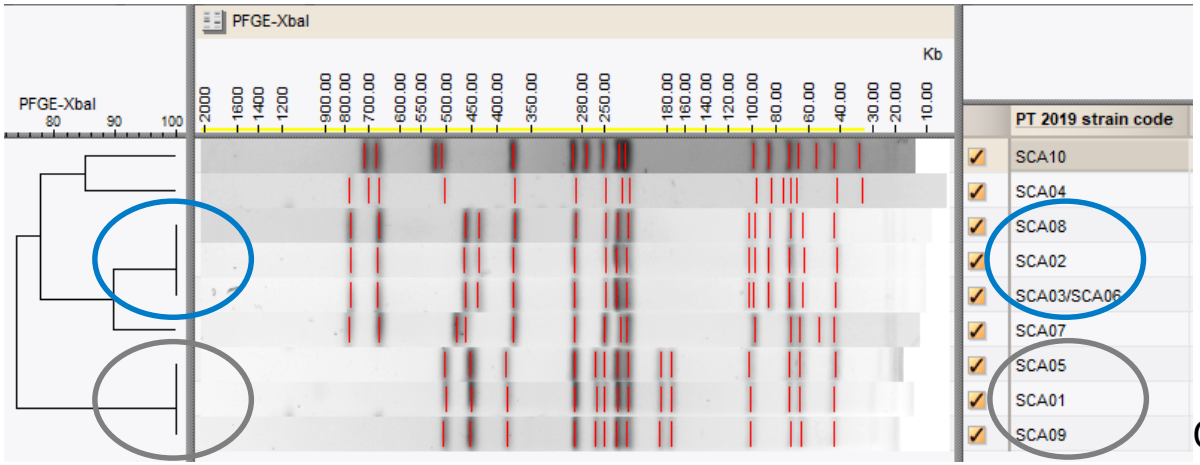
Evaluation PFGE-based cluster analysis

- 6 participants
- Original data (< 2015) analysis in BioNumerics
- Cluster definition from "SOP for the analysis of data in the joint EFSA-ECDC molecular typing database for the purpose of outbreak detection and assessment", October 2015
  - zero bands difference in PFGE with the *Xba*I enzyme using the Dice similarity coefficient with tolerance and optimisation of 1.5%

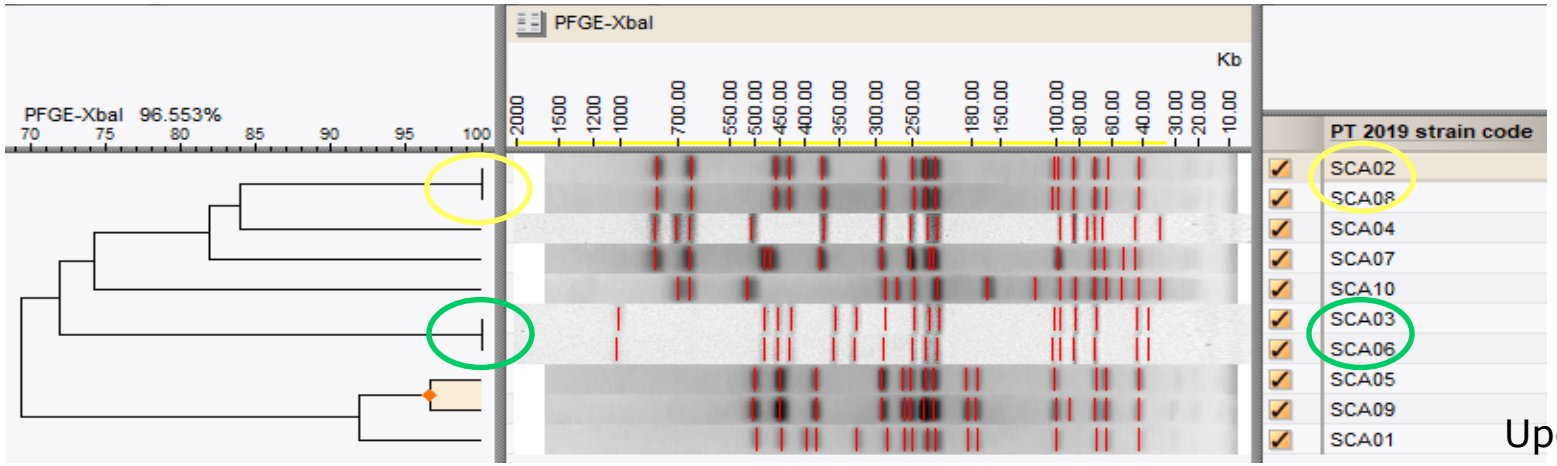
- *"Expected (REF) results" :*
  - ***2*** *clusters: 4 strains SCA02, SCA03/SCA06, SCA08 and 3 strains SCA01, SCA05, SCA09*
- *Updated REF results, based on the combined participants' results*
  - ***2*** *clusters: 2 strains SCA02, SCA08 and the 2 technical duplicates SCA03/SCA06*



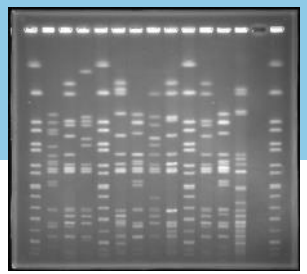
Evaluation PFGE-based cluster analysis



Original REF results



Updated REF results



Evaluation PFGE-based cluster analysis

- Results *sec* compared to the "expected results"
  - REF, as pre-defined by the EURL-*Salmonella* for this PT
  - Technical duplicates SCA03 and SCA06 within one cluster

Labcode	Number of clusters reported	Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5	Cluster identification as expected	SCA03/SCA06 within 1 cluster
REF-Updated	**2**	**SCA03, SCA06**	**SCA02, SCA08**				**REF-Updated**	**REF-Updated**
11	1		SCA02; SCA08				No	No
8	2	SCA3-SCA6	SCA2-SCA8				Yes	Yes
25	2	SCA02, SCA08, SCA07, SCA03, SCA06, SCA04, SCA10		SCA01, SCA09, SCA05			No	Yes
21	3	SCA03,SCA06	SCA02,SCA08	SCA01,SCA05,SCA09			No	Yes
34	3	SCA03, SCA06	SCA02, SCA08	SCA01, SCA05, SCA09			No	Yes
14	5	SCA06, SCA03	SCA02, SCA08, SC07	SCA01, SCA09, SCA05	SCA04	SCA10	No	Yes

SCA03, SCA06	Cluster 1, PFGE-based analysis
SCA02, SCA08	Cluster 2, PFGE-based analysis
	Nota Bene

- 1 x reported completely as expected
- 4 x also the 3 (STMmono) strains SCA01, SCA05, SCA09 as a cluster
- 2 x all strains assigned into clusters

"Cluster" definition ?!

Evaluation MLVA-based cluster analysis

- 8 participants
- Original data (< 2015) analysis in BioNumerics
- Cluster definition from "SOP for the analysis of data in the joint EFSA-ECDC molecular typing database for the purpose of outbreak detection and assessment", October 2015
  - no loci with a different number of repeats by MLVA;

- *"Expected (REF) results" :*
  - ***2** clusters:*

2-12-7-9-212	SCA10	
2-20-8-11-212	SCA04	
3-11-9-NA-211	SCA05	
3-13-9-NA-211	SCA01	Cluster 2, MLVA-based analysis
3-13-9-NA-211	SCA09	Cluster 2, MLVA-based analysis
3-14-17-25-311	SCA08	
3-16-7-17-311	SCA03	Cluster 1, MLVA-based analysis
3-16-7-17-311	SCA06	Cluster 1, MLVA-based analysis
3-16-17-18-311	SCA02	
5-9-14-9-211	SCA07	

Individual results MLVA (for information only)

- loci:STTR9, STTR5, STTR6, STTR10, STTR3

Labcode\Strain	SCA01	SCA02	SCA03	SCA04	SCA05
REF	**3-13-9-NA-211**	**3-16-17-18-311**	**3-16-7-17-311**	**2-20-08-11-212**	**3-11-9-NA-211**
11	03-13-09-NA-211	03-16-17-18-311	03-16-07-17-311	02-20-08-11-212	03-11-09-NA-211
13	03-13-09-NA-211	03-16-17-18-311	03-16-07-17-311	02-20-08-11-212	03-11-09-NA-211
14	3-13-9-00-211	3-16-17-18-311	3-16-7-17-311	2-20-8-11-212	3-11-9-00-211
20	3 13 9 NA 211	3 16 17 18 311	3 16 7 17 311	2 20 8 11 212	3 11 9 NA 211
21	03-13-09-00-211	03-16-17-18-311	03-16-07-17-311	02-20-08-11-212	03-11-09-00-211
22	3-13-9-00-0211	3-16-17-18-0311	3-16-7-17-0311	2-20-8-11-0212	3-11-9-00-0211
26	03-13-09-00-211	03-16-17-18-311	03-16-07-17-311	02-20-08-11-212	03-11-09-00-211
34	03-13-09-NA-0211	03-16-17-18-0311	03-16-07-17-0311	02-20-08-11-0212	03-11-09-NA-0211

Labcode\Strain	SCA06	SCA07	SCA08	SCA09	SCA10
REF	**3-16-7-17-311**	**5-9-14-9-211**	**3-14-17-25-311**	**3-13-9-NA-211**	**2-12-7-9-212**
11	03-16-07-17-311	05-09-14-09-211	03-14-17-25-311	03-13-10-NA-211	02-12-07-09-212
13	03-16-07-17-311	05-09-14-09-211	03-14-17-25-311	03-13-09-NA-211	02-12-07-09-212
14	3-16-7-17-311	5-9-14-9-211	3-14-17-25-311	3-13-9-00-211	2-12-7-9-212
20	3 16 7 17 311	5 9 14 9 211	3 14 17 25 311	3 13 9 NA 211	2 12 7 9 212
21	03-16-07-17-311	05-09-14-09-211	03-14-17-25-311	03-13-09-00-00	02-12-07-09-00
22	3-16-7-17-0311	5-9-14-9-0211	3-14-17-25-0311	3-13-9-00-0211	2-12-7-9-0212
26	03-16-07-17-311	05-09-14-09-211	03-14-17-25-311	03-13-09-00-211	02-12-07-09-212
34	01-16-07-17-0311	05-09-14-09-0211	03-14-17-25-0311	03-13-09-NA-0211	02-12-07-09-0212

Evaluation MLVA-based cluster analysis

- Results *sec* compared to the "expected results"
  - REF, as pre-defined by the EURL-*Salmonella* for this PT
  - Technical duplicates SCA03 and SCA06 within one cluster

Labcode	Number of clusters reported	Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5	Cluster identification as expected	SCA03/SCA06 within 1 cluster
REF	**2**	**SCA03, SCA06**	**SCA01, SCA09**				**REF**	**REF**
11	1	SCA03; SCA06					No	Yes
13	2	SCA03;SCA06	SCA01;SCA09;SCA05				No	Yes
20	2	SCA03 and SCA06	SCA01 and SCA09				Yes	Yes
22	2	SCA03, SCA06	SCA01, SCA09				Yes	Yes
26	2	SCA03, SCA06	SCA01, SCA09				Yes	Yes
34	3	SCA03, SCA06	SCA01, SCA05, SCA09	SCA02, SCA08			No	Yes
14	5	SCA08, SCA02, SCA03, SCA06	SCA05, SCA01, SCA09	SCA07	SCA10	SCA04	No	Yes
21	Not done						not applicable	not applicable

SCA03, SCA06	Cluster 1, MLVA-based analysis
SCA01, SCA09	Cluster 2, MLVA-based analysis

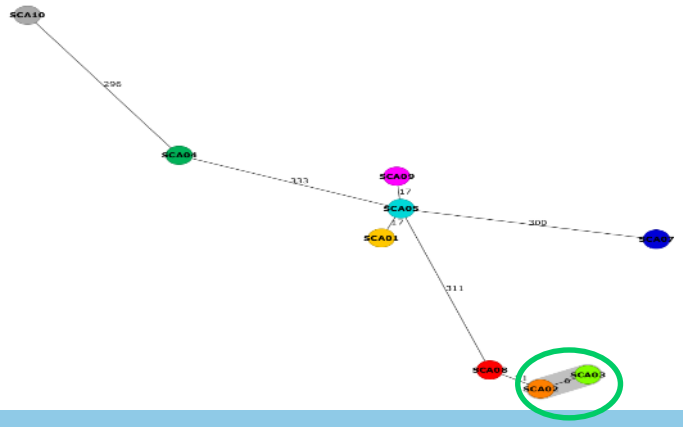
3-11-9-NA-211	SCA05	
3-13-9-NA-211	SCA01	Cluster 2, MLVA-based analysis
3-13-9-NA-211	SCA09	Cluster 2, MLVA-based analysis

- 3 x reported completely as expected
- 4 x also the 3 (STMmono) strains SCA01, SCA05, SCA09 as a cluster
- 1 x all strains assigned into clusters

"Cluster" definition ?!

Evaluation WGS-based cluster analysis

- 14 participants
- REF Selection from 15 original strains
  - DNA extraction, library preparation, sequencing performed externally; WGS platform: Illumina NovaSeq
  - Raw data processing: in-house developed pipeline *(assembly_pipeline:* https://github.com/Papos92), which includes the SPAdes assembler.
  - Cluster analysis: Ridom SeqSphere+, using the cgMLST Enterobase v2.0 scheme
- Cluster definitions for WGS ?! (no standard reference)
  - Cluster alert was set at 7 *in this particular pilot*
- *"Expected (REF) results" :*
  - ***1*** *cluster:*
  - *3 strains SCA02, SCA03/SCA06*



WGS protocols used by the 14 participants

Labcode	DNA extraction, library preparation and sequencing performed	WGS platform used	Data analysis used	Tool used for analysis	Method used for phylogenetic analysis
15	In-house	Illumina MiniSeq	cg-MLST-based	Ridom SeqSphere	Minimum Spanning Tree (MST)
29	In-house	Illumina MiSeq	cg-MLST-based	Ridom SeqSphere	Minimum Spanning Tree (MST)
22	In-house	Illumina MiSeq	cg-MLST-based	Ridom SeqSphere	MST, pairwise comparison
EL	Outsourced	Illumina NovaSeq	cg-MLST-based	Ridom SeqSphere	MST, pairwise comparison
17	In-house	Illumina MiSeq	cg-MLST-based	Ridom SeqSphere	Neighbor joining (NJ)
26	In-house	Illumina NextSeq	cg-MLST-based	Ridom SeqSphere	Neighbor joining (NJ)
28	1: in-house ; 2,3 : outsourced	Illumina NextSeq	cg-MLST-based	Ridom SeqSphere	Neighbor joining (NJ)
11	In-house	Illumina MiSeq	cg-MLST-based	in-house galaxy	Neighbor joining (NJ)
27	In-house	Illumina MiSeq	cg-MLST-based	BioNumericsCenter for Genomic Epidemiology	Single Linkage
34	In-house	Illumina MiSeq	cg-MLST-based	inhouse automated CHEWBACCA based Pipeline	Single linkage hierarchical clustering
14	In-house	Illumina MiSeq	SNP-assembly-based	KSNP3	Neighbor joining (NJ)
32	In-house	Illumina NextSeq	SNP-reference-based	Snippy + Gubbins	Maximum likelihood (ML)
20	In-house	Illumina MiSeq	SNP-reference-based	In-house pipeline	Minimum spanning tree
9	1,2: in-house; 3: outsourced;	Illumina MiSeq	SNP-reference-based	Trimmomatic, Spades, NDtree, Seqsero	Neighbor joining (NJ)
16	In-house	Illumina MiSeq	SNP-reference-based	BWA, FreeBayes, vcflib i vcf-kit, Disty McMatrixface	Neighbor joining (NJ)

Data quality criteria as used by the participants

- Variety in naming, as well as in thresholds
- Complete overview will be included in the Full Report PT Typing 2019

	QL/QT	Criterium	Threshold
6x:	Quali	Confirmation of genus (e.g. MLST/Python, K-mer finder, KRAKEN2)	
4x:	Quali	Serotyping (e.g. SISTR, SeqSero)	
7x:	Quali	Contamination check (e.g. Kmer ID, Confindr, KRAKEN)	

2x:	Quali	GC content	%/similar between strains
1x:	Quali	GC percentage	51.9-52.2%
1x:	Quant	GC%	> 49% AND < 53%

	QL/QT	Criterium	Threshold
1x:	Quali	MLST (Bionumerics)	7 Loci Achtman
1x:	Quali	Perc. Good cgMLST Targets	> 99 %
1x:	Quali	Percent matching targets in S. enterica cgMLST scheme	More than 90%
1x:	Quant	% of good cgMLST targets	>90% of 3002 targets
1x:	Quant	7-loci MLST mean coverage	30 x
1x:	Quant	allele calling	cgMLST found and called > 95 %
1x:	Quant	allele calling result	percentage of good targets ~ 98%
1x:	Quant	cgMLST genes found	>95%
1x:	Quant	Core Genome	At least 98%
1x:	Quant	Reference coverage	>90%

	QL/QT	Criterium	Threshold
1x:	Quant	Average assembly coverage	more than 10 reads
1x:	Quant	Coverage	10 x
1x:	Quant	Coverage	minimum 20-30x
1x:	Quant	Coverage	> 25x
4x:	Quant	Coverage (depth)	> 30X
1x:	Quant	Coverage	min 30X, max 100X
2x:	Quant	Coverage	50 x
1x:	Quant	Mean coverage	10
1x:	Quant	Median coverage	>20x
1x:	Quant	raw reads theoretical coverage	> 30X

	QL/QT	Criterium	Threshold
1x:	Quant	N50	> 10 000
1x:	Quant	N50	> 15 000
1x:	Quant	N50	> 30 000
2x:	Quant	N50	50 000
1x:	Quant	N50	80 000
1x:	Quant	N50	> 100 000
1x:	Quant	N50	> 400 000

	QL/QT	Criterium	Threshold
1x:	Quant	Number of contigs assembly	< 60
1x:	Quant	number of contigs	< 200
1x:	Quant	Number of contigs	200 bases (contigs shorter than 200 bases have to be ignored)
1x:	Quant	Number of contigs	250
1x:	Quant	Number of contigs	< 300
1x:	Quant	Number of contigs	400 or Less
1x:	Quant	Number of contigs	< 500

Evaluation WGS-based cluster analysis

- Results *sec* compared to the "expected results"

Labcode (method used)	Number of clusters reported	Cluster 1	Cluster 2	Cluster 3	Cluster identification as expected	SCA03/SCA06 within 1 cluster
REF (cg-MLST)	**1**	**SCA02, SCA03, SCA06**			**REF**	**REF**
29 (cg-MLST)	0				No	No
11 (cg-MLST)	1	SCA01; SCA02			No	No
15 (cg-MLST)	1	SCA03 - SCA06			No	Yes
17 (cg-MLST)	1	SCA03, SCA06			No	Yes
20 (SNP-reference)	1	SCA03 and SCA06			No	Yes
22 (cg-MLST)	1	SCA03, SCA06			No	Yes
26 (cg-MLST)	1	SCA03, SCA06			No	Yes
27 (cg-MLST)	2	SCA06, SCA03 and SCA02			Yes	Yes
9 (SNP-reference)	2	SCA02, SCA03, SCA06	SCA01, SCA05, SCA09		No	Yes
16 (SNP-reference)	2	SCA08, SCA03, SCA02, SCA06 and SCA07, SCA10, SCA04	SCA09, SCA05, SCA01		No	Yes
34 (cg-MLST)	2	SCA02, SCA03, SCA06, SCA08	SCA01, SCA05, SCA09		No	Yes
14 (SNP-assembly)	3	SCA08, SCA02, SCA06, SCA033, (SCA07)	SCA05, SCA01, SCA09	SCA10, SCA04	No	Yes
28 (cg-MLST)	3	SCA08, SCA02, SCA03, SCA06	SCA05, SCA01, SCA09	SCA07, SCA04, SCA10	No	Yes
32 (SNP-reference)	4	SC8, SC2, SC6, SC3	SC7, SC5, SC1, SC9	SC4, SC10	No	Yes

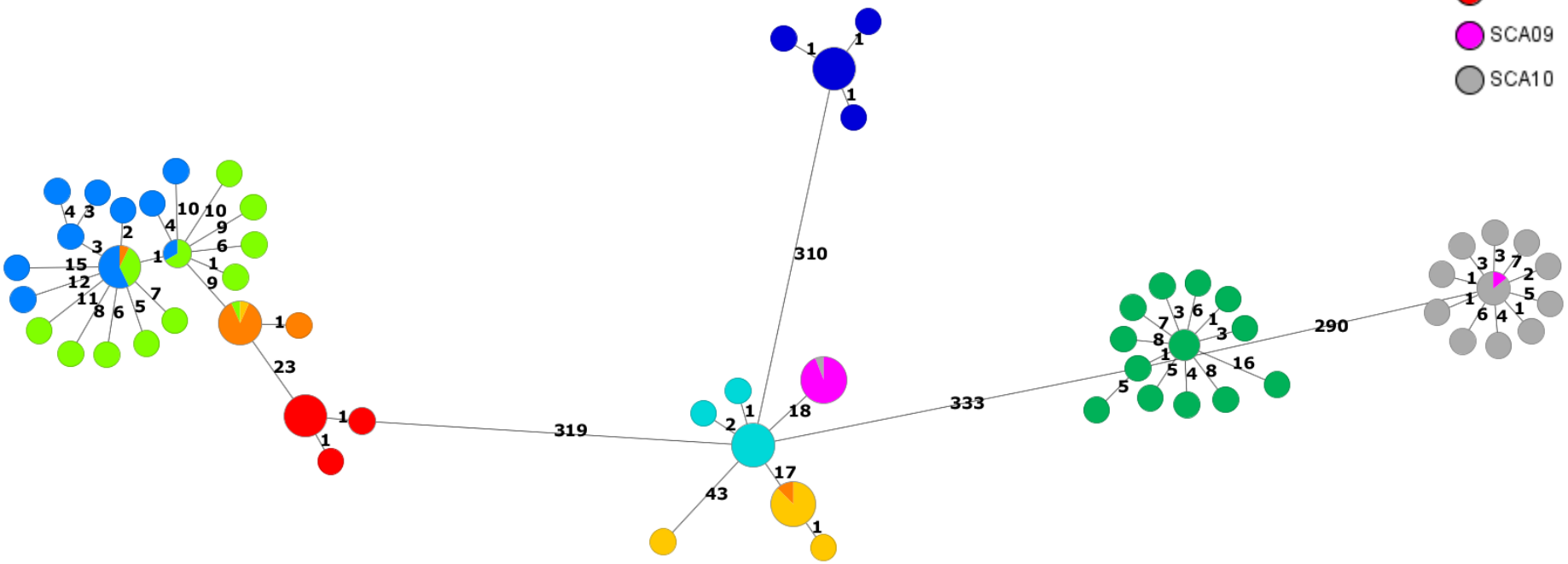
SCA02, SCA03, SCA06	Cluster 1, WGS-based analysis
11	Analysis of Lab11 WGS data shows that most likely strain SCA01 was processed twice: as SCA01 and as SCA02

- 1 x reported completely as expected
- 5 x SCA03 and SCA06 as 2 strains within 1 cluster
- 7 x cluster containing SCA03/SCA06, plus other(s)
- 6 x also the 3 (STMmono) strains SCA01, SCA05, SCA09 as a cluster
- 3 x all strains assigned into clusters

"Cluster" definition ?!

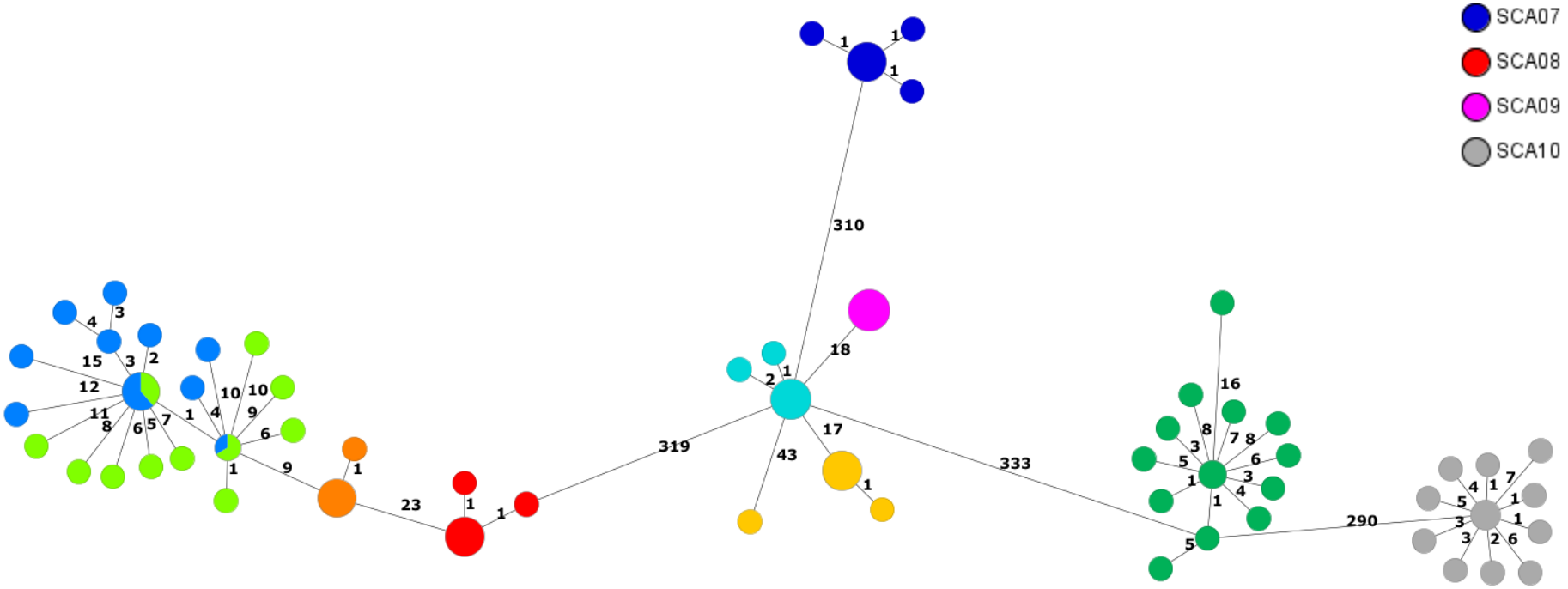
Evaluation WGS-based cluster analysis data

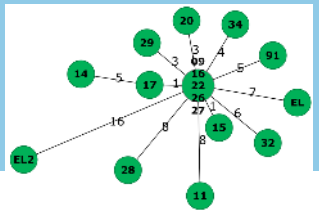
- MST of all strains from all participants' processed raw data
  - Assembly_pipeline; Ridom SeqSphere+, cgMLST including all 3002 targets, pairwise ignoring missing values; NB: Allele differences not on scale
  - Lab 17: swap between SCA01-SCA02 and SCA09-SCA10
  - Lab 16: swap between SCA02-SCA03
  - Lab 11: SCA01 processed as SCA01 and SCA02



Evaluation WGS-based cluster analysis data

- *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:*





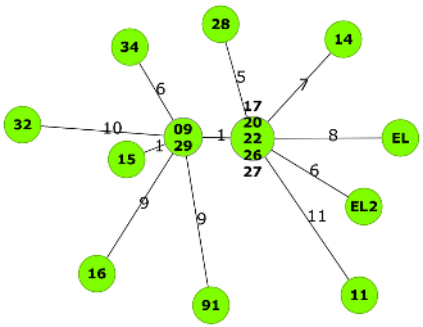
Genetic divergence within the SCA PT 2019 strains

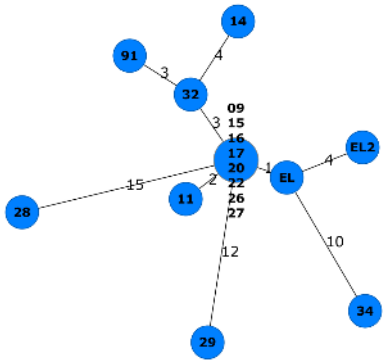
- Variability in PFGE results between original (<2015) and PT strains
- Variability in WGS results for PT strains, especially SCA03/SCA06, SCA04, SCA10
  - Technical reasons (e.g. variety in protocols used by participants)?
  - Biological reasons (e.g. instability of strains)?

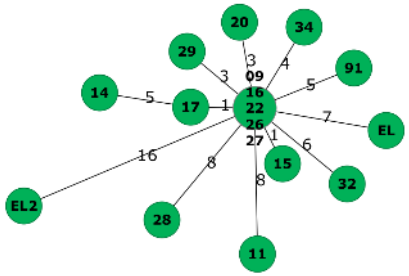
- Further investigations at the EURL-*Salmonella* (still ongoing)
  - technical/*in silico* track, WGS data will be dry-lab investigated in more detail
    - Quality assessment all WGS data (number and length of contigs, read depth)
    - Homogeneity check (subset will be mapped and visually investigated with either Artemis or IGV).
    - If possible, SNP analysis (subset will be investigated on micro-evolution in any other than MLST loci parts of the genome)
  - biological track, potential micro-evolution will be wet-lab investigated.
    - Variability in *fresh* strains (next PT), results after prolonged storage conditions (Transport tubes/-70°C)

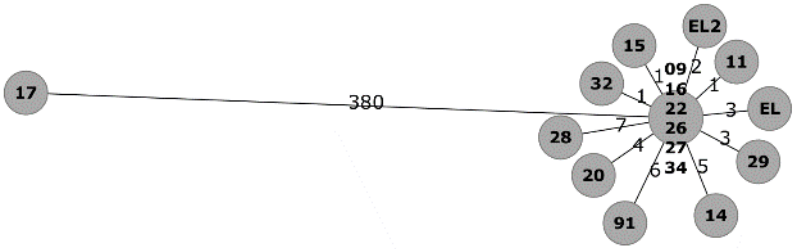
Further investigations, technical track example

- MST of strains SCA3, SCA6, SCA4, SCA10 from all participants' processed raw data (indicated in **bold** numbers)
  - EL: November 2019, EL2, February 2020
  - Assembly_pipeline; Ridom SeqSphere+, cgMLST including all 3002 targets, pairwise ignoring missing values; NB: Allele differences not on scale





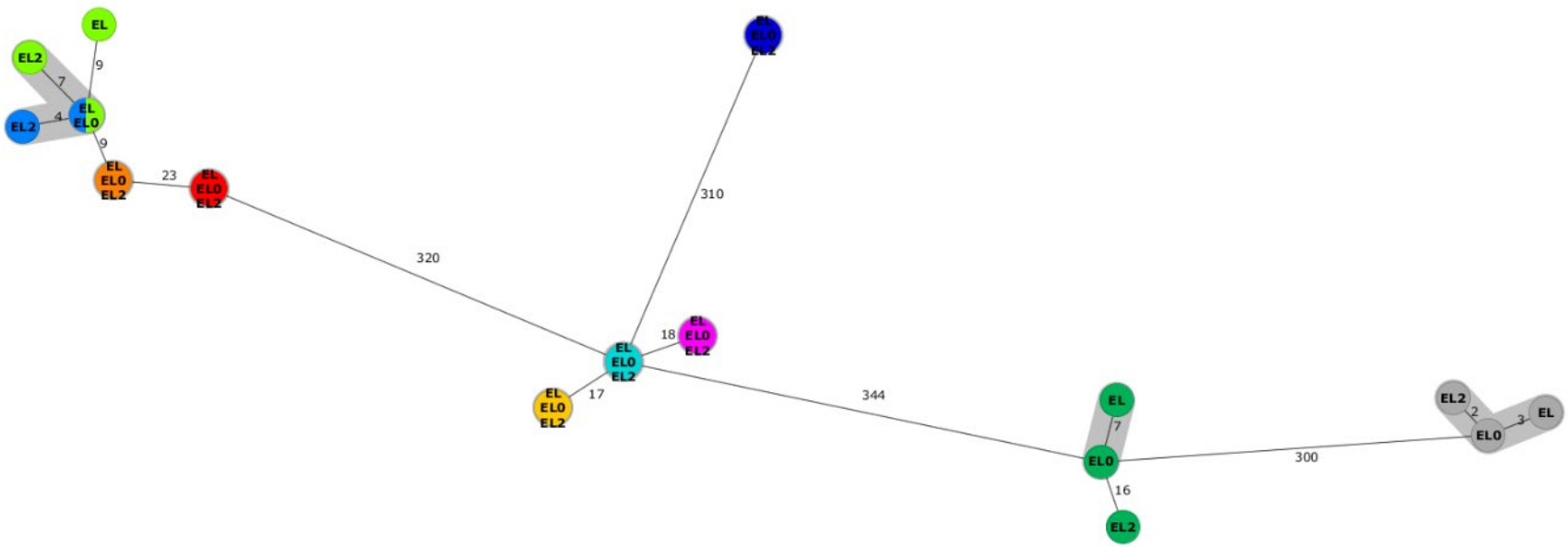




- E.g. What do Labs 22, 26, and 27 have in common?

Further investigations, biological track example

- MST of all strains, EURL-*Salmonella*
  - EL0: September 2019, EL: November 2019, EL2, February 2020
  - Assembly_pipeline; Ridom SeqSphere+, cgMLST including all 3002 targets, pairwise ignoring missing values; NB: Allele differences not on scale
- To be added:
  - EL3: September 2020, cultured from transport tubes (November 2019),
  - EL4: September 2020, cultured from -70°C (November 2019)



Conclusions first pilot PT Cluster Analysis

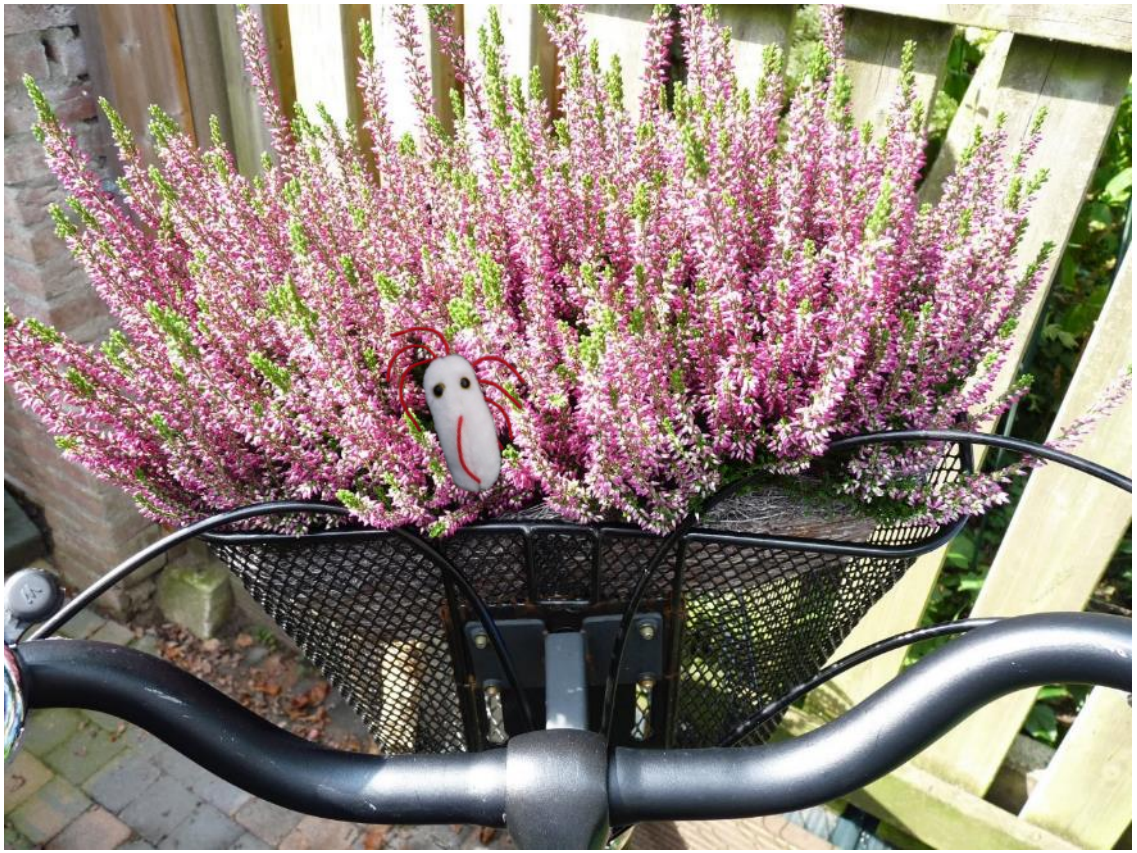
- 18 participants overall
  - per method: 6 PFGE, 8 MLVA, 14 WGS
- A lot of information and interesting data obtained!
- Strain selection and potential variability
- Cluster definitions
- No performance criteria were set for this PT
  - However, technical duplicates SCA03/SCA06 were expected to be assigned within 1 cluster
  - PFGE: 5/6, MLVA: 8/8, WGS: 12/14

- **Second pilot on (optional) Cluster Analysis in 2020**
  - using PFGE and/or MLVA and/or WGS
  - **Minimum of 5 participants per method required**
  - ***Defined* cluster analysis** on 10 *Salmonella* strains
    - Simulation of an outbreak-related request from the EURL-*Salmonella* (EFSA/ECDC) to the NRL-network
    - A WGS reference sequence may be provided

Provisional Planning of 25th Typing Study (2020)

Week	Date	Subject
39	Week of 21 September	Emailing of the link to the registration form for the typing study. Please **register by 16 October 2020** at the latest.
43	Week of 19 October	Emailing of the protocol 2020.
45	2 November	Shipment of the parcels to the participants as Biological Substance Category B (UN 3373).
45	Week of 2 November	*Upon receipt*: Starting the identification of the strains, according to the usual practice of the laboratory. Sending the link for the result form on serotyping to the participants. Sending the link for the result form on PFGE and/or MLVA and/or WGS Cluster Analysis to the participants in a separate email.
50	11 December 2020 at the latest	Deadline for completing the electronic submission of **serotyping** results: **11 December 2020** After this deadline, the result form for serotyping will be closed.
	29 January 2021 at the latest	Deadline for completing the electronic submission of **PFGE/MLVA/WGS Cluster Analysis** results: **29 January 2021**

Thank you for your attention and your participation !



Hartelijk dank !