EXTERNAL SCIENTIFIC REPORT

APPROVED: 13 December 2016

doi:10.2903/sp.efsa.2017.EN-1151

Closing gaps for performing a risk assessment on *Listeria monocytogenes* in ready-to-eat (RTE) foods: activity 3, the comparison of isolates from different compartments along the food chain, and from humans using whole genome sequencing (WGS) analysis

Eva Møller Nielsen[1], Jonas T. Björkman[1], Kristoffer Kiil[1], Kathie Grant[2], Tim Dallman[2], Anaïs Painset[2], Corinne Amar[2], Sophie Roussel[3], Laurent Guillier[3], Benjamin Félix[3], Ovidiu Rotariu[4], Francisco Perez-Reche[4], Ken Forbes[4], Norval Strachan[4]

[1]Statens Serum Institut, Copenhagen, Denmark; [2]Public Health England, Colindale, UK; [3]Anses, Maison-Alfort, France; [4]University of Aberdeen, UK

Abstract

This report presents the results of the project "Closing gaps for performing a risk assessment on *Listeria monocytogenes* in ready-to-eat (RTE) foods: activity 3, the comparison of isolates from different compartments along the food chain, and from humans using whole genome sequencing (WGS) analysis". The main objective was to compare *L. monocytogenes* isolates collected in the EU from ready-to-eat (RTE) foods, compartments along the food chain and from human cases by the use of WGS. A total of 1,143 *L. monocytogenes* isolates were selected for the study, including 333 human clinical isolates and 810 isolates from the food chain. The isolates were whole genome sequenced. The phylogeny showed a clear delineation between *L. monocytogenes* lineages and between clonal complexes within lineages. A range of typing methods were applied to the sequence data, providing the framework to answer questions on genetic diversity and epidemiological relationships. Retrospective analysis of nine outbreaks showed that WGS is a powerful tool in national and international outbreak investigations as WGS can accurately rule isolates in or out of outbreaks. Source attribution models showed bovine reservoir to be the main source of human disease although

Consortium partners – key persons

- Statens Serum Institut:
  - Eva Møller Nielsen, project leader
  - Jonas Larsson
  - Kristoffer Kiil
- Public Health England:
  - Tim Dallman
  - Kathie Grant
  - Anais Painset
  - Corinne Amar
- Anses:
  - Sophie Roussel
  - Laurent Guillier
  - Benjamin Felix
- University of Aberdeen:
  - Ken Forbes
  - Norval Strachan

Main objective

To compare *L. monocytogenes* isolates collected in the EU from RTE foods, compartments along the food chain and humans using whole genome sequencing (WGS) analysis.

Specific objective 1

- to carry out the molecular characterisation of a selection of *L. monocytogenes* isolates from different sources, i.e. RTE foods, compartments along the food chain (e.g. food producing animals, food processing environment), and humans employing WGS analysis.

In total 1143 genomes were analysed:

Food, RTE baseline study

Food, other

Food production chain

Sporadic clinical

Outbreaks, human and food

Specific objective 2

Specific objective 2: to analyse the WGS typing data of the selected *L. monocytogenes* isolates with three goals:

i. to explore the genetic diversity of *L. m* within and between the different sources and human origin;

ii. to assess the epidemiological relationship of *L. monocytogenes* from the different sources and of human origin considering the genomic information and the metadata available for each isolate;

iii. to identify the presence of putative markers conferring the potential to survive/multiply in the food chain and/or cause disease in humans (e.g. virulence and antimicrobial resistance).

Specific objective 3

Specific objective 3: to perform a retrospective analysis of outbreak strains (i.e. using a subset of epidemiologically linked human and food isolates) to investigate the suitability of WGS as a tool in outbreak investigations.

The outcome of this analysis should provide an evaluation on the advantages and limitations of employing WGS data for investigating outbreaks of food-borne listeriosis.

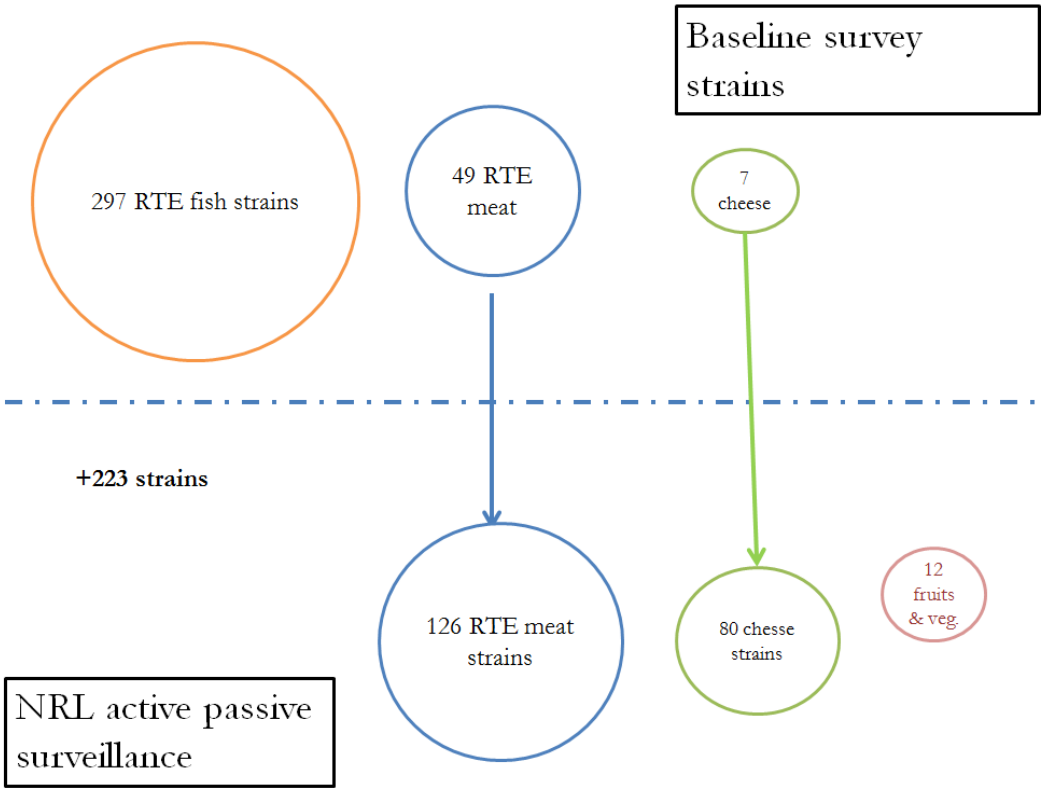
Report contents

1. Introduction
2. Isolate collection
3. Methodologies
4. Sequencing and Phylogenetic Analysis
5. Retrospective analysis of outbreaks
6. Genetic diversity
7. Epidemiological relationship: Source Attribution
8. Epidemiological relationship – linking of genetically related isolates
9. Putative markers
10. Conclusion

2. Isolate collection

Isolates from different food sources

- Baseline EFSA survey (mainly smoked fish)
- Consortium provided 223 additional strains



Isolates from different food sources

- Baseline EFSA survey (mainly smoked fish)
- + Consortium provided 223 additional strains (annexes 2+3)

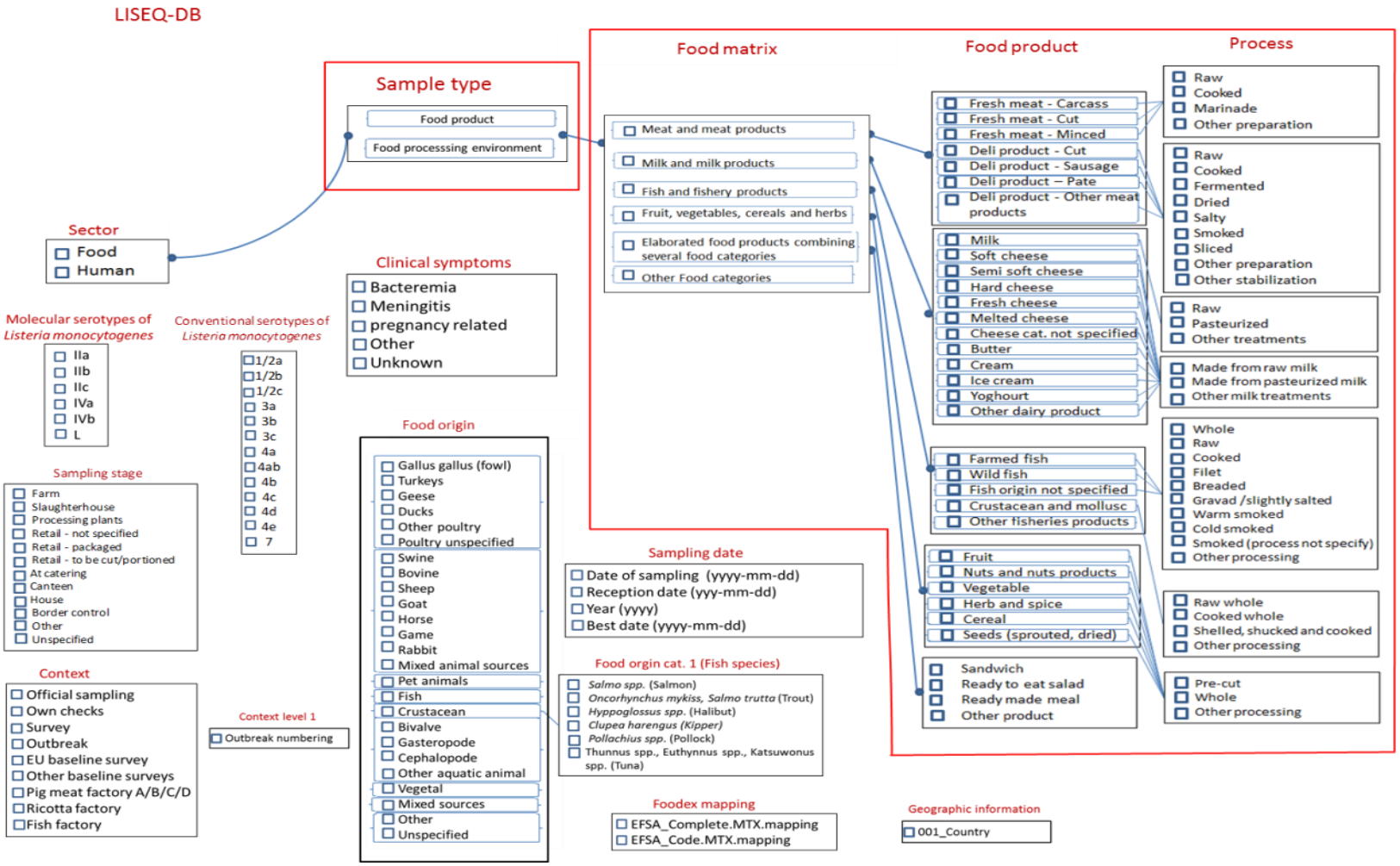
+ Food production chain strains (for persistence markers, see part 9)

Clinical isolates

- Sporadic (for source attribution, epidemiological investigation): 262 isolates
- Strains from outbreaks

	Human	Food	Source of infection
Outbreak 1	5	10	Beef
Outbreak 2	5	3	Crab meat
Outbreak 3	5	4	Sandwiches
Outbreak 4	2	2	Ox tongue
Outbreak 5	9	1	Unknown
Outbreak 6	4	1	Rakfisk
Outbreak 7	13	6	Foie gras
Outbreak 8	4	9	Cheese
Outbreak 9	25	0	Brie cheese

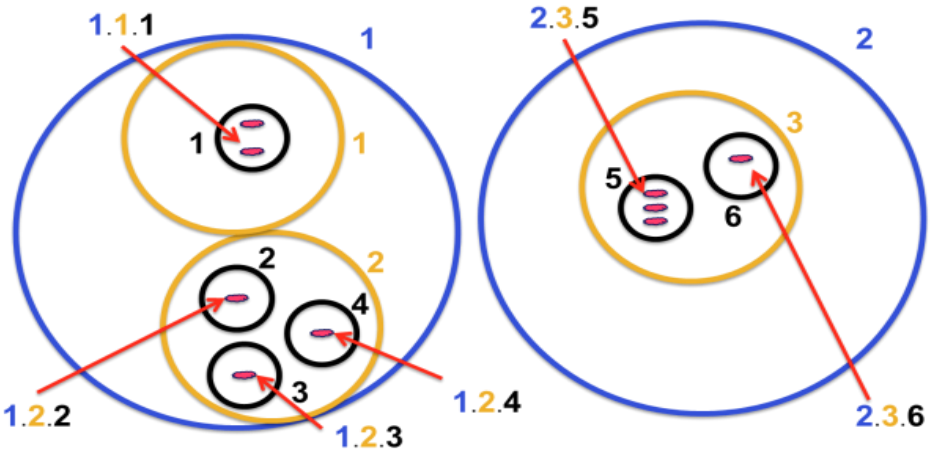
Database

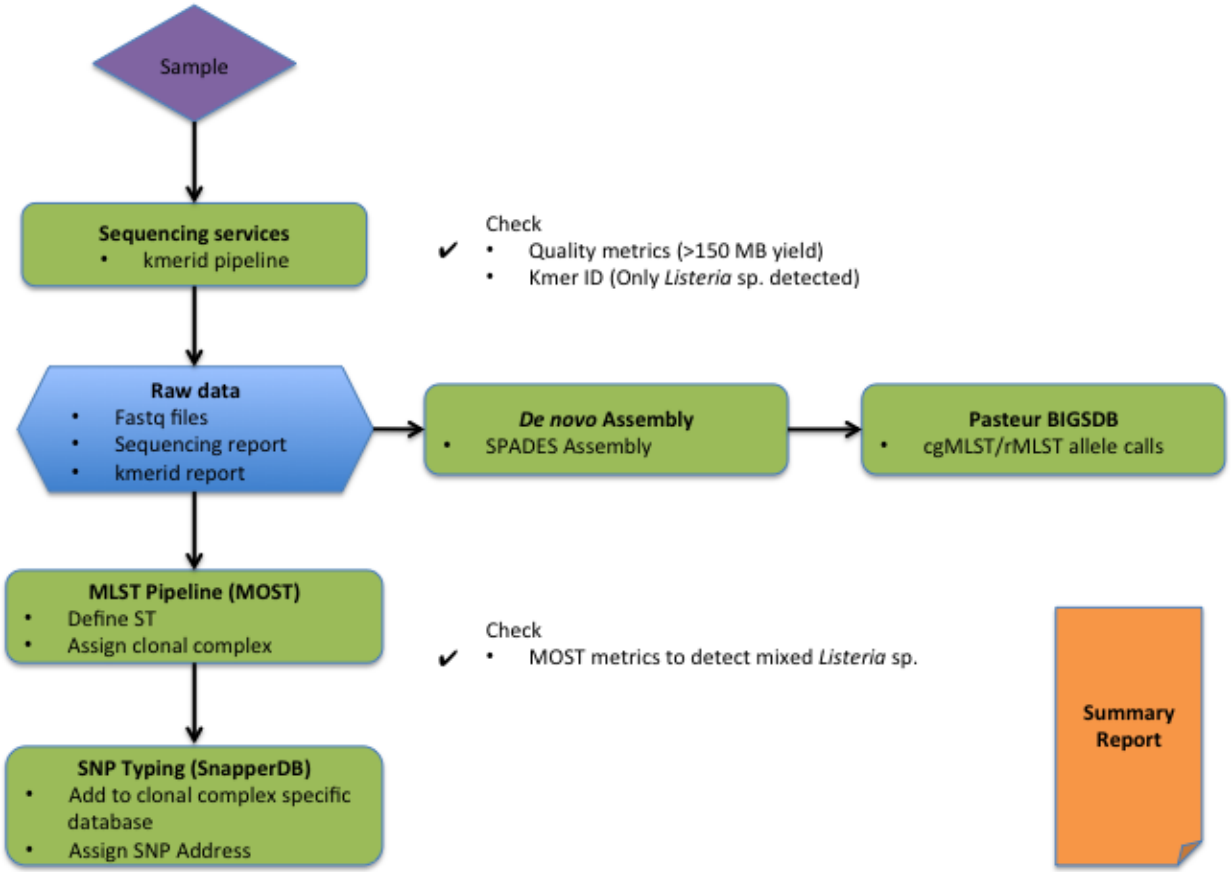

LISEQ-DB
Sample type
Food product
Food processsing environment
Sector
Food
Human
Clinical symptoms
Bacteremia
Meningitis
pregnancy related
Other
Unknown
Molecular serotypes of Listeria monocytogenes
IIa
IIb
IIc
IVa
IVb
L
Conventional serotypes of Listeria monocytogenes
1/2a
1/2b
1/2c
3a
3b
3c
4a
4ab
4b
4c
4d
4e
7
Sampling stage
Farm
Slaughterhouse
Processing plants
Retail - not specified
Retail - packaged
Retail - to be cut/portioned
At catering
Canteen
House
Border control
Other
Unspecified
Context
Official sampling
Own checks
Survey
Outbreak
EU baseline survey
Other baseline surveys
Pig meat factory A/B/C/D
Ricotta factory
Fish factory
Context level 1
Outbreak numbering
Food origin
Gallus gallus (fowl)
Turkeys
Geese
Ducks
Other poultry
Poultry unspecified
Swine
Bovine
Sheep
Goat
Horse
Game
Rabbit
Mixed animal sources
Pet animals
Fish
Crustacean
Bivalve
Gasteropode
Cephalopode
Other aquatic animal
Vegetal
Mixed sources
Other
Unspecified
Food matrix
Meat and meat products
Milk and milk products
Fish and fishery products
Fruit, vegetables, cereals and herbs
Elaborated food products combining several food categories
Other Food categories
Food product
Fresh meat - Carcass
Fresh meat - Cut
Fresh meat - Minced
Deli product - Cut
Deli product - Sausage
Deli product – Pate
Deli product - Other meat products
Milk
Soft cheese
Semi soft cheese
Hard cheese
Fresh cheese
Melted cheese
Cheese cat. not specified
Butter
Cream
Ice cream
Yoghourt
Other dairy product
Farmed fish
Wild fish
Fish origin not specified
Crustacean and mollusc
Other fisheries products
Fruit
Nuts and nuts products
Vegetable
Herb and spice
Cereal
Seeds (sprouted, dried)
Sandwich
Ready to eat salad
Ready made meal
Other product
Process
Raw
Cooked
Marinade
Other preparation
Raw
Cooked
Fermented
Dried
Salty
Smoked
Sliced
Other preparation
Other stabilization
Raw
Pasteurized
Other treatments
Made from raw milk
Made from pasteurized milk
Other milk treatments
Whole
Raw
Cooked
Filet
Breaded
Gravad /slightly salted
Warm smoked
Cold smoked
Smoked (process not specify)
Other processing
Raw whole
Cooked whole
Shelled, shucked and cooked
Other processing
Pre-cut
Whole
Other processing
Sampling date
Date of sampling (yyyy-mm-dd)
Reception date (yyy-mm-dd)
Year (yyyy)
Best date (yyyy-mm-dd)
Food orgin cat. 1 (Fish species)
Salmo spp. (Salmon)
Oncorhynchus mykiss, Salmo trutta (Trout)
Hyppoglossus spp. (Halibut)
Clupea harengus (Kipper)
Pollachius spp. (Pollock)
Thunnus spp., Euthynnus spp., Katsuwonus spp. (Tuna)
Foodex mapping
EFSA_Complete.MTX.mapping
EFSA_Code.MTX.mapping
Geographic information
001_Country

3. Methodologies – to be presented within each result section

4. Sequencing and Phylogenetic Analysis

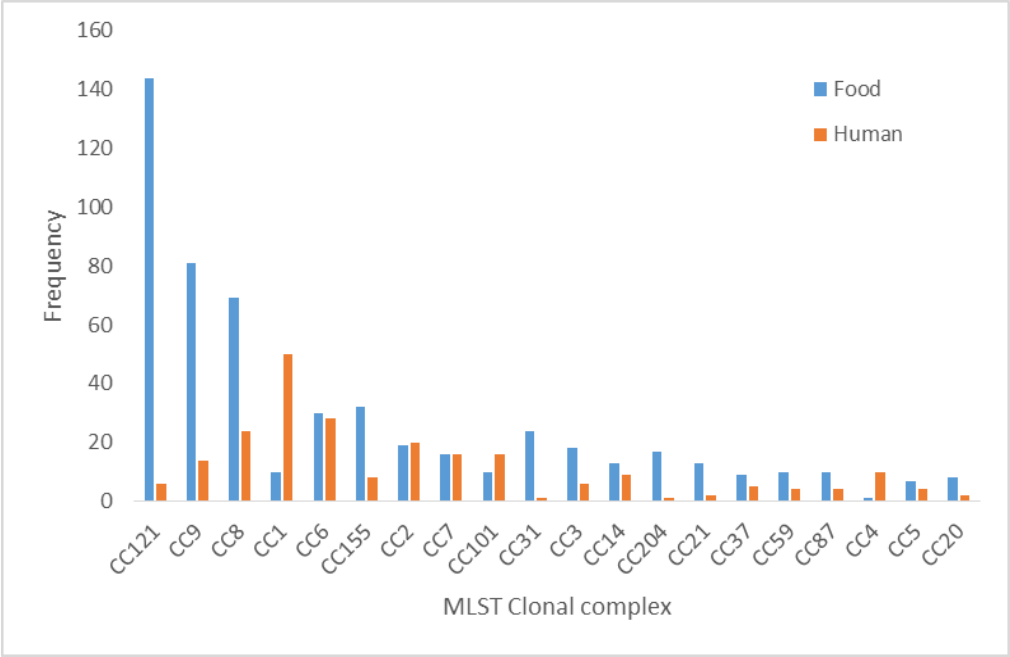
Sequencing and Phylogenetic Analysis



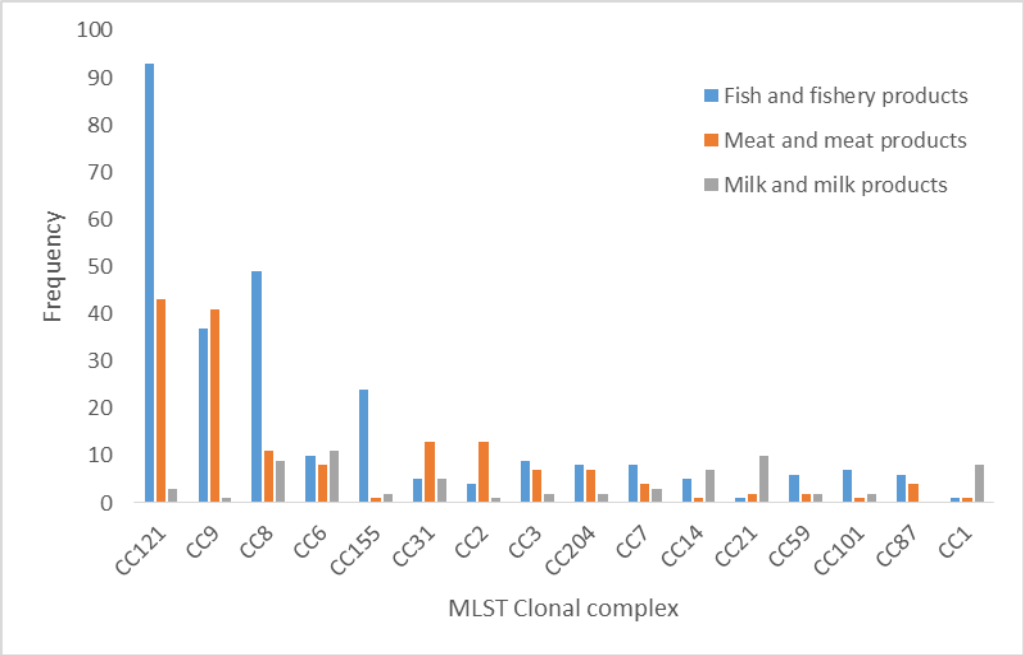


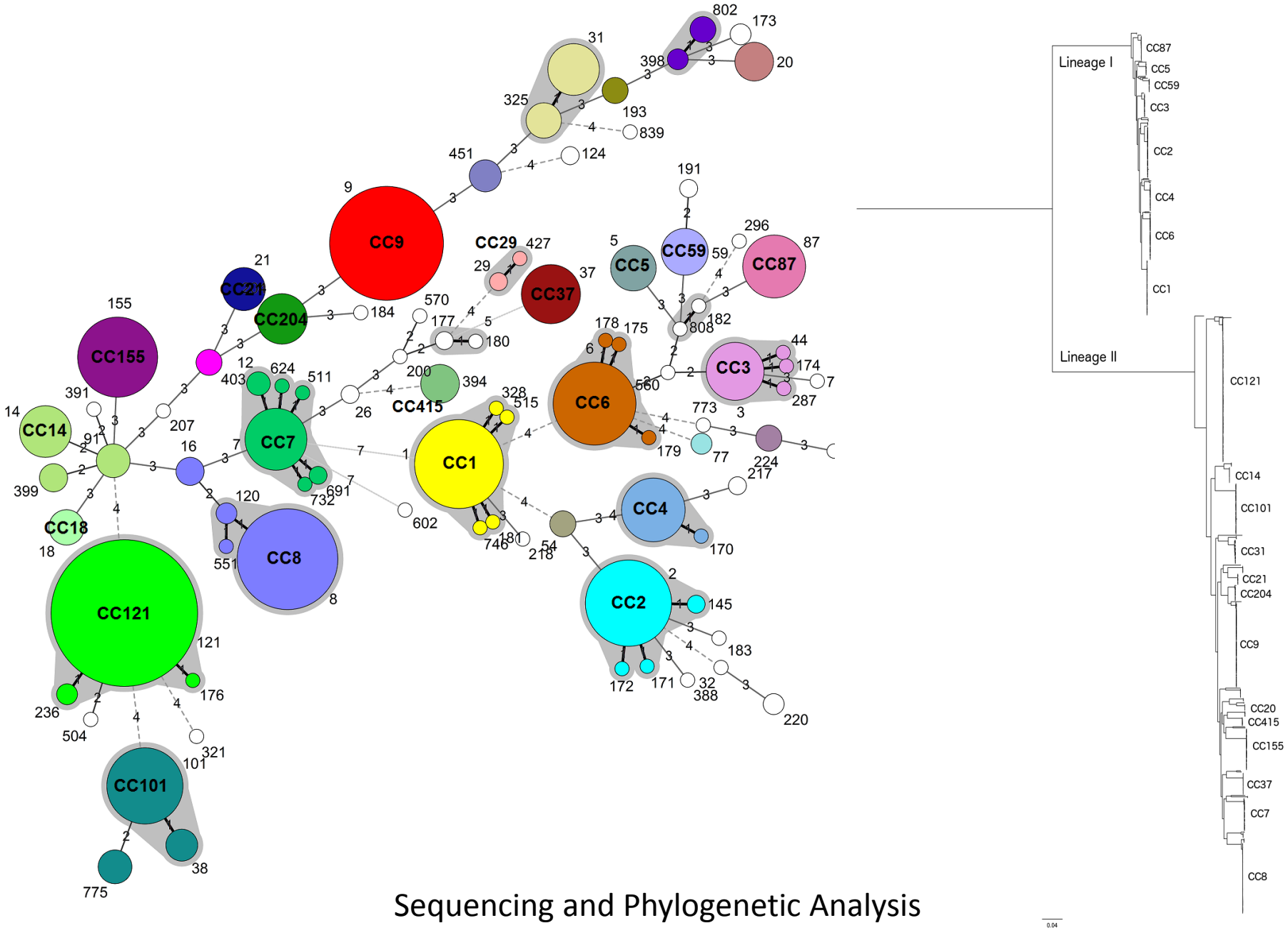
Sequencing Analysis

Distribution of clonal complexes in ready-to-eat food and from sporadic human infections



Distribution of clonal complexes from the three major food product categories





Sequencing and Phylogenetic Analysis

5. Retrospective analysis of outbreaks

Outbreaks

- Nine outbreaks from four countries
- Defined by epidemiology and current generation typing (PFGE or fAFLP)
- National outbreaks without a previously known international component

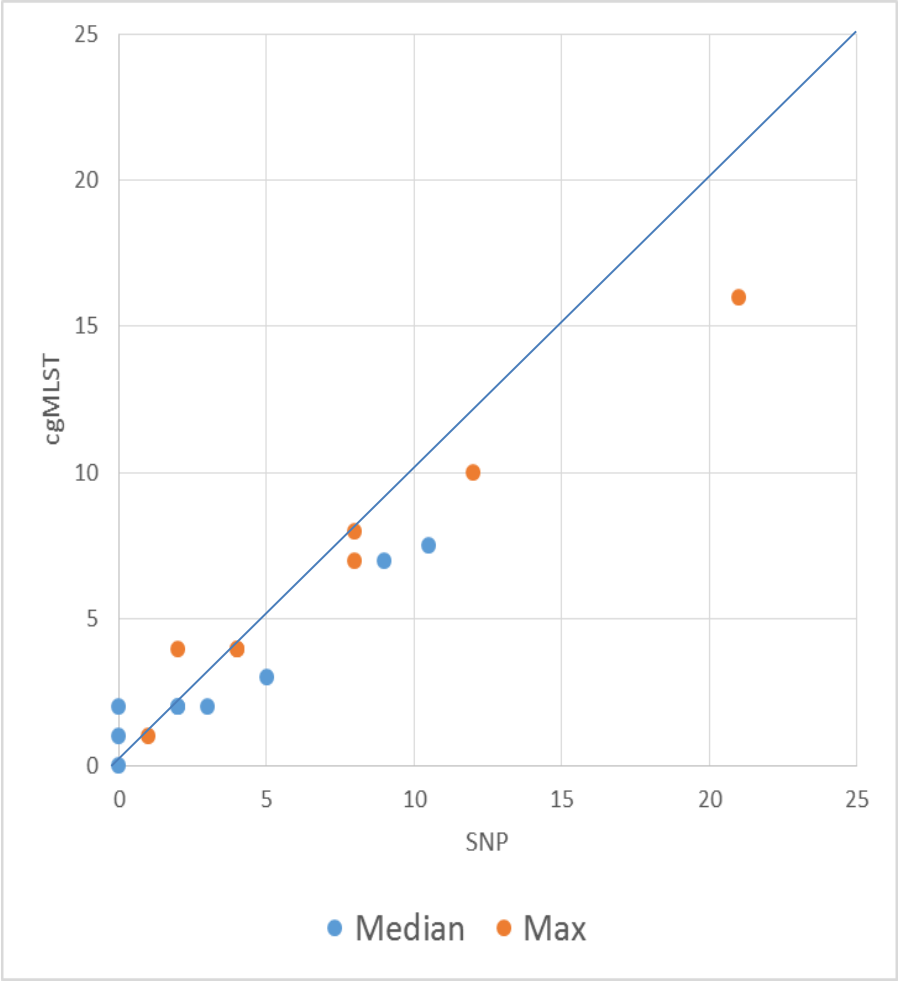
Methods

- SNP analysis based on pairwise comparisons produced by PHE
  - SNP trees are maximum parsimony trees with only SNP positions covered in all genomes.
  - SNP distance matrices are based on pairwise comparisons with pairwise deletions in case of an ambiguous base call.
- cgMLST is based on the 1748 loci Pasteur scheme
  - cgMLST trees are Minimum Spanning trees

Outbreak number		#1	#2	#3	#4	#5	#6	#7	#8	#9
	Clonal complex	CC155	CC1	CC7	CC59	CC415	CC398	CC87	ST14	CC4
	Years	2012-13	2007-13	2013-14	2009-11	2013-14	2013	2013-14	2012	2012
All isolates within CC	n human / food	13/41	55/17	22/20	6/12	9/3	8/1	17/16	6/16	34/2
SNP	Median	111	174	252	90	3	38	38	214	126
cgMLST	Median	57	73	59	46	16	19,5	20	23	54
SNP	Max	174	259	1358	243	93	65	74	337	183
cgMLST	Max	118	119	113	119	50	36	41	47	88
Outbreak isolates	n human / food	5/8	5/3	4/4	2/2	8/1	4/1	13/6	4/9	24/0
SNP	Median	0	10,5	2	9	3	0	5	2	0
cgMLST	Median	1	7,5	2	7	2	0	3	2	2
SNP	Mode	0	16	1	9	3	0	5	0	0
cgMLST	Mode	0	5	2	10	1	0	3	1	2
SNP	Max	2	21	4	12	4	1	8	8	2
cgMLST	Max	51	16	4	10	4	1	7	8	4

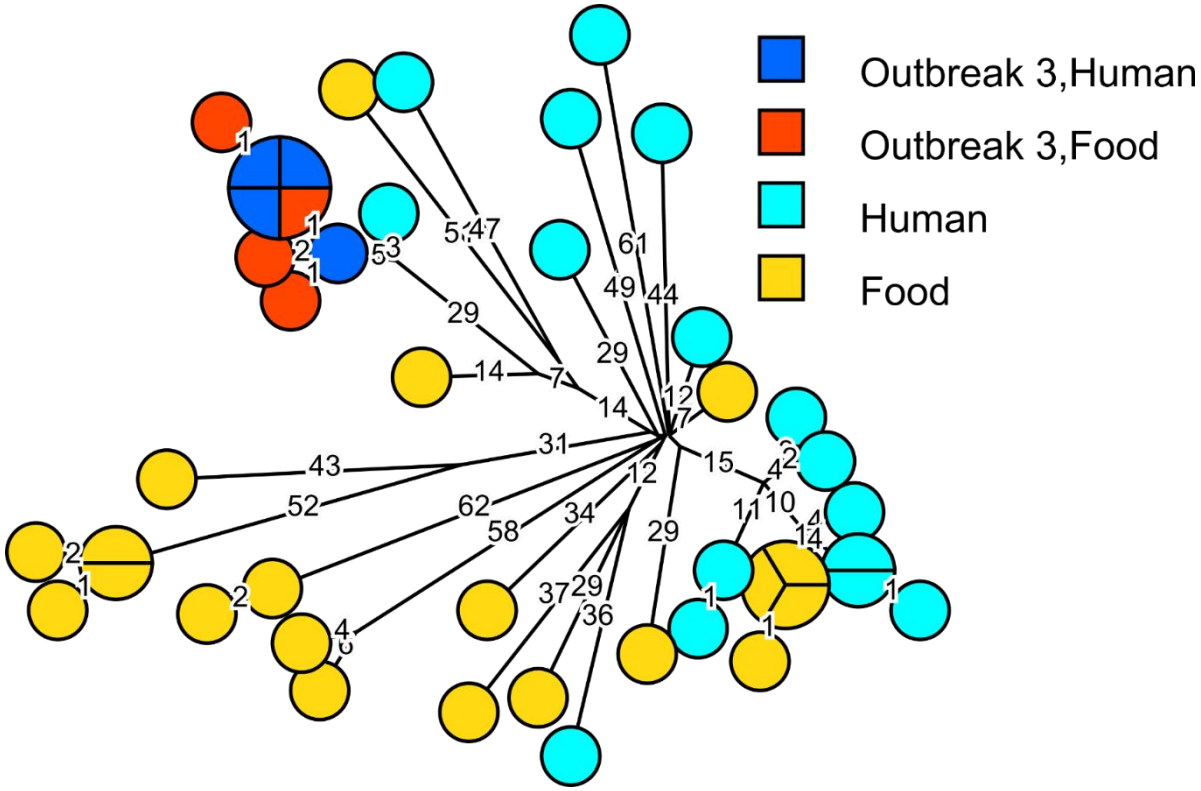
SNP vs cgMLST

- One outlier not shown
- Good concordance
- Slightly more resolution on the SNP analysis



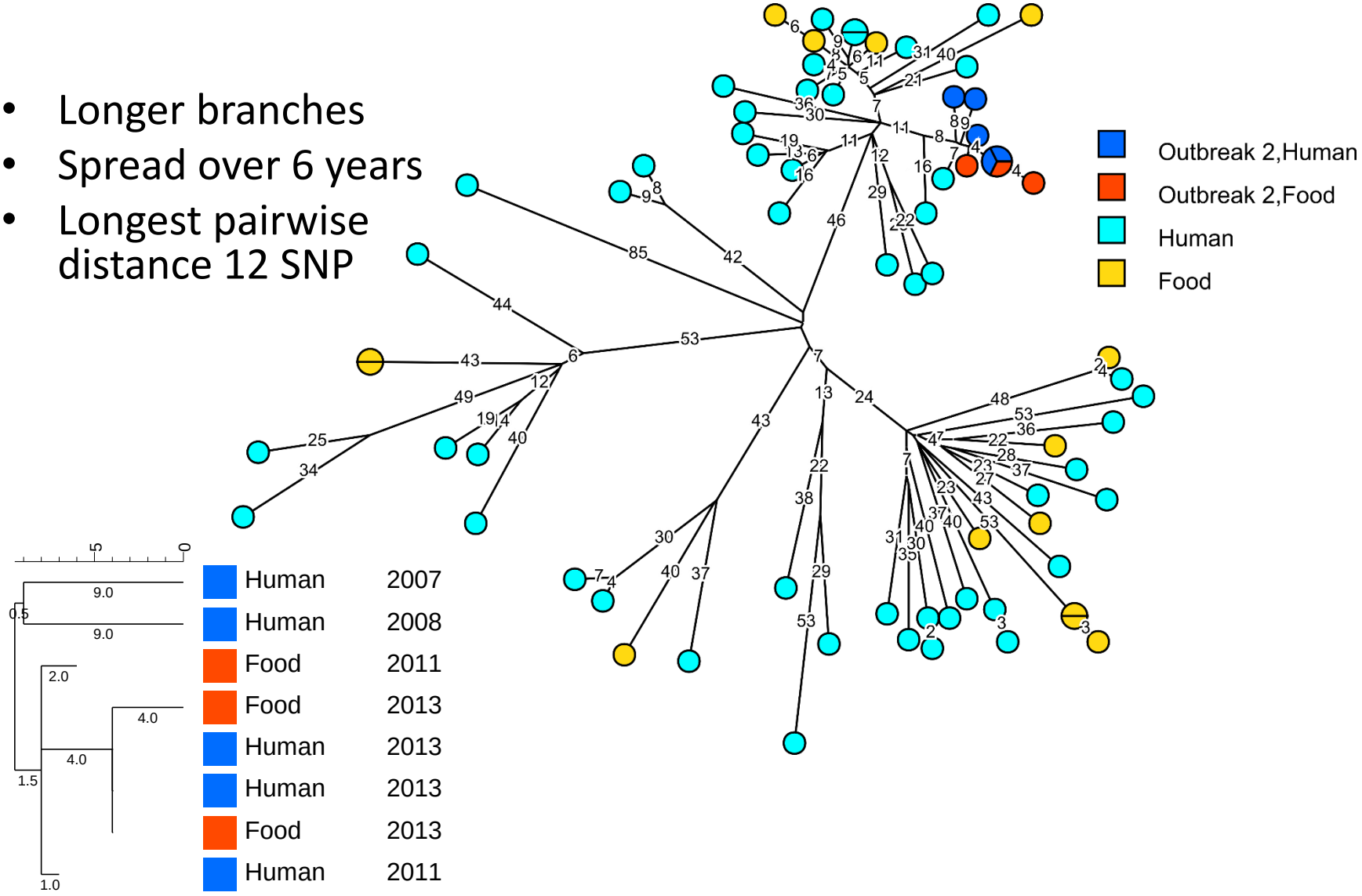
Point source outbreak

- Short branches
- One extra isolate possibly connected



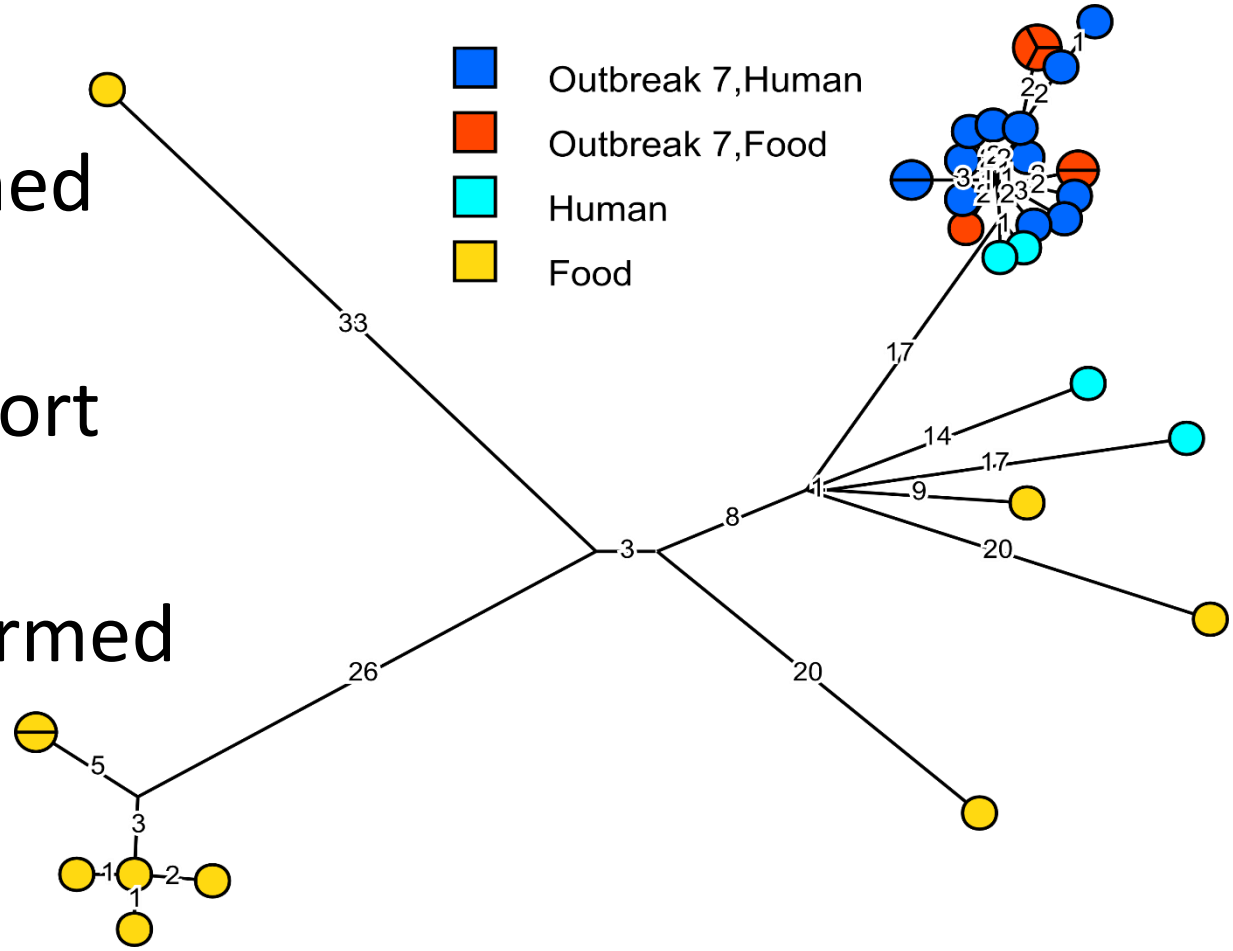
Extended outbreak in time

- Longer branches
- Spread over 6 years
- Longest pairwise distance 12 SNP



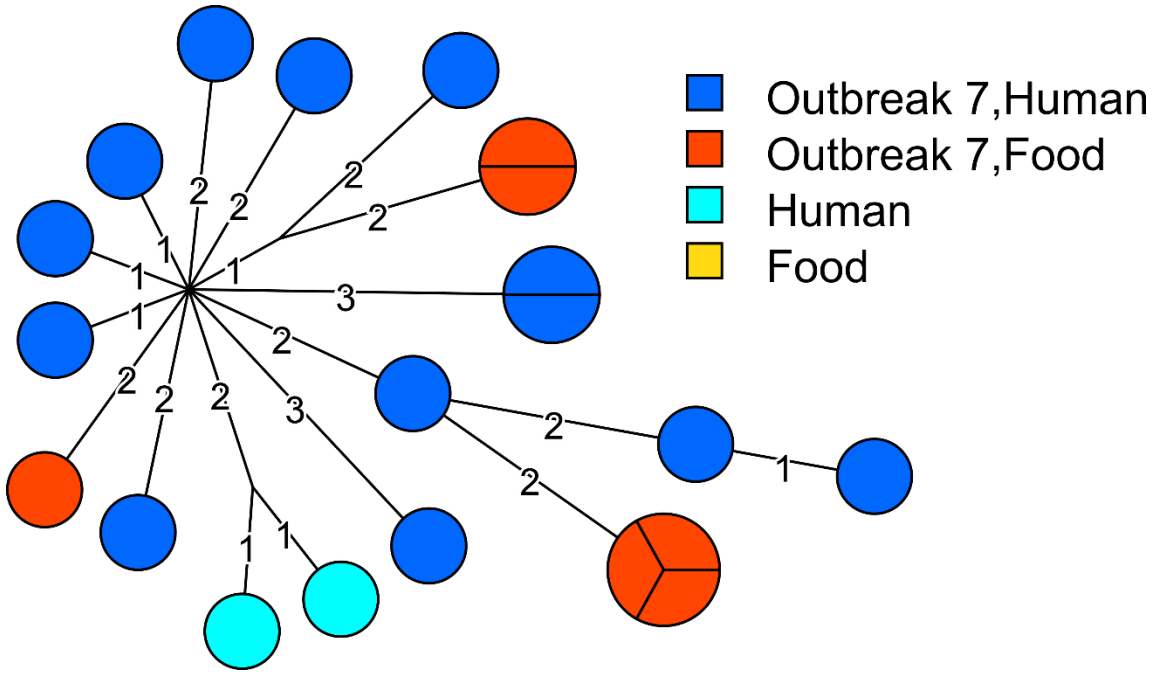
Extended distances, but not in time

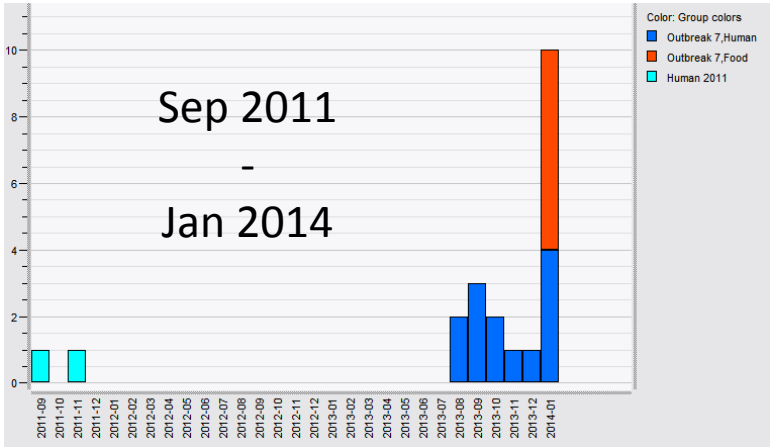
- Rare type with confirmed outbreak
- Relatively short time
- Source confirmed by epi



Extended distances, but not in time

- Very few isolates are identical
- Still an outbreak over a relatively short time





Outbreak Conclusions

- The WGS analysis corroborated the epidemiology
  - 6/9 clusters show a typical point-source-like pattern
    - Median pairwise distance of ≤5 SNP
    - Maximum pairwise distance ≤10 SNP
  - Two outbreaks are extended in time and show more variation
  - One shows more variation, but is not extended in time
- No international aspects detected
- No extra matching food isolates detected
- cgMLST concordant with SNP

6. Genetic diversity

Analyses Carried Out

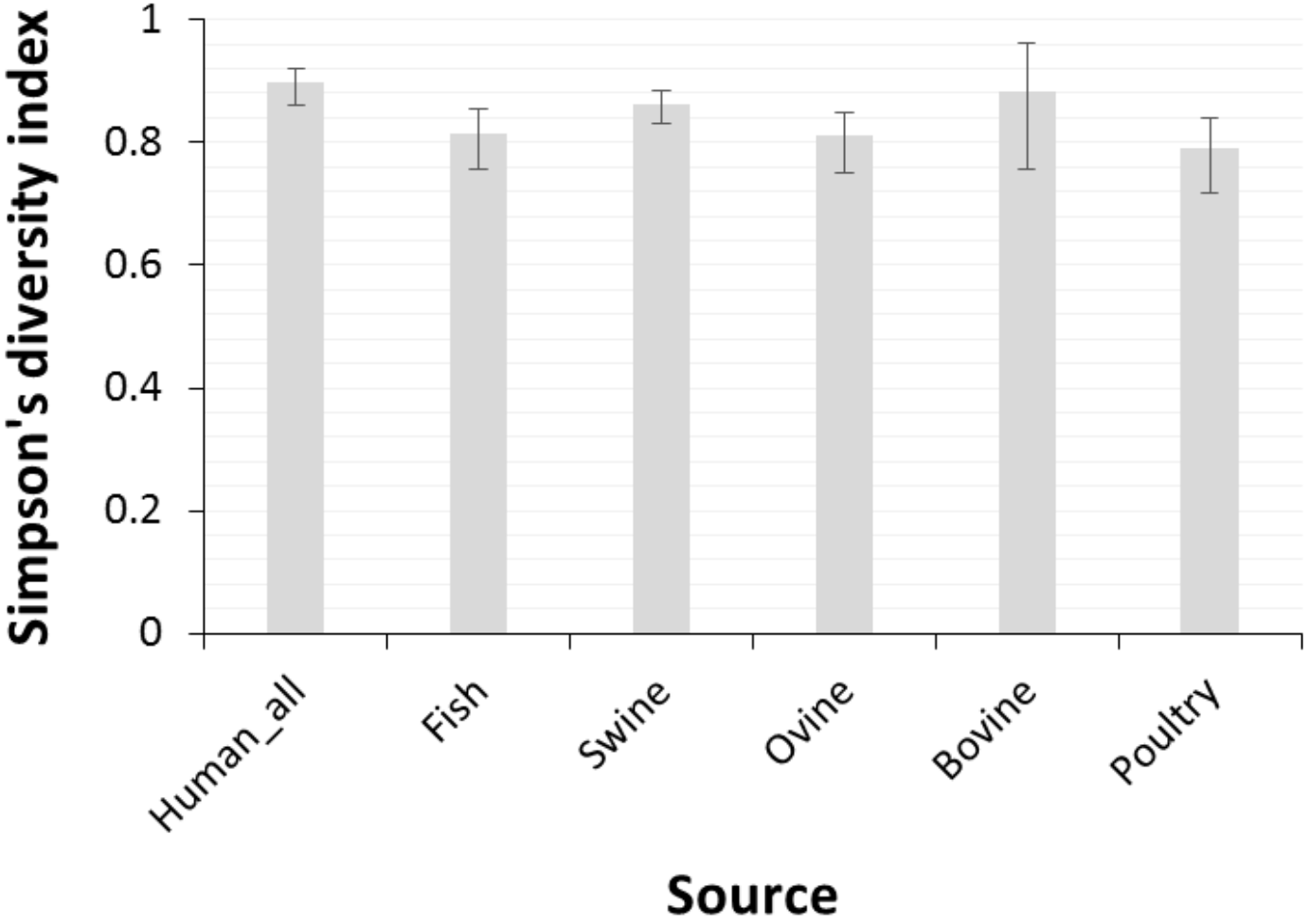
Level of molecular analysis	Simpson's Diversity	Rarefaction	Nei's genetic distance
7 locus MLST	√	√	√
30 locus rMLST	√	√	√
1748 cgMLST			√
39,529 cgSNP			√

Genomes Used

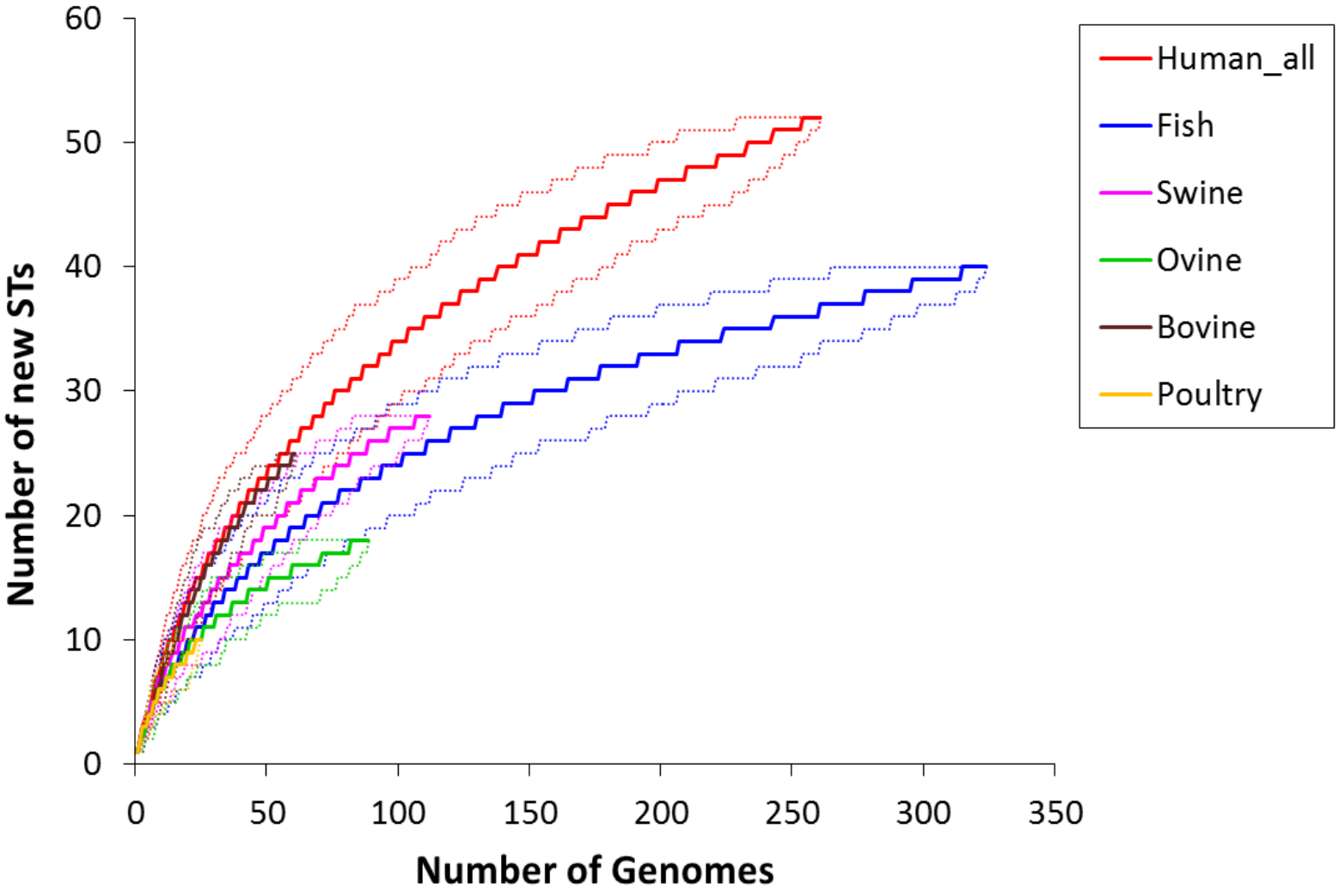
Human and source	Number of Genomes	Number of genomes with 7 locus MLST and not part of an outbreak
Human*	333	261
Mixed	30	27
Poultry*	32	25
Bovine*	80	61
Shellfish	3	0
Swine*	114	112
Fish*	325	324
Unspecified	101	101
Vegetable	5	5
Ovine*	117	89
Caprine	3	3
Total	1143	1011

*denotes used in source attribution comprising 872 genomes

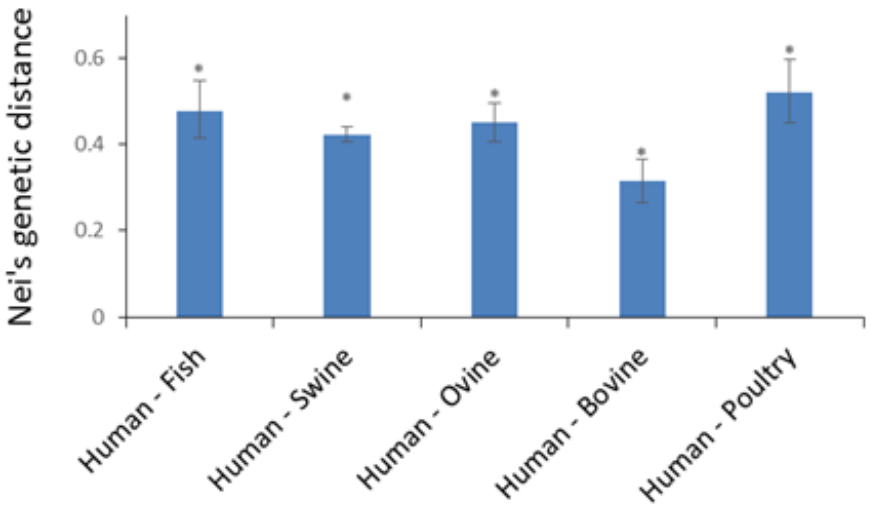
Simpson's diversity Index for 7 locus MLST

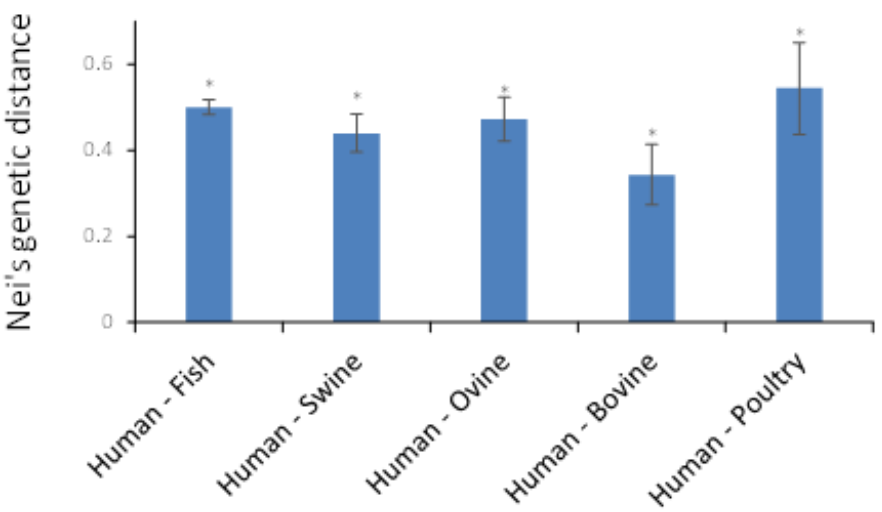


Rarefaction for 7 locus MLST



Nei for 7 locus MLST & 1748 locus cgMLST



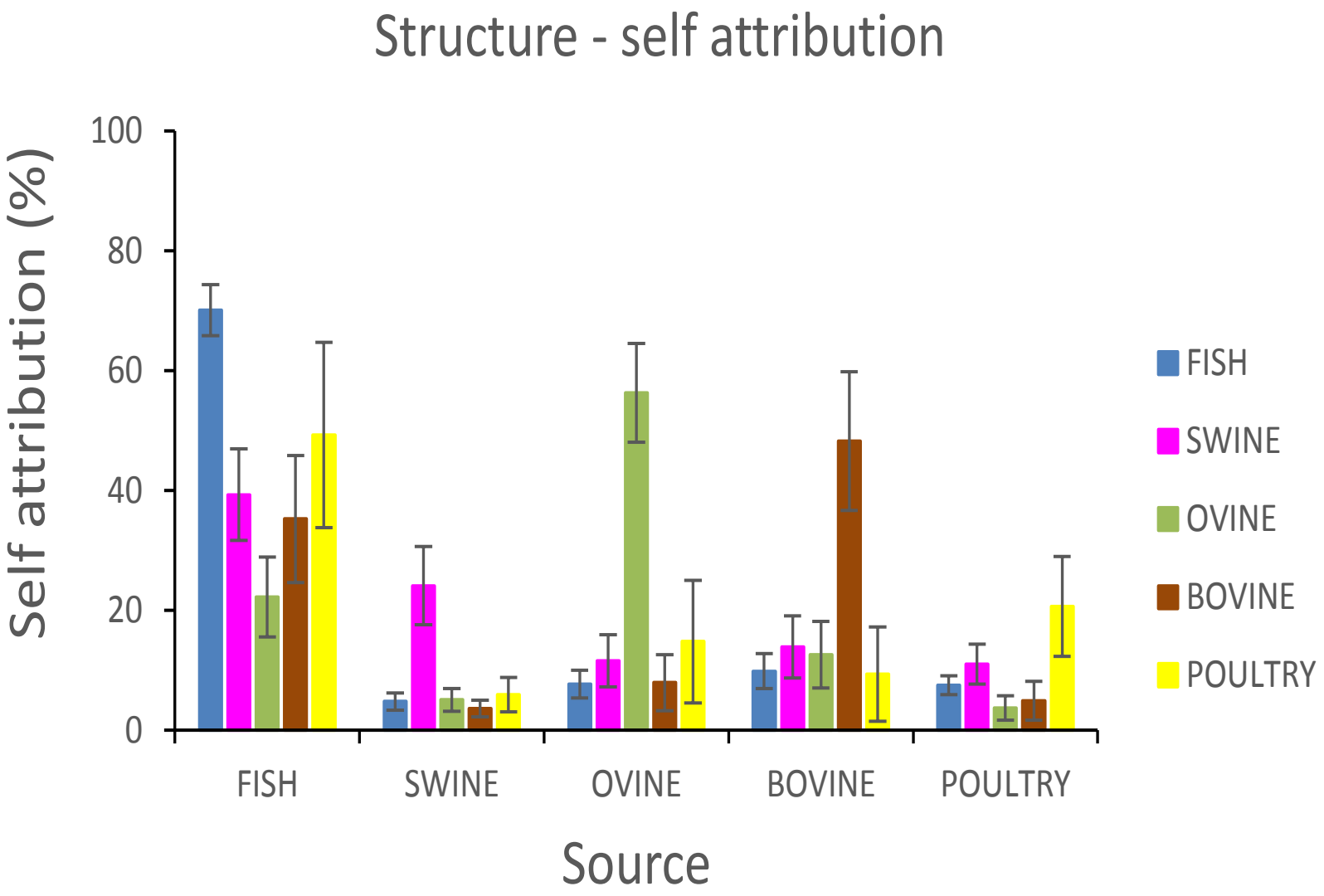


7. Epidemiological relationship: Source Attribution

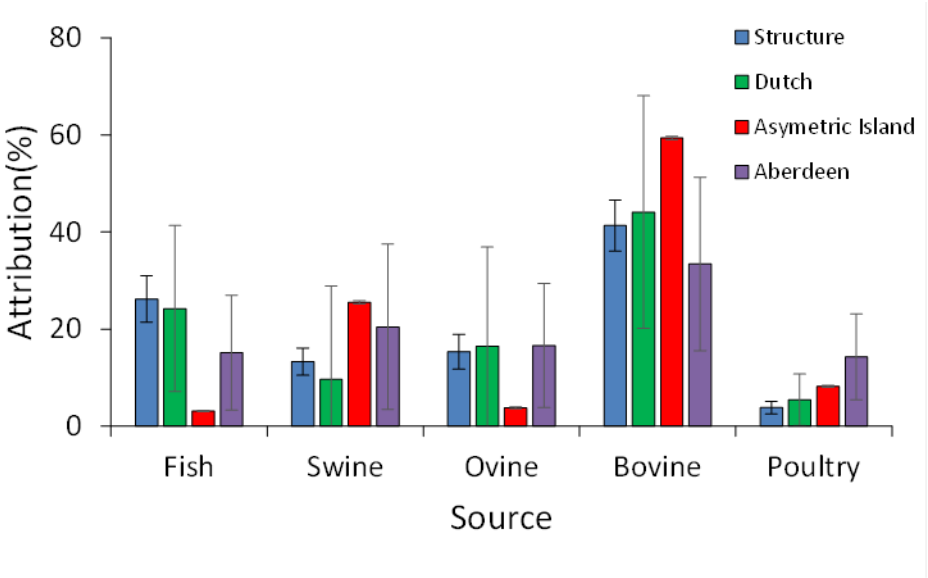
Analyses performed

Number of sources	Level of molecular analysis (number of loci)	STRUCTURE	Dutch	Asymmetric Island	Hald	Aberdeen
5 sources	ST(1)	nd	nd	nd	√	nd
	MLST(7)	√	√	√	np	√
	rMLST(30)	√	√	√	np	√
	cgMLST(1748)	√	√	np	np	√
	cgSNP(15,000 Dutch, 39,529 Aberdeen)	np	√	np	np	√
4 sources (excluding poultry)	ST(1)	nd	nd	nd	nd	nd
	MLST(7)	√	√	√	np	√
	rMLST(30)	√	√	√	np	√
	cgMLST(1748)	√	√	np	np	√
	cgSNP(15,000 Dutch, 39,529 Aberdeen)	np	√	np	np	√

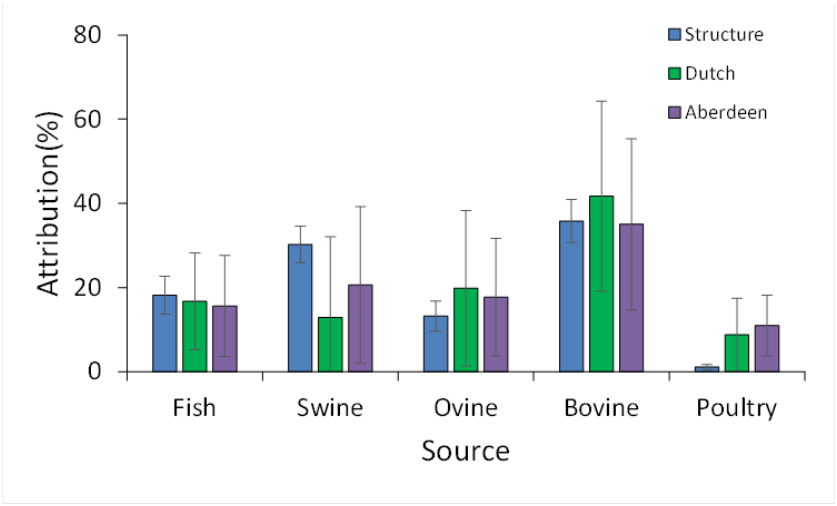
Self Attribution



Source attribution (5 sources)



30 locus rMLST

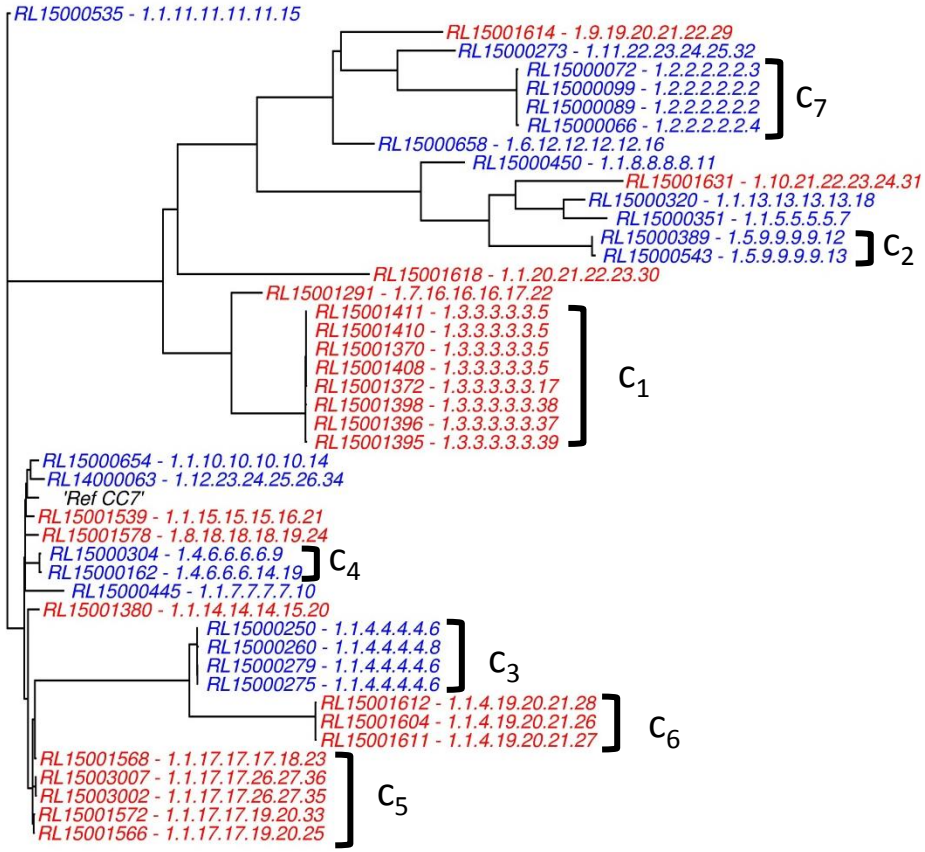


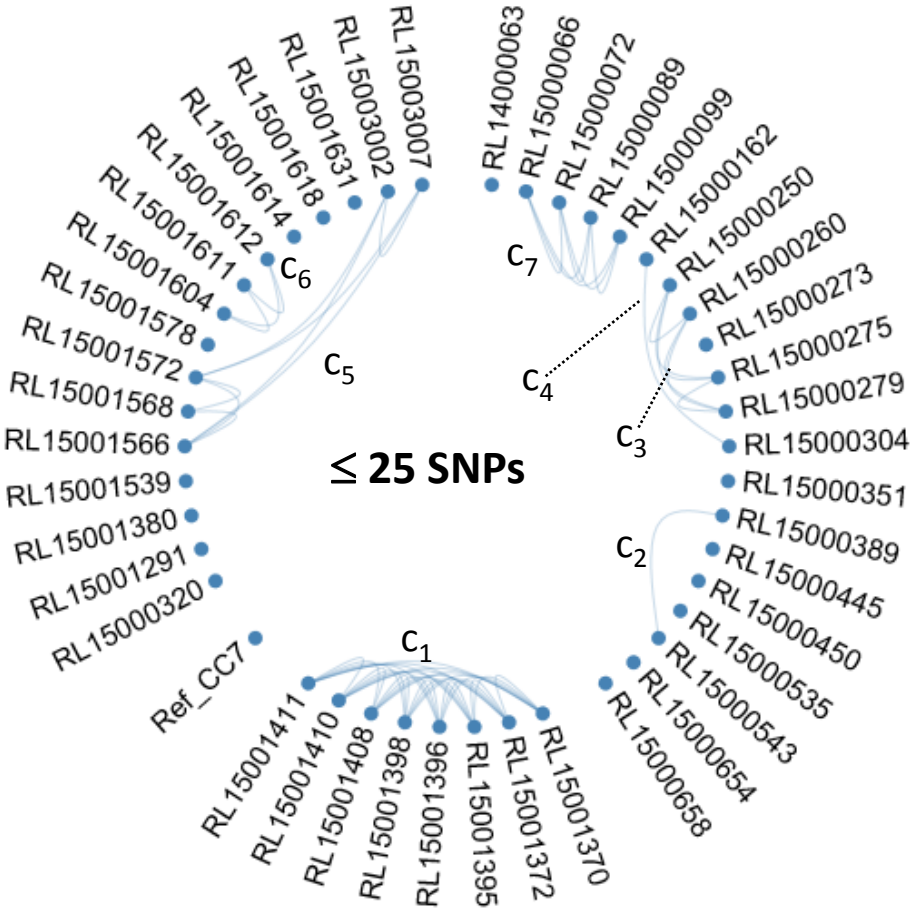
1748 locus cgMLST

8. Epidemiological relationship – linking of genetically related isolates

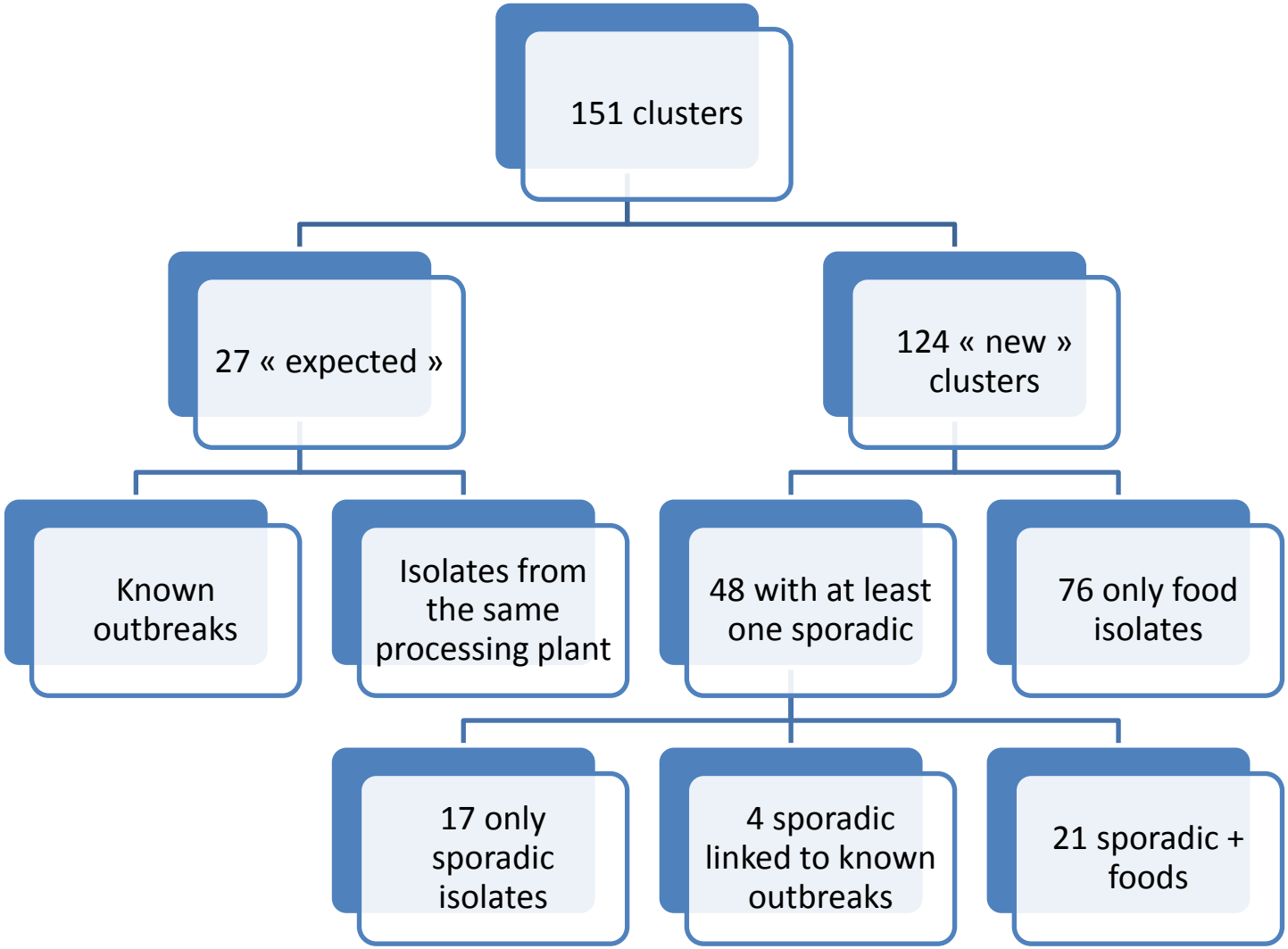
Definition of genetically clustered strains

- Method



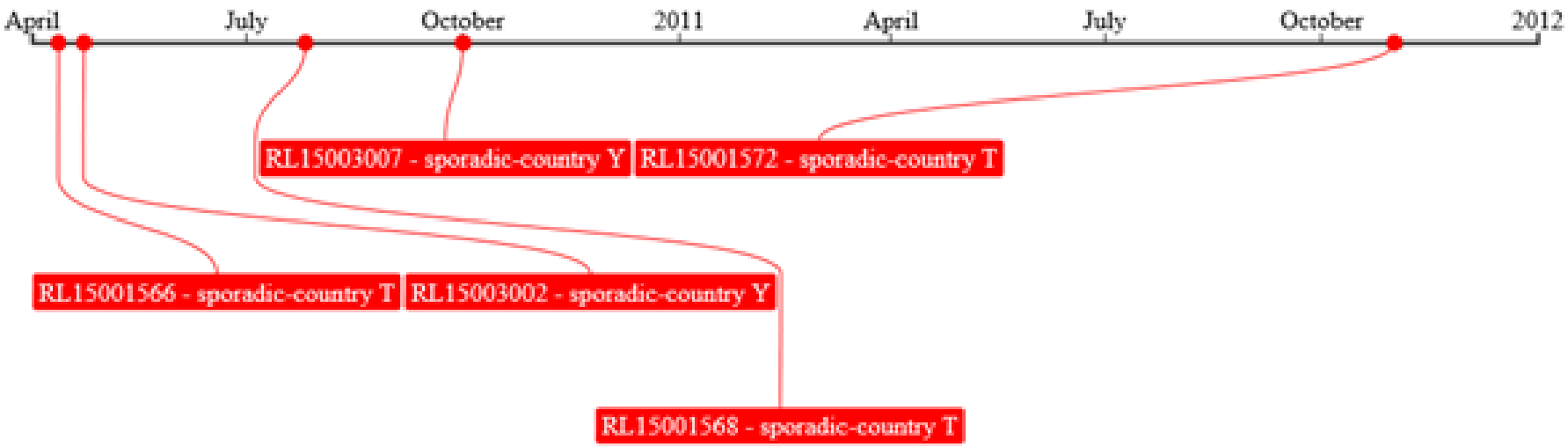


Relation established between the isolates of 21 CCs



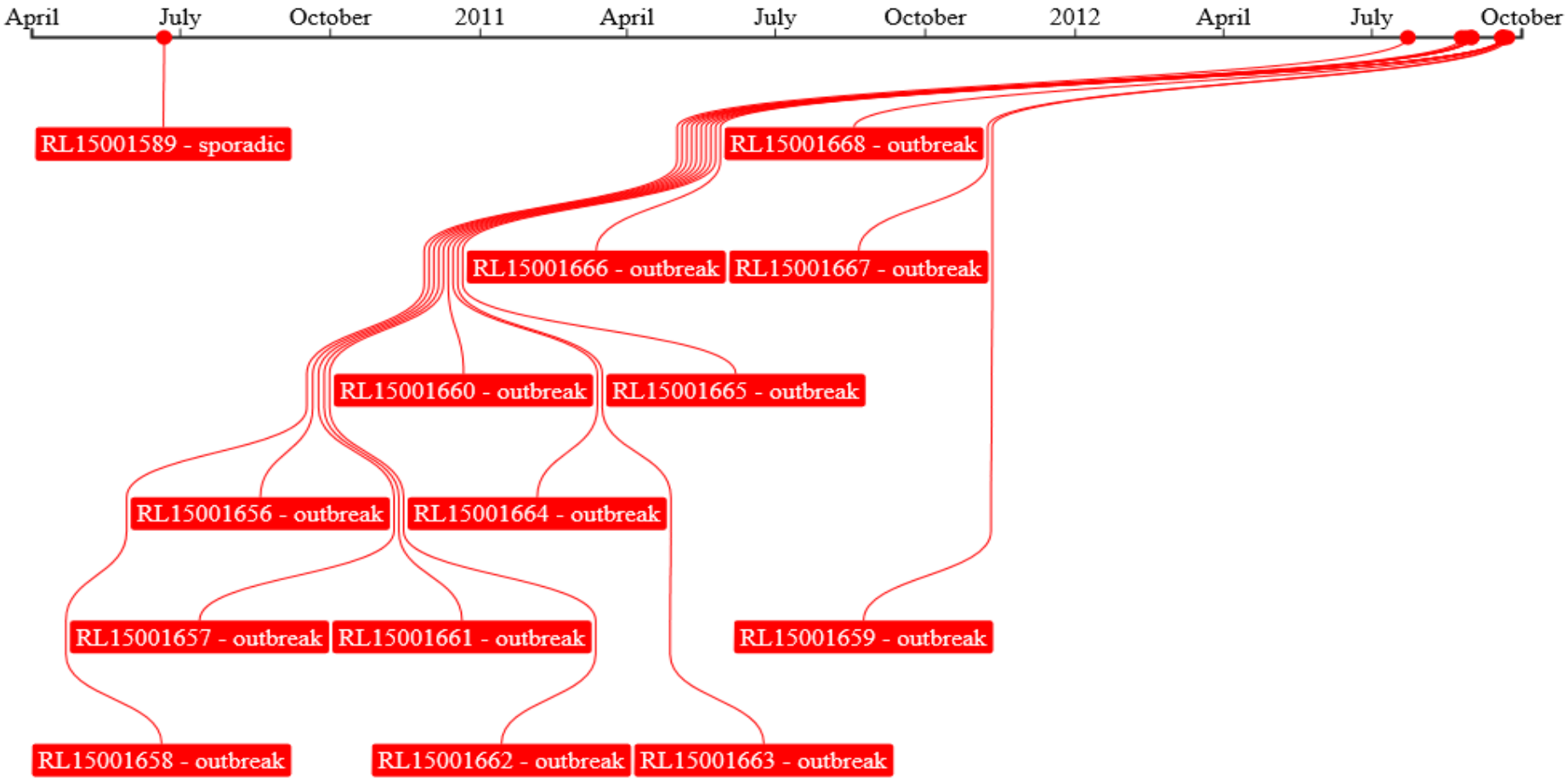
Relation established between the isolates of 21 CCs

- Example 1: sporadic cases together



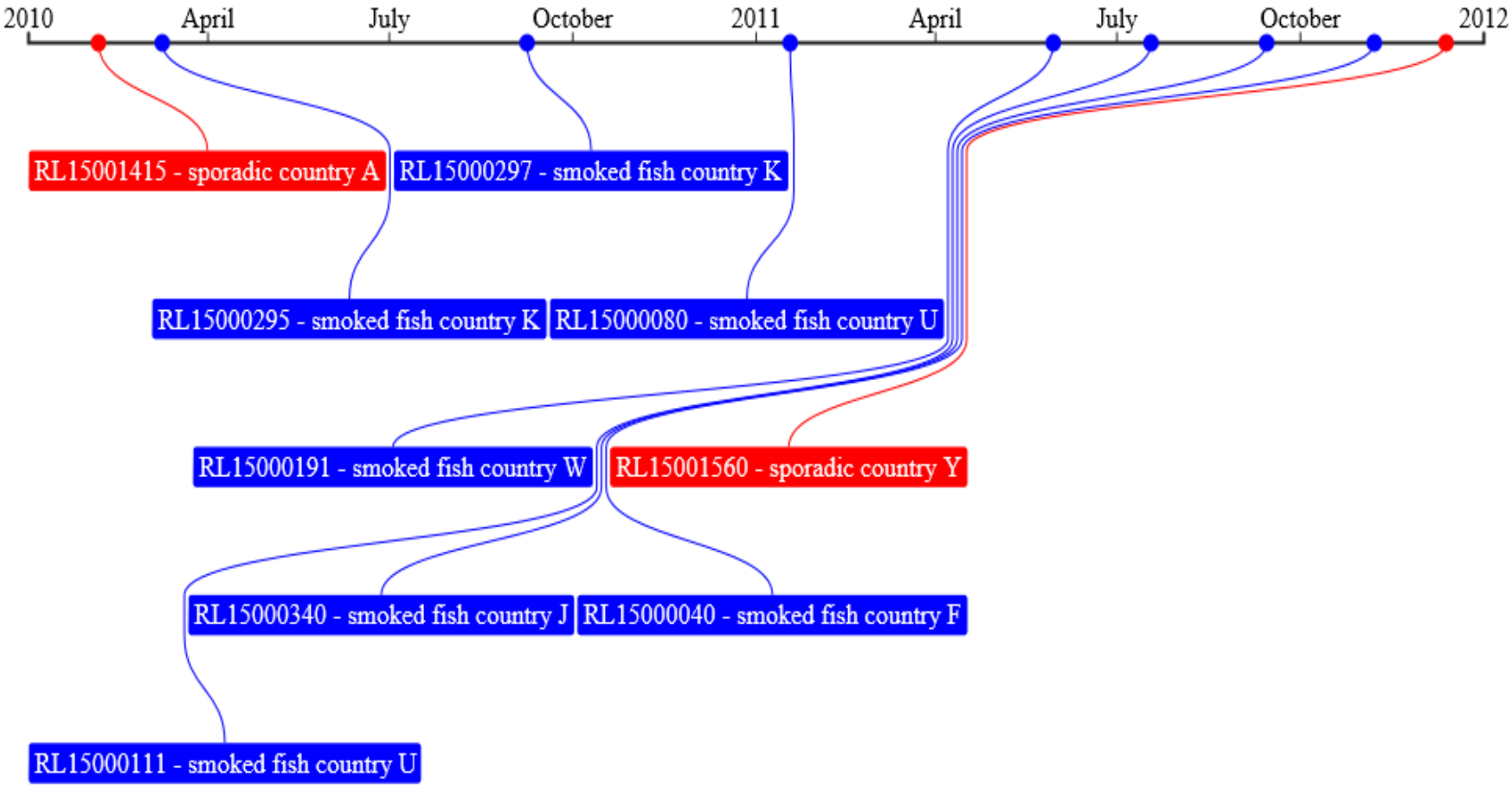
Relation established between the isolates of 21 CCs

- Example 2: sporadic case link to a known outbreak (CC14)



Relation established between the isolates of 21 CCs

- Example 3: sporadic case and RTE food isolates



9. Putative Markers

Antibiotic resistance genes
Virulence factors
Genes implicated in persistence
Markers of host association

Objectives

- Extensive literature exists on stress response and virulence factors of *L. monocytogenes*

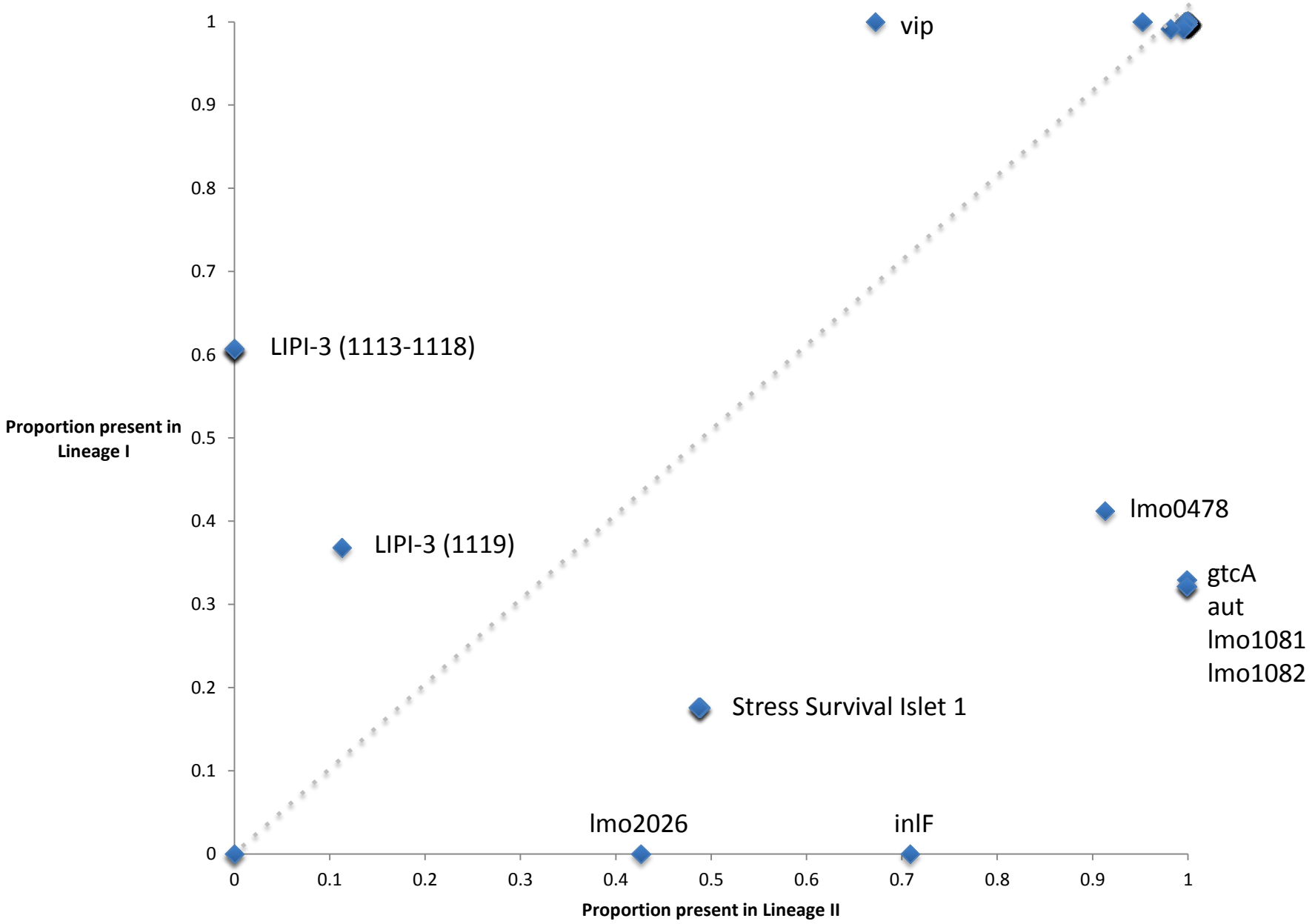
Characterization

- Presence/absence of known markers
  - Virulence, antibio-resistance
  - Persistence in food processing environment
- Identification of *de novo* genetic makers of host association

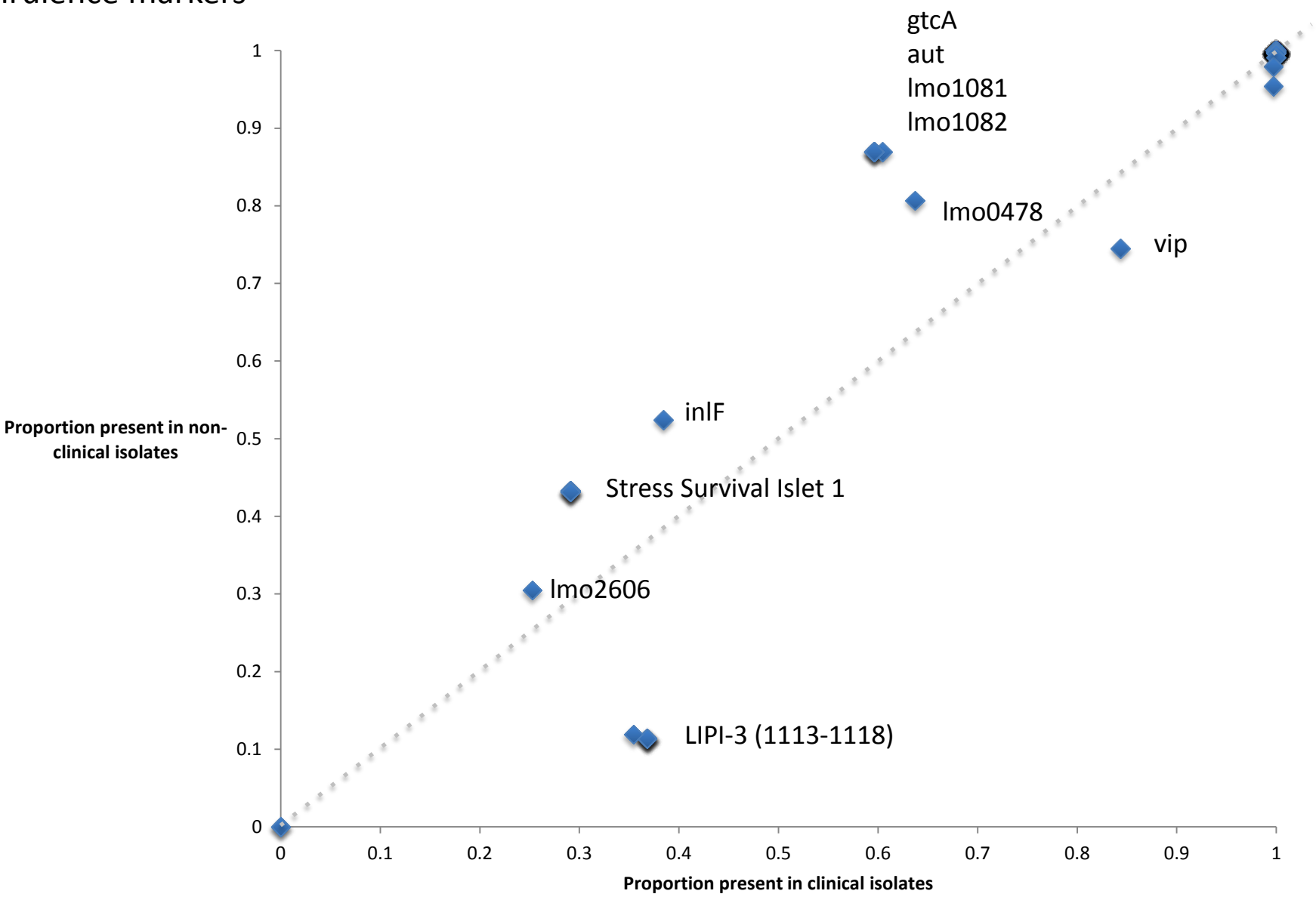
Percent of isolates in the study harbouring the assayed resistance genes

Gene	% Detection
tetM	0.6
tetS	0
bcrA	4.9
bcrB	4.9
bcrC	4.7
emrE	0.3
qacA	0.5
qacC	18.3
Tn6188qac	18.5
penA	0

Virulence markers



Virulence markers



Persistence

Methodology:

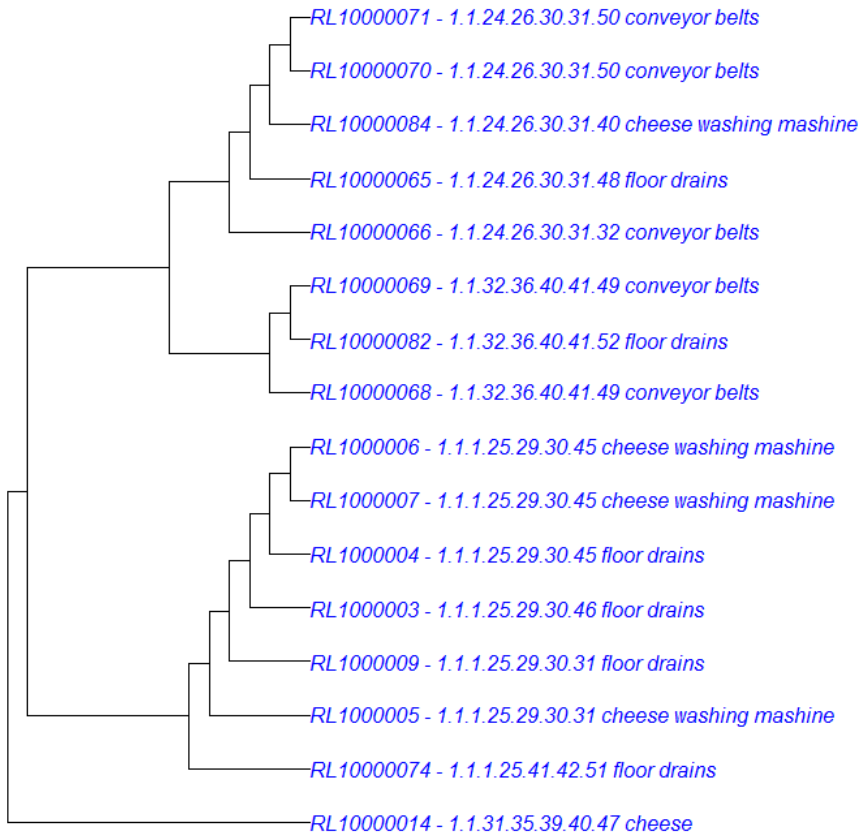
- SNPs to investigate a priori thought persistent isolates (2 cheese processing plants, 1 smoked salmon, 1 meat)
- Presence/absence of putative markers of persistence

Confirm persistence phenotype

- Example cheese factory (CC2)

3 different clusters in processing plant:

Not the one in final product



Presence/absence for markers

- Example pork persistent versus non persistent
- Presence/absence of putative markers is not specific to « persisent phenotype »

Group	RL number	Presence of potential markers for persistence														
		lmo0204	lmo0673	lmo0435	lmo1460	lmo2504	lmo1288	lmo2016	lmo1879	lmo0676	lmo0679	lmo0696	lmo0706	lmo0686	lmo0699	NC_019556.1 FliY
Persistent	RL1500 0543	x	x		x	x	x	x	x	x	x	x	x	x	x	x
	RL1500 0542	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x
	RL1500 0541	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x
	RL1500 0540	x	x		x	x	x	x	x	x	x	x	x	x	x	x
	RL1500 0539	x	x		x	x	x	x	x	x	x	x	x	x	x	x
	RL1500 0538	x	x		x	x	x	x	x	x	x	x	x	x	x	x
	RL1500 0393	x	x		x	x	x	x	x	x	x	x	x	x	x	x
	RL1500 0392	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x
	RL1500 0391	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x
	RL1500 0390	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x
	RL1500 0389	x	x		x	x	x	x	x	x	x	x	x	x	x	x
	RL1500 0388	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x
	RL1500 0387	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x
Non persistent	RL1500 0361	x	x		x	x	x	x	x	x	x	x	x	x	x	x
	RL1500 0362	x	x		x	x	x	x	x	x	x	x	x	x	x	x
	RL1500 0363	x	x		x	x	x	x	x	x	x	x	x	x	x	x
	RL1500 0364	x	x		x	x	x	x	x	x	x	x	x	x	x	x
	RL1500 0365	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x
	RL1500 0366	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x
	RL1500 0367	x	x		x	x	x	x	x	x	x	x	x	x	x	x
	RL1500 0368	x	x		x	x	x	x	x	x	x	x	x	x	x	x
	RL1500 0369	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x
	RL1500 0370	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x
	RL1500 0371	x	x		x	x	x	x	x	x	x	x	x	x	x	x
	RL1500 0372	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x
	RL1500 0373	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x
	RL1500 0374	x	x		x	x	x	x	x	x	x	x	x	x	x	x
	RL1500 0375	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x
	RL1500 0376	x	x		x	x	x	x	x	x	x	x	x	x	x	x
	RL1500 0377	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x
	RL1500 0378	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x

10. Conclusions

1. WGS of *Listeria monocytogenes*

- A unique EU-wide WGS dataset with associated metadata of *L. monocytogenes* human and food isolates that can be use for further studies
- Optimised method for routine high throughput WGS of *L. monocytogenes*
- Investigated the phylogeny of the isolates thus providing the framework for further analysis on genetic diversity and potential epidemiological associations

2.i. Genetic Diversity

- High level of genetic diversity within all sources using both 7 locus MLST and 20 locus rMLST
- Rarefaction demonstrated that only a small portion of *L. monocytogenes* diversity sampled in this project
- Human isolates more diverse than other sources; those from bovine source closest genetically to human isolates

2.ii. Use of WGS to investigate epidemiological relationships

- Source attribution- indicated that bovine reservoir appears to be main source of human isolates, other sources contributed and CIs high
  - New approaches need to be developed for source attribution across the genome
- In conjunction with metadata - numerous consistent genetic linkages between unlinked strains were identified including 124 novel clusters

2.iii. Using WGS to identify putative genetic markers

- AMR markers – WGS for rapid monitoring
- Virulence – known markers in all isolates
- Persistence – use of WGS for monitoring presence of persistent strains but no specific markers for specific persistent phenotype
- Markers of host association – cgSNP and accessory genomes likely to be useful source of host associated polymorphisms

3. WGS for Outbreaks

- Ability to cluster cases previously shown to be epidemiologically linked
- Ability to link previously unassigned cases to known OBs
- Detected sensitive and specific potential links between cases and/or foods that require epidemiological investigations
- Overall potential to detect more OB and more cases
- Evidence to support the use of WGS analysis for *L. monocytogenes surveillance* and for outbreak detection and investigation across Europe

With thanks to EFSA
for funding this work