STATENS SERUM INSTITUT

Listeria whole-genome-sequencing
EFSA Project

Anses
Statens Serum Institut
Public Health England
University of Aberdeen

STATENS SERUM INSTITUT

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.

OC/EFSA/BIOCONTAM/2014/01

Consortium partners – key persons

- Statens Serum Institut:
  - Eva Møller Nielsen, project leader
  - Jonas Larsson
  - Kristoffer Kiil
- Anses
  - Sophie Roussell
  - Laurent Guillier
  - Yannick Blanchard
- Public Health England
  - Kathie Grant
  - Tim Dallman
  - Corinne Amar
- University of Aberdeen
  - Kenn Forbes
  - Norval Strachan

Main objective of EFSA tender

STATENS SERUM INSTITUT

- **Main objective:**

The main objective is 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

STATENS SERUM INSTITUT

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 processsing environment), and humans employing WGS analysis.

- Database with relevant and available metadata will be established
- 1000 L.m strains will be whole-genome sequenced using Illumina platforms (HiSeq, MiSeq, NextSeq).
- The majority will be performed at PHE, high-throughput sequencing facility.
- QC-pipeline, assembly

Typing based on WGS

STATENS SERUM INSTITUT

- Phylogenetic analysis (SNP-trees), overall framework to assess the diversity in sources

- Gene based typing
  - MLST (7-locus; established nomenclature)
  - Extended MLST (25-locus; nomenclature possible)
  - Core-genome MLST (>1700 loci)

These analyses form the basis for the further epidemiological analysis in objectives 2 and 3.

Specific objective 2

STATENS SERUM INSTITUT

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;

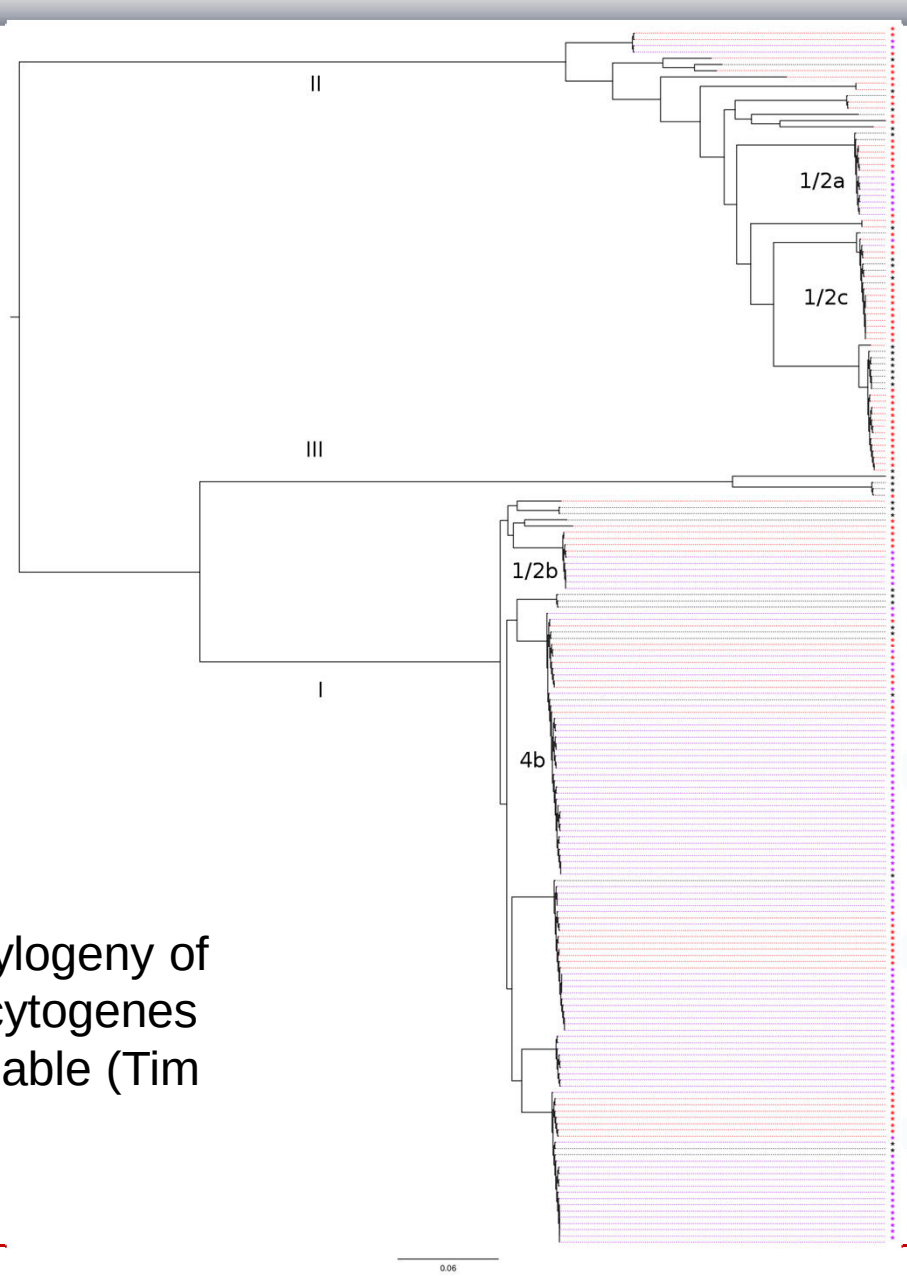
iii. to identify the presence of putative markers conferring the potential to survive/multiply in the food chain and/or cause disease in humans (e.g. virulence and antimicrobial resistance).

Genetic diversity

STATENS SERUM INSTITUT

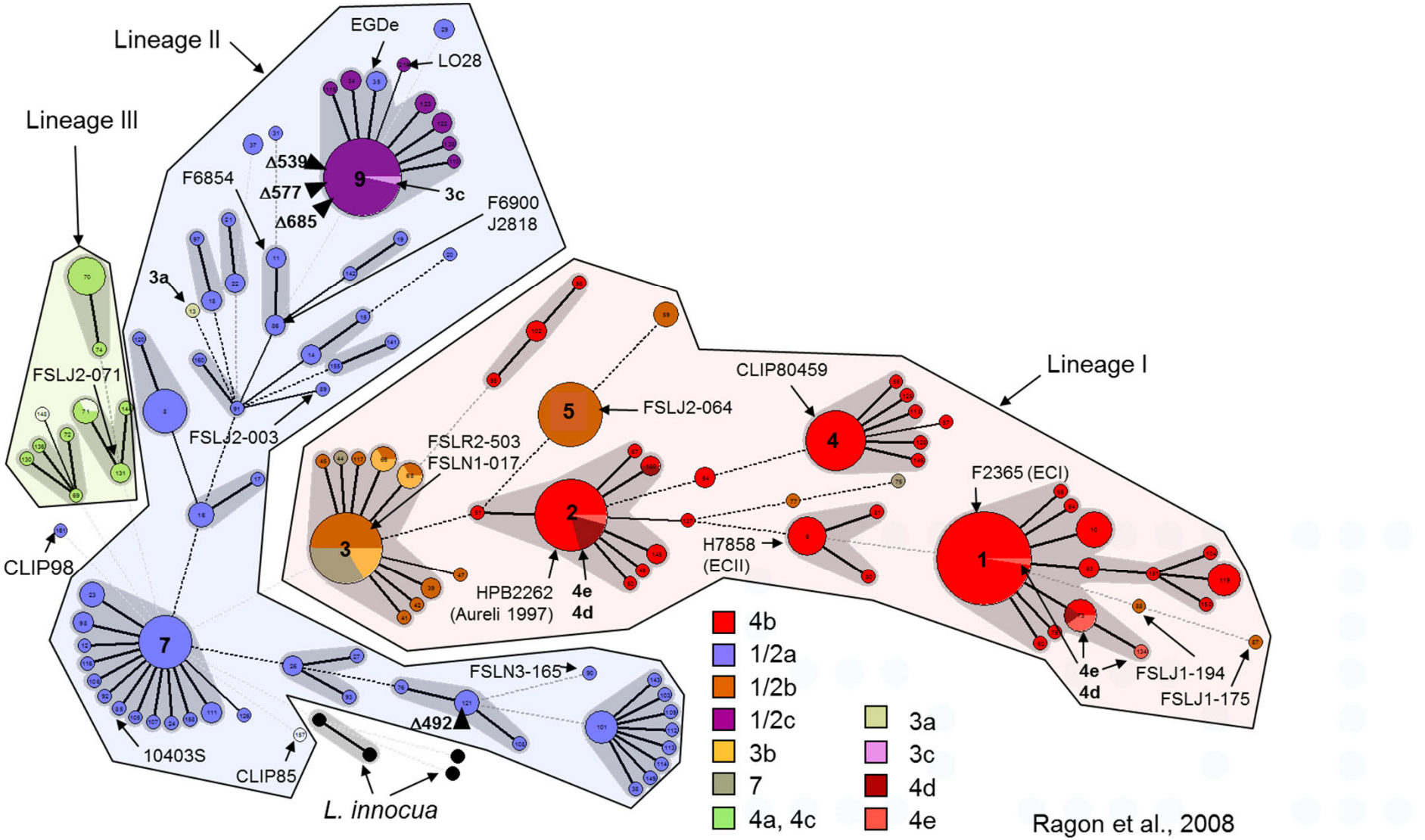
- SNP-trees highlighting the source, geography and time of isolation
- Diversity and frequency of types depending on source/geography/time, frequency of overlapping types between sources and human origins.
- Type distributions in the various sources, the genotype diversity and genetic relatedness: diversity index, geographical ditribution, etc
  - Types defined at the level of MLST
  - WGS types defined at higher discriminatory level when suitable
    - extended MLST (SSI method)
    - cgMLST (method by Institut Pasteur & co)

Listeria phylogeny WGS



Maximum likelihood SNP phylogeny of 195 strains of Listeria monocytogenes from PHE, SSI, publicly available (Tim Dallman; PHE)

Listeria phylogeny (MLST)



Putative markers for potential to survive/multiply in the food chain and/or cause disease in humans

- *Allele based association: A*llele information from wgMLST used for identification of loci or specific alleles that are over- or under-represented in specific sources/reservoirs or in human infections.

- *Genome-wide association:* Identify genes showing evidence of association with a particular phenotype (e.g. clinical listeriosis)

- *Antibiotic resistance genes:* Presence/absence of known AMR genes.

Specific objective 3: Outbreaks

STATENS SERUM INSTITUT

- *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.*
  - *provide an evaluation on the advantages and limitations of WGS for investigating outbreaks of food-borne listeriosis*

- Analysis of WGS data from outbreak strains and similar background isolates
  - explore the diversity between epidemiologically linked isolates
  - phylogenetic relationship of the isolates using SNP and wgMLST data
  - evaluating the thresholds for cluster definitions
  - demonstrating how WGS methods work for well-described outbreaks.

Selection of isolates

STATENS SERUM INSTITUT

- Human clinical isolates, sporadic 250
- Outbreaks (5-8) 100
- BLS isolates 349
  - Fish 293
  - Meat 49
  - Cheese 7
- Supplementary food isolates 236
- Selected processing plants 100
  (raw, environment, food)

- Selection of strains related to food sold in many EU-countries
  - Meat and meat products (France)
  - Cheeses and cheese plants over several years (Italy, UK)
  - Fish processing industry – fish and processing environment (Scotland)

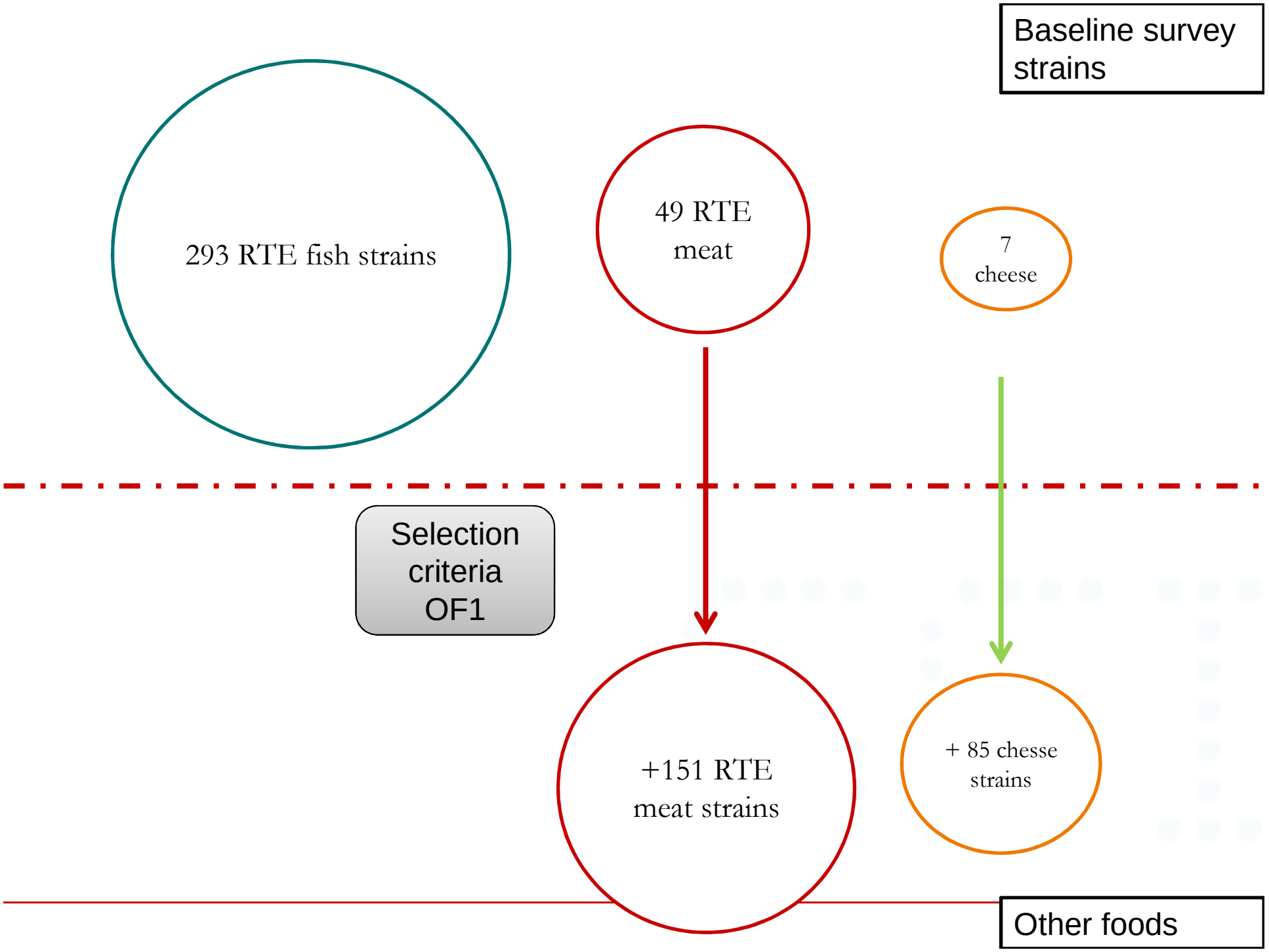
Food and food chain isolates

STATENS SERUM INSTITUT

- Baseline survey: 707 *L. monocytogenes* isolated from
  - 72 RTE meat products,
  - 619 fish products (310 at sampling and 309 at end of shelf life stages)
  - 16 cheeses
- Other isolates, provided by consortium or collaborators.
  - Supplement to BLS
  - Other food sources
  - Food processing chains

Selection of strains related to food sold in many EU-countries:

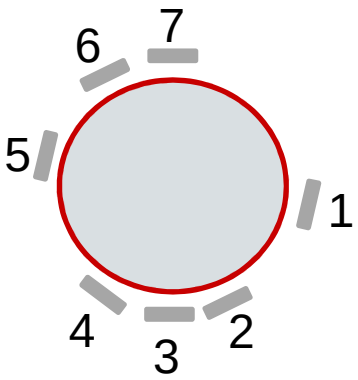
- Fish processing industry – fish and processing environment (Scotland)
- Meat and meat products (France, 2010-2014)
- Cheeses and cheese plants over several years (Sardinia, Italia)


Baseline survey strains
293 RTE fish strains
49 RTE meat
7 cheese
Selection criteria OF1
+151 RTE meat strains
+ 85 chesse strains
Other foods

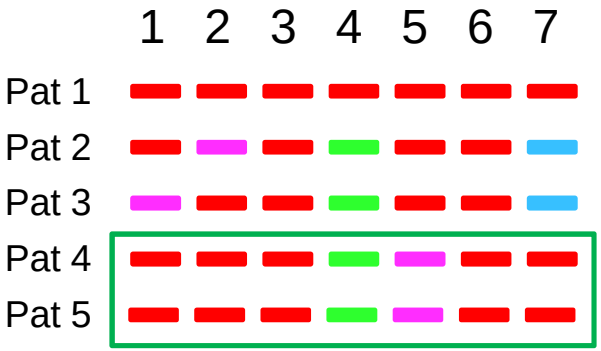
Whole-genome-sequencing (WGS), SNPs, MLST

STATENS SERUM INSTITUT

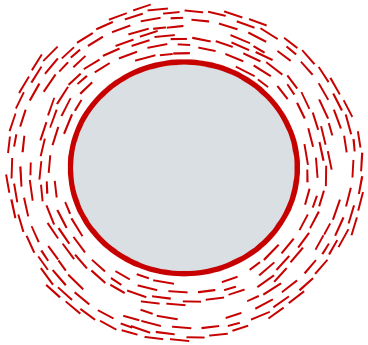
MLST – approx 3500 basepairs



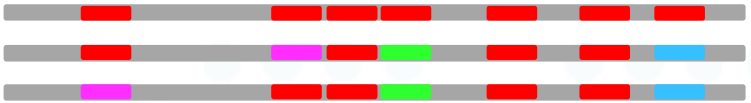
7 selected regions
comparison of isolates



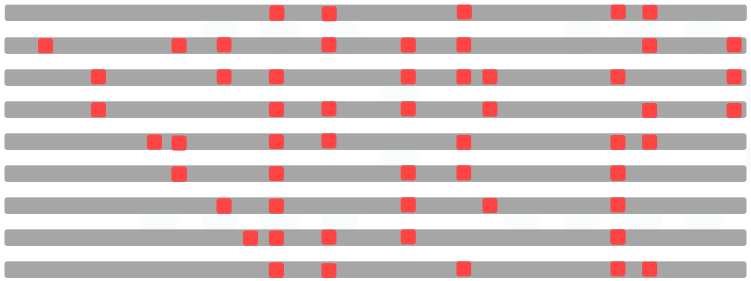
WGS – MLST + SNP analysis



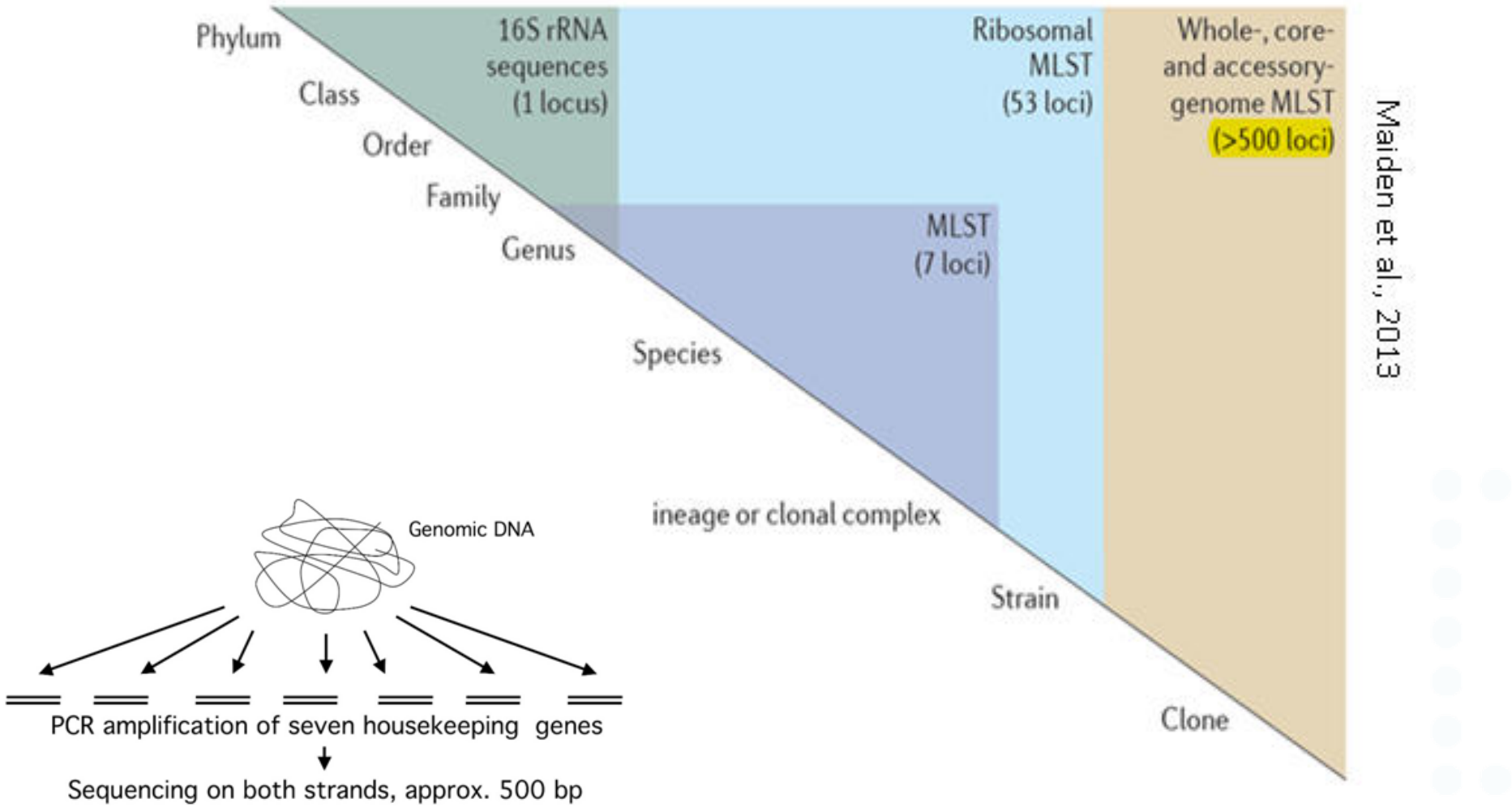
MLST: 3500 bp = 0.13% of the genome



SNP: 2.8 mio bp = 98% of the genome



MLST – cgMLST – wgMLST



Extended MLST (xtMLST)

STATENS SERUM INSTITUT

- Relatively few additional loci increase discriminatory power dramatically

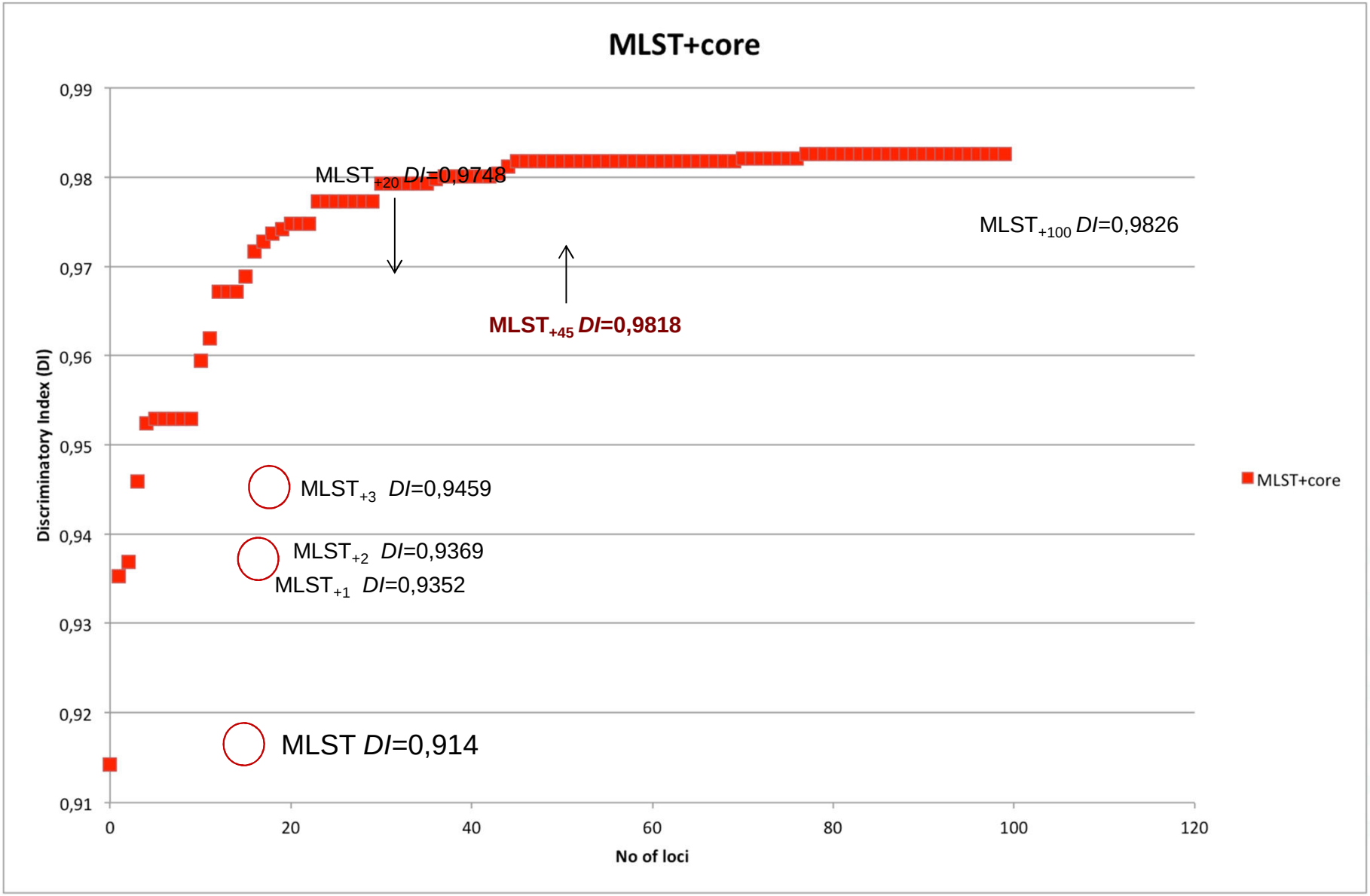
- Simpel nomenclature possible
  - String of e.g. 25 numbers

- 25-locus MLST developed on retrospetive data
  - Selected outbreak and sporadic cases 2002-2012

- Evaluated on independent surveillance data
  - 2013-2014 human infections

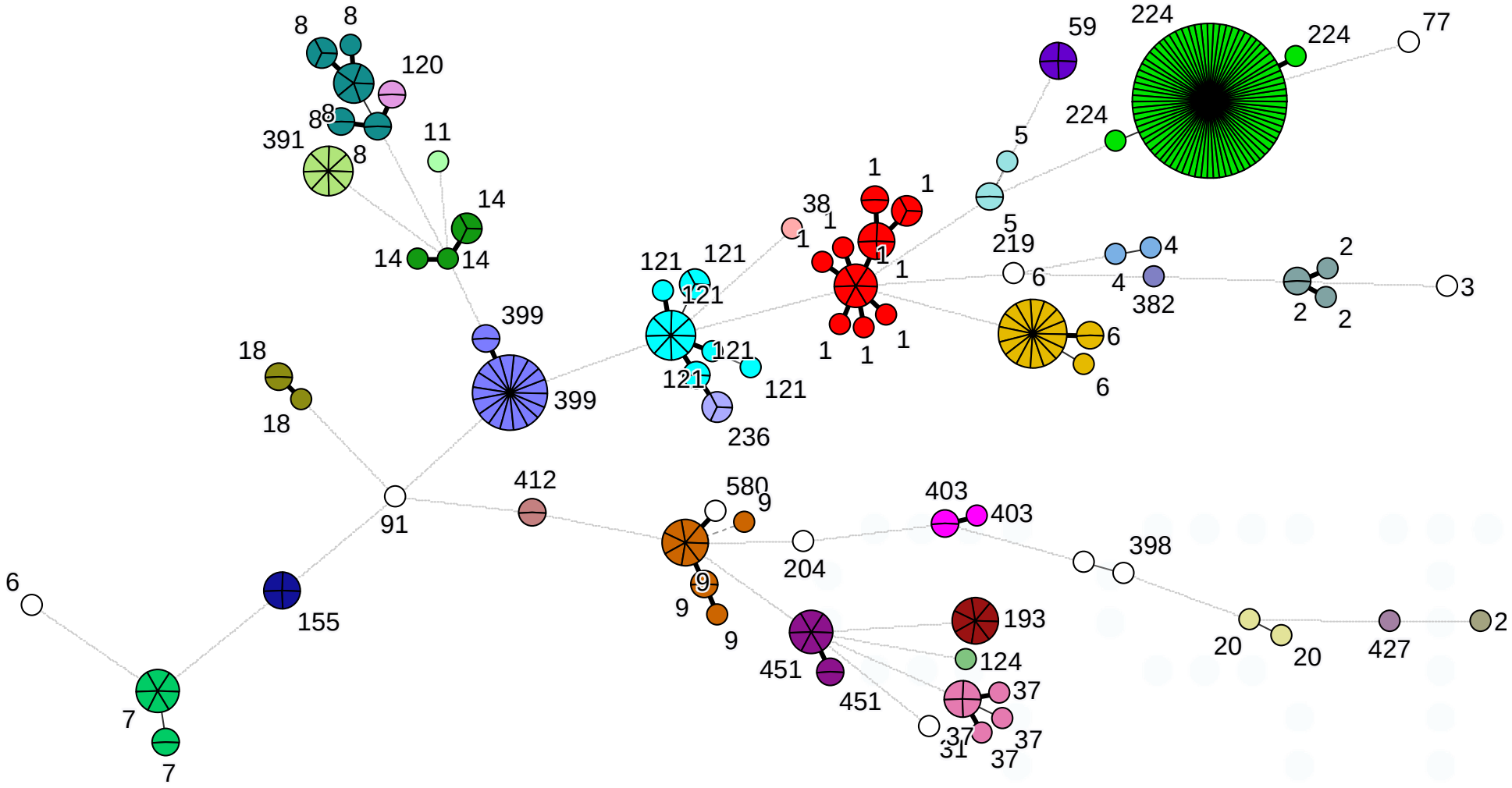
MLST+100 CORE GENES

STATENS SERUM INSTITUT



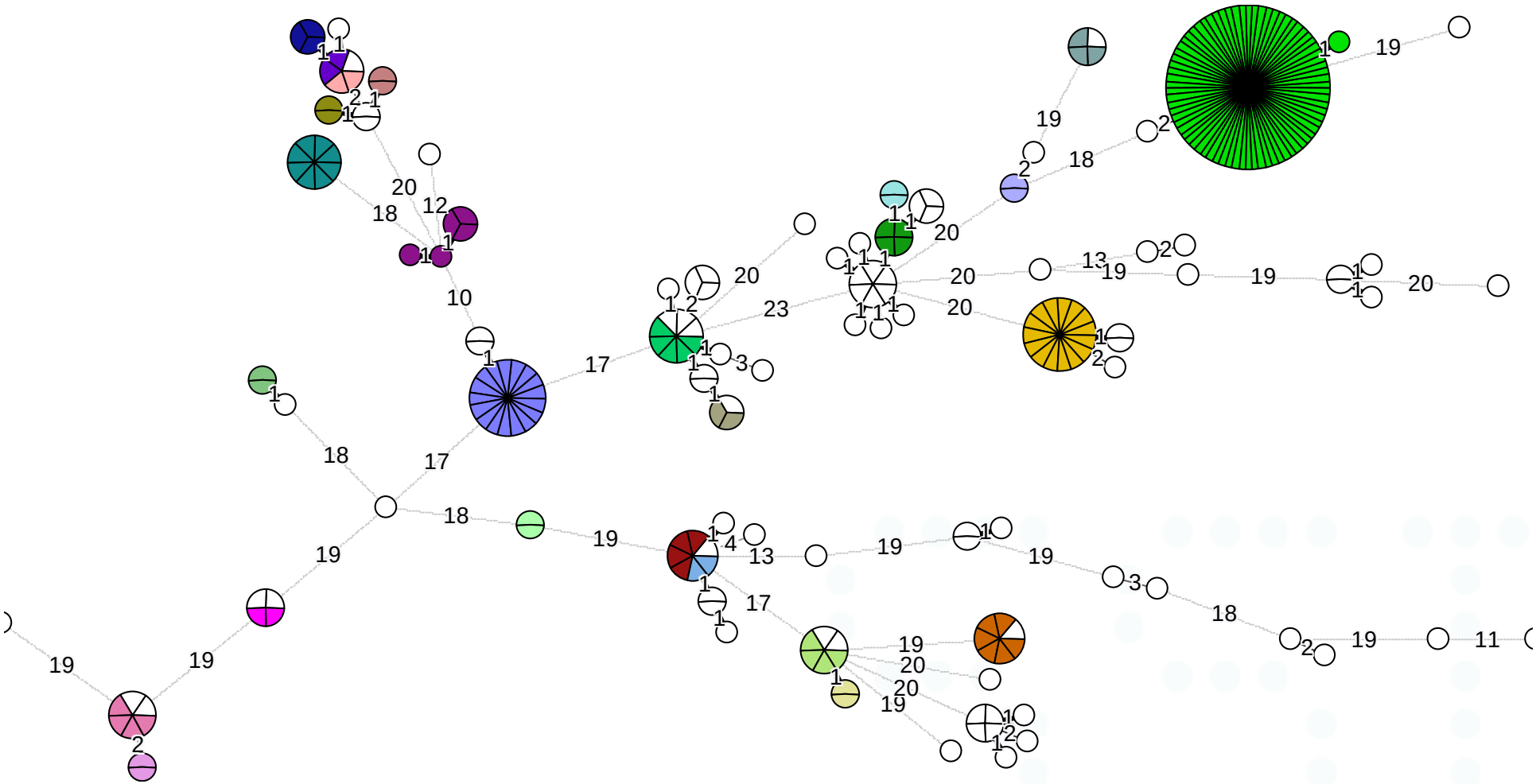
xtMST tree: coloured by 7-locus MLST (n=264)

STATENS SERUM INSTITUT



xtMST tree: coloured by SNP cluster

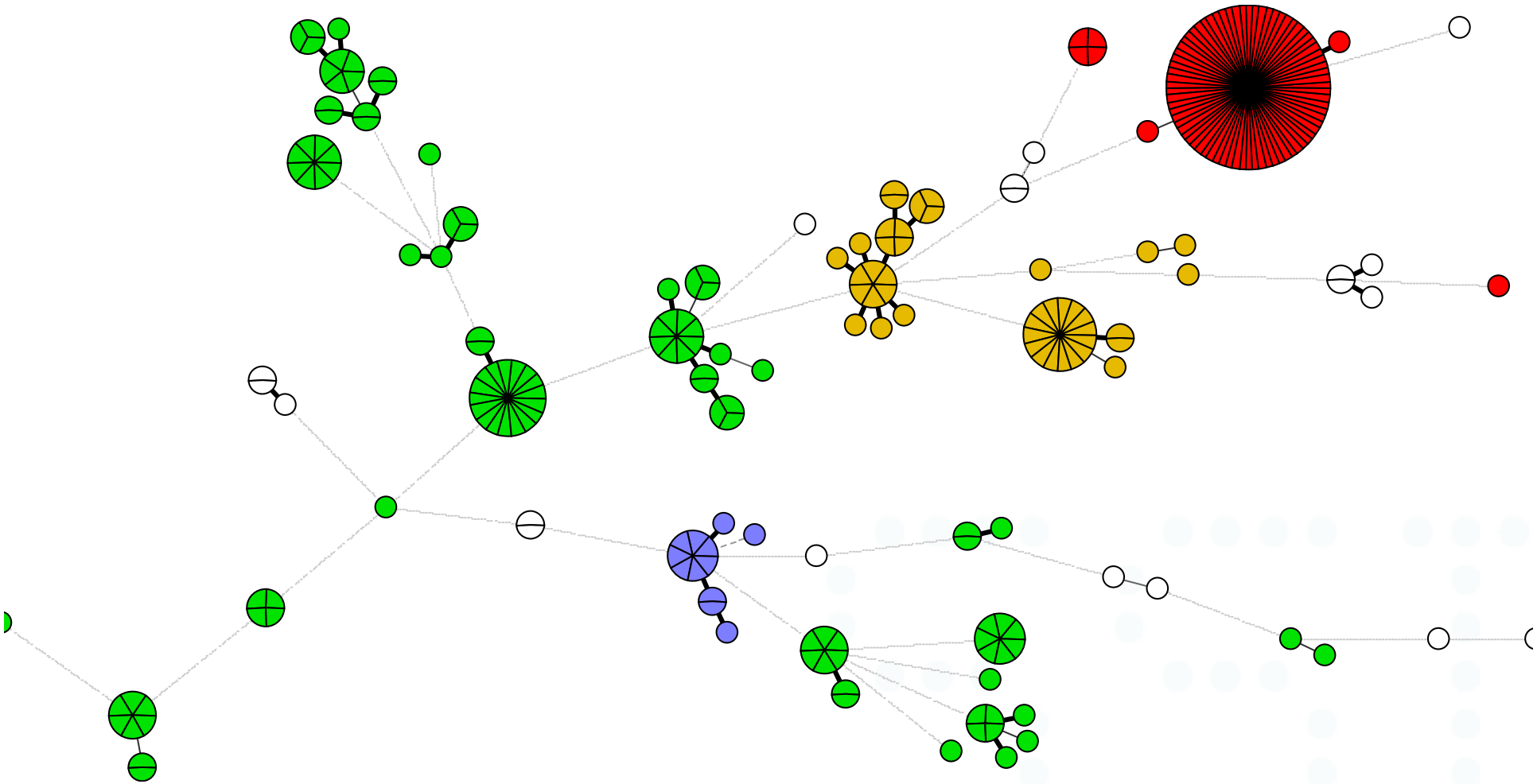
STATENS SERUM INSTITUT



Numbers: allele differences

xtMST tree: coloured by molecular serotype

STATENS SERUM INSTITUT



Two-year project

STATENS SERUM INSTITUT

- Contract signed October 7, 2014
- Selection of isolates
  - Database, including information on isolates
- Collection of isolates, agreements with isolate providers
- Performing WGS from March 2015
  - First data analysis April-May 2015
- WGS, final set including QC and re-runs
  - Final data set for further analysis ready by July 2015
- Data analysis, objectives 2 and 3:
  - August 2015 to spring 2016
- Report to EFSA by September 7, 2016.

Data sharing and collaborations

STATENS SERUM INSTITUT

- Isolate providers:
  - Sequence data in return
  - Set their data into context
  - Publications

- EFSA owns data and data analysis