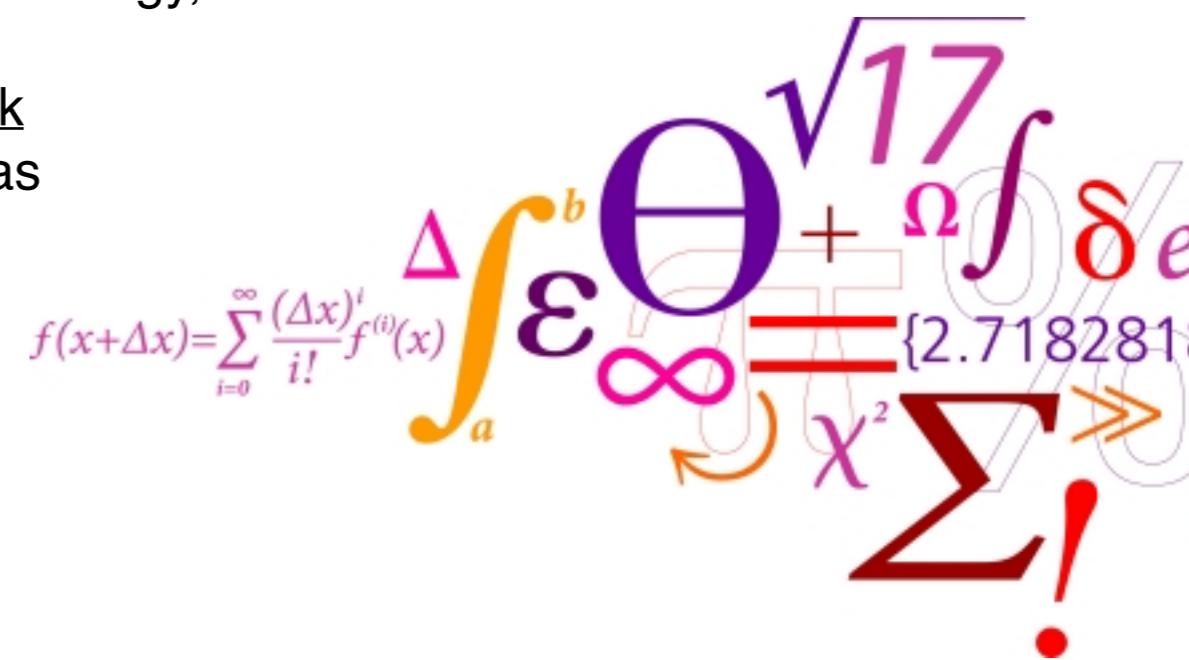


Pan-genome analysis

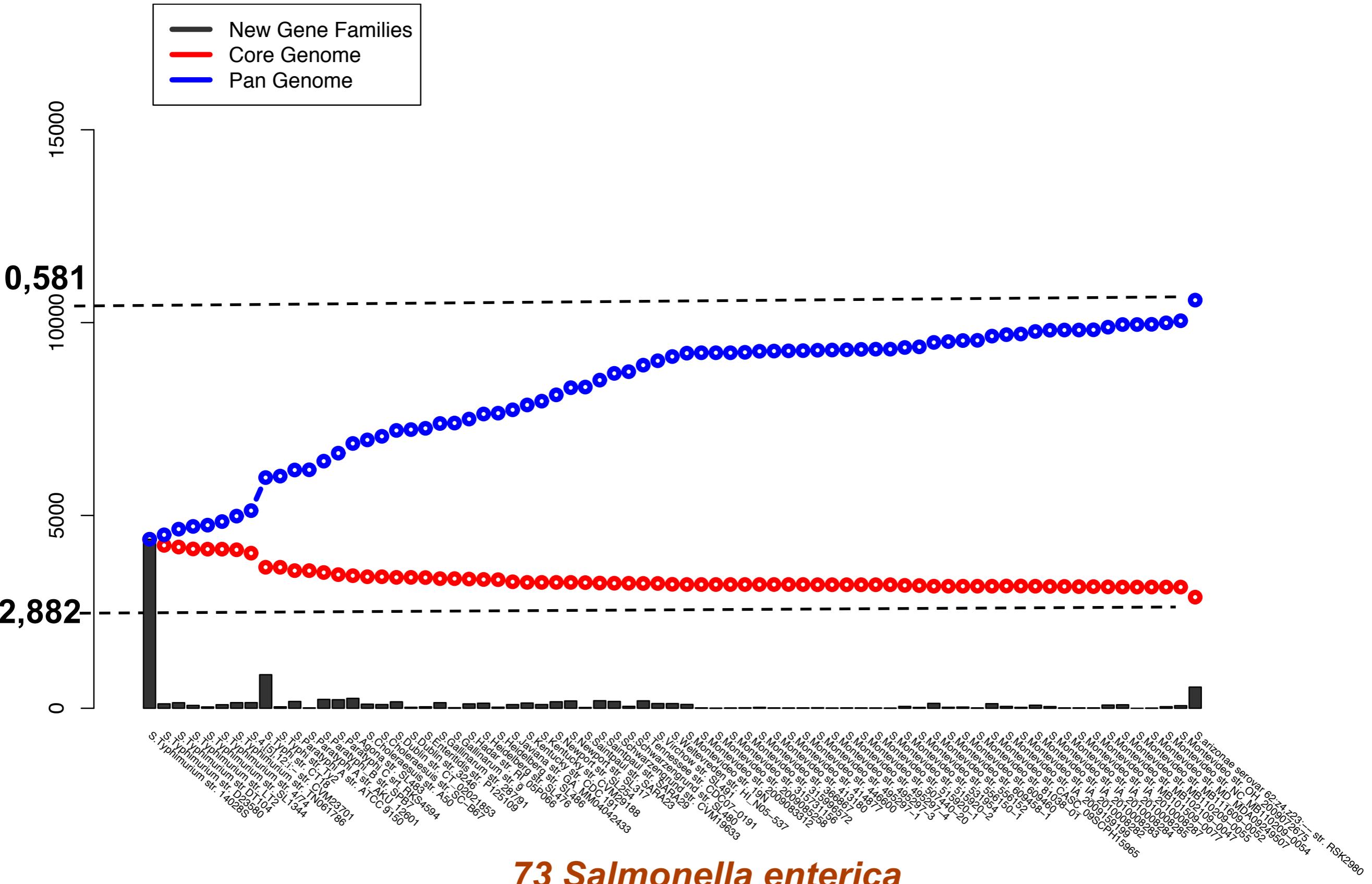
Pimplapas Leekitcharoenphon (Shinny)
Research Group of Genomic Epidemiology, DTU-Food

pile@food.dtu.dk
 @ShinnyPimplapas

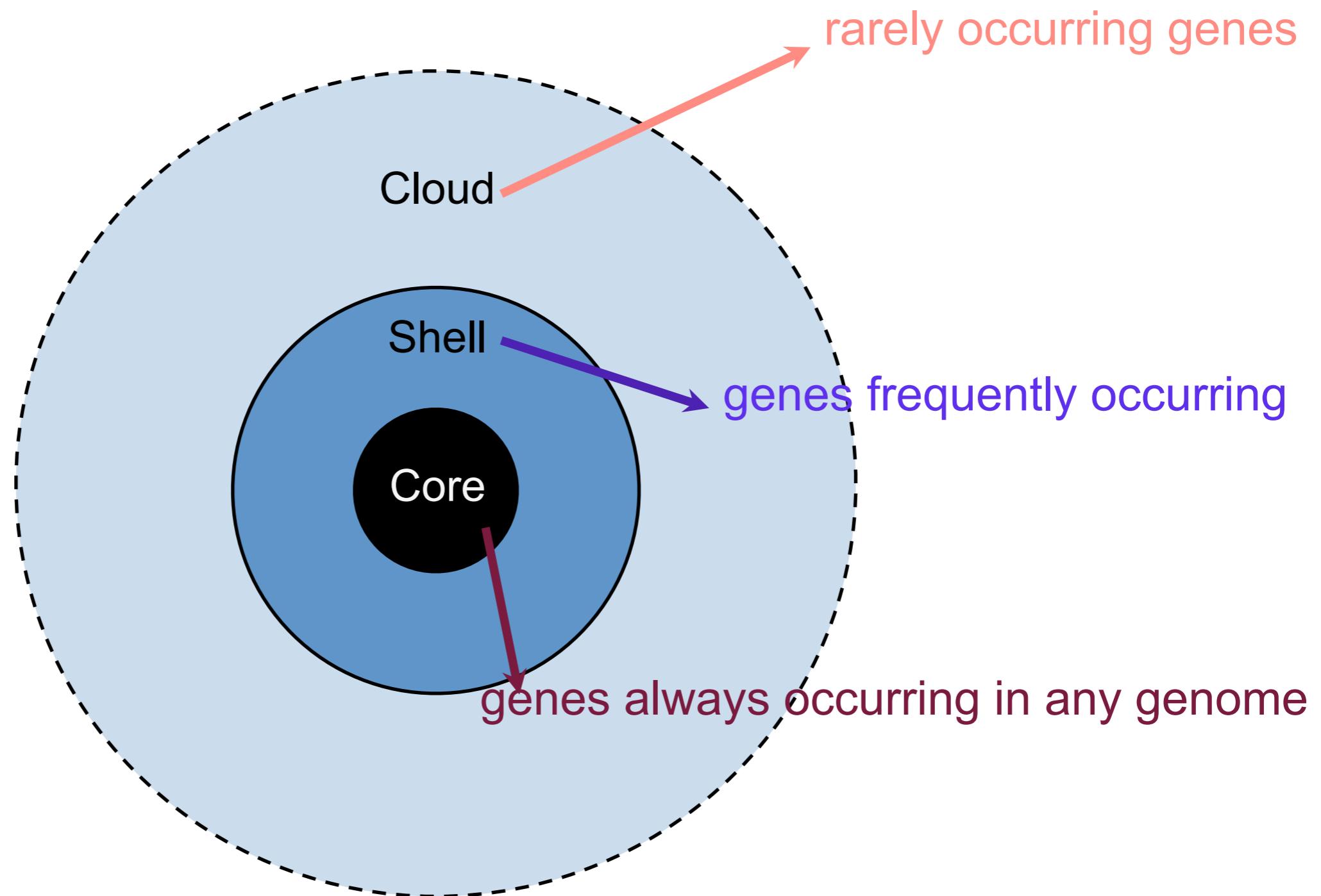
$$f(x+\Delta x) = \sum_{i=0}^{\infty} \frac{(\Delta x)^i}{i!} f^{(i)}(x)$$


Infectious Disease Bioinformatics course, March 2022

Pan - Core Genome



Bacterial pan-genome



Roary: the pan genome pipeline

<https://sanger-pathogens.github.io/Roary/>



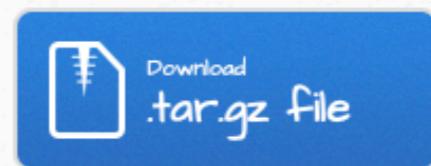
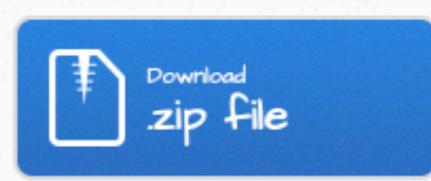
/ Abstract

By Andrew Page based on version 3.11.2 (22-Jan-2018)

Roary is a high speed stand alone pan genome pipeline, which takes annotated assemblies in GFF3 format (produced by Prokka (Seemann, 2014)) and calculates the pan genome. Using a standard desktop PC, it can analyse datasets with thousands of samples, something which is computationally infeasible with existing methods, without compromising the quality of the results. 128 samples can be analysed in under 1 hour using 1 GB of RAM and a single processor. To perform this analysis using existing methods would take weeks and hundreds of GB of RAM. Roary is not intended for metagenomics or for comparing extremely diverse sets of genomes.

/ Citation and further details of the method

Andrew J. Page, Carla A. Cummins, Martin Hunt, Vanessa K. Wong, Sandra Reuter, Matthew T. G. Holden, Maria Fookes, Daniel Falush, Jacqueline A. Keane, Julian Parkhill, "Roary: Rapid large-scale prokaryote pan genome analysis", Bioinformatics, 2015;31(22):3691-3693
[doi:10.1093/bioinformatics/btv421](https://doi.org/10.1093/bioinformatics/btv421)



Roary is maintained by [sanger-pathogens](#).

This page was generated by [GitHub Pages](#) using the Architect theme by [Jason Long](#).

Roary installation

<https://github.com/sanger-pathogens/Roary>

- [Introduction](#)
- [Installation](#)
 - [Required dependencies](#)
 - [Optional dependencies](#)
 - [Ubuntu/Debian](#)
 - [Debian Testing](#)
 - [Ubuntu 14.04/16.04](#)
 - [Ubuntu 12.04](#)
 - [Bioconda - OSX/Linux](#)
 - [Galaxy](#)
 - [GNU Guix](#)
 - [Virtual Machine - OSX/Linux/Windows](#)
 - [Docker - OSX/Linux/Windows/Cloud](#)
 - [Installing from source \(advanced Linux users only\)](#)
 - [Ancient systems and versions of perl](#)
 - [Running the tests](#)
 - [Versions of software we test against](#)
- [Usage](#)
- [License](#)
- [Feedback/Issues](#)
- [Citation](#)
- [Further Information](#)

Roary on Computerome

```
# Load all required modules
```

```
module load tools
module load parallel/20170822
module load prank/140603
module load mcl/14-137
module load cd-hit/4.6.1
module load mafft/7.245
module load fasttree/2.1.9
```

```
# Check all available option in roary
```

```
roary -h
```

Path to FASTA files /home/projects/course_23262/course/week09/pangenome/data

```
GCA_000008285.fna
GCA_000021185.fna
GCA_000026705.fna
GCA_000168635.fna
GCA_000168815.fna
GCA_000196035.fna
```

/// roary

Usage: roary [options] *.gff

Options:

-p INT	number of threads [1]
-o STR	clusters output filename [clustered_proteins]
-f STR	output directory [.]
-e	create a multiFASTA alignment of core genes using PRANK
-n	fast core gene alignment with MAFFT, use with -e
-i	minimum percentage identity for blastp [95]
-cd FLOAT	percentage of isolates a gene must be in to be core [99]
-qc	generate QC report with Kraken
-k STR	path to Kraken database for QC, use with -qc
-a	check dependancies and print versions
-b STR	blastp executable [blastp]
-c STR	mcl executable [mcl]
-d STR	mcxdeblast executable [mcxdeblast]
-g INT	maximum number of clusters [50000]
-m STR	makeblastdb executable [makeblastdb]
-r	create R plots, requires R and ggplot2
-s	dont split paralogs
-t INT	translation table [11]
-ap	allow paralogs in core alignment
-z	dont delete intermediate files
-v	verbose output to STDOUT
-w	print version and exit
-y	add gene inference information to spreadsheet, doesnt work with -e
-iv STR	Change the MCL inflation value [1.5]
-h	this help message

Prokka for gene annotation

```
# Load required module
```

```
module load prokka/1.12
```

```
# Example of running prokka
```

```
prokka --kingdom Bacteria --outdir prokka_GCA_000008285 --genus  
Listeria --locustag GCA_000008285 GCA_000008285.fna
```

Try to run prokka on Computerome. Example of prokka is on DTU Learn (prokka.sh)

Path to gff files /home/projects/course_23262/course/week09/pangenome/data/all_gff/*.gff

Running roary

#running roary

```
roary -f ./output -e -n -v -z -o clustered_proteins /home/projects/course_23262/course/week08/pangenome/data/all_gff/*.gff
```

Try to run roary on Computerome. Example of roary is on DTU Learn (roary.sh)

* Run roary as a group because it takes time.

Roary output

```
/home/projects/course_23262/course/week09/pangenome/roary_output/output
```

Output files

// summary_statistics.txt

Number of genes in the core and accessory. A text file with an overview of the genes and how frequently they occur in the input isolates. If the number of core genes is 0 it can indicate you have some contamination. Likewise if the total number of genes is very high.

Core genes (99% \leq strains \leq 100%):	1830
Soft core genes (95% \leq strains < 99%):	157
Shell genes (15% \leq strains < 95%):	1471
Cloud genes (0% \leq strains < 15%):	2987
Total genes:	6445

// gene_presence_absence.csv

The gene presence and absence spreadsheet lists each gene and which samples it is present in. The view below shows how it looks in Excel.

Gene	Non-unique-Gene name	Annotation	No. Initiating	No. Responses	Avg responses per Initiating	Genomic Fragment	Order within Fragment	Accessory Fragment	Accessory Order with Accessory QC	DRR&ATL	DRR&ATL	N	N	DRR&ATL
grna_00		Hypothetical protein	34	34	1	3400				DRR&ATL	DRR&ATL	2262	2262	DRR&ATL
grna_01		Uridylate/diphosphate hydrolase	34	34	1	3400				DRR&ATL	DRR&ATL	2263	2263	DRR&ATL
grna_02		Biotin kinase 3	34	34	1	3400				DRR&ATL	DRR&ATL	2264	2264	DRR&ATL
grna_03		26S/20S ribosomal 2'-O-methyltransferase	34	34	1	3400				DRR&ATL	DRR&ATL	2265	2265	DRR&ATL
grna_04		Adenosine triphosphate synthase	34	34	1	3400				DRR&ATL	DRR&ATL	2266	2266	DRR&ATL
grna_05		"Hypothetical protein" rpsC	34	34	1	3400				DRR&ATL	DRR&ATL	2267	2267	DRR&ATL
grna_06		Wolbachia protein	34	34	1	3400				DRR&ATL	DRR&ATL	2268	2268	DRR&ATL
grna_07		Alpha/beta hydrolase protein rplP	34	34	1	3400				DRR&ATL	DRR&ATL	2269	2269	DRR&ATL
grna_08		Analysis of unknown function (NUF100)	34	34	1	3400				DRR&ATL	DRR&ATL	2270	2270	DRR&ATL
grna_09		Undecaprenyl pyrophosphate lyase	34	34	1	3400				DRR&ATL	DRR&ATL	2271	2271	DRR&ATL
grna_10		GDP binding protein HK	34	34	1	3400				DRR&ATL	DRR&ATL	2272	2272	DRR&ATL
grna_11		"Hypothetical protein"	34	34	1	3400				DRR&ATL	DRR&ATL	2273	2273	DRR&ATL
grna_12		2'-O-methyltransferase -4R4M-(pro-1')-methyltransferase TrnPO	34	34	1	3400				DRR&ATL	DRR&ATL	2274	2274	DRR&ATL
grna_13		Signal peptide 1.5	34	34	1	3400				DRR&ATL	DRR&ATL	2275	2275	DRR&ATL
grna_14		Translative operon transcriptional repressor	34	34	1	3400				DRR&ATL	DRR&ATL	2276	2276	DRR&ATL
grna_15		Hydroxide lyase/hydrolase	34	34	1	3400				DRR&ATL	DRR&ATL	2277	2277	DRR&ATL
grna_16		Arginase dehydrogenase	34	34	1	3400				DRR&ATL	DRR&ATL	2278	2278	DRR&ATL
grna_17		Arginase (NAD+-dependent) multi enzyme 3	34	34	1	3400				DRR&ATL	DRR&ATL	2279	2279	DRR&ATL
grna_18		Arginase	34	34	1	3400				DRR&ATL	DRR&ATL	2280	2280	DRR&ATL
grna_19		Probable diglycine oxidase homolog	34	34	1	3400				DRR&ATL	DRR&ATL	2281	2281	DRR&ATL
grna_20		Hypothetical protein	34	34	1	3400				DRR&ATL	DRR&ATL	2282	2282	DRR&ATL
grna_21		Hypothetical protein	34	34	1	3400				DRR&ATL	DRR&ATL	2283	2283	DRR&ATL
grna_22		Ornithine decarboxylase	34	34	1	3400				DRR&ATL	DRR&ATL	2284	2284	DRR&ATL
grna_23		"Hypothetical protein"	34	34	1	3400				DRR&ATL	DRR&ATL	2285	2285	DRR&ATL
grna_24		Conformational switcher serine large chain	34	34	1	3400				DRR&ATL	DRR&ATL	2286	2286	DRR&ATL
grna_25		Oxidoreductase/transferase	34	34	1	3400				DRR&ATL	DRR&ATL	2287	2287	DRR&ATL
grna_26		short chain dehydrogenase	34	34	1	3400				DRR&ATL	DRR&ATL	2288	2288	DRR&ATL
grna_27		Aromatic amino acid	34	34	1	3400				DRR&ATL	DRR&ATL	2289	2289	DRR&ATL
grna_28		Gamma-glutamyl hydrolase	34	34	1	3400				DRR&ATL	DRR&ATL	2290	2290	DRR&ATL
grna_29		Gamma-glutamyl hydrolase	34	34	1	3400				DRR&ATL	DRR&ATL	2291	2291	DRR&ATL
grna_30		Tetra- and pentapeptide hydrolase	34	34	1	3400				DRR&ATL	DRR&ATL	2292	2292	DRR&ATL
grna_31		U11/U12 small nucleolar RNA	34	34	1	3400				DRR&ATL	DRR&ATL	2293	2293	DRR&ATL
grna_32		U11/U12 small nucleolar RNA	34	34	1	3400				DRR&ATL	DRR&ATL	2294	2294	DRR&ATL
grna_33		U11/U12 small nucleolar RNA	34	34	1	3400				DRR&ATL	DRR&ATL	2295	2295	DRR&ATL
grna_34		U11/U12 small nucleolar RNA	34	34	1	3400				DRR&ATL	DRR&ATL	2296	2296	DRR&ATL
grna_35		U11/U12 small nucleolar RNA	34	34	1	3400				DRR&ATL	DRR&ATL	2297	2297	DRR&ATL
grna_36		U11/U12 small nucleolar RNA	34	34	1	3400				DRR&ATL	DRR&ATL	2298	2298	DRR&ATL
grna_37		U11/U12 small nucleolar RNA	34	34	1	3400				DRR&ATL	DRR&ATL	2299	2299	DRR&ATL
grna_38		U11/U12 small nucleolar RNA	34	34	1	3400				DRR&ATL	DRR&ATL	2300	2300	DRR&ATL
grna_39		U11/U12 small nucleolar RNA	34	34	1	3400				DRR&ATL	DRR&ATL	2301	2301	DRR&ATL
grna_40		U11/U12 small nucleolar RNA	34	34	1	3400				DRR&ATL	DRR&ATL	2302	2302	DRR&ATL
grna_41		U11/U12 small nucleolar RNA	34	34	1	3400				DRR&ATL	DRR&ATL	2303	2303	DRR&ATL
grna_42		U11/U12 small nucleolar RNA	34	34	1	3400				DRR&ATL	DRR&ATL	2304	2304	DRR&ATL
grna_43		U11/U12 small nucleolar RNA	34	34	1	3400				DRR&ATL	DRR&ATL	2305	2305	DRR&ATL
grna_44		U11/U12 small nucleolar RNA	34	34	1	3400				DRR&ATL	DRR&ATL	2306	2306	DRR&ATL
grna_45		U11/U12 small nucleolar RNA	34	34	1	3400				DRR&ATL	DRR&ATL	2307	2307	DRR&ATL
grna_46		U11/U12 small nucleolar RNA	34	34	1	3400				DRR&ATL	DRR&ATL	2308	2308	DRR&ATL
grna_47		U11/U12 small nucleolar RNA	34	34	1	3400				DRR&ATL	DRR&ATL	2309	2309	DRR&ATL
grna_48		U11/U12 small nucleolar RNA	34	34	1	3400				DRR&ATL	DRR&ATL	2310	2310	DRR&ATL
grna_49		U11/U12 small nucleolar RNA	34	34	1	3400				DRR&ATL	DRR&ATL	2311	2311	DRR&ATL
grna_50		U11/U12 small nucleolar RNA	34	34	1	3400				DRR&ATL	DRR&ATL	2312	2312	DRR&ATL
grna_51		U11/U12 small nucleolar RNA	34	34	1	3400				DRR&ATL	DRR&ATL	2313	2313	DRR&ATL
grna_52		U11/U12 small nucleolar RNA	34	34	1	3400				DRR&ATL	DRR&ATL	2314	2314	DRR&ATL
grna_53		U11/U12 small nucleolar RNA	34	34	1	3400				DRR&ATL	DRR&ATL	2315	2315	DRR&ATL
grna_54		U11/U12 small nucleolar RNA	34	34	1	3400				DRR&ATL	DRR&ATL	2316	2316	DRR&ATL
grna_55		U11/U12 small nucleolar RNA	34	34	1	3400				DRR&ATL	DRR&ATL	2317	2317	DRR&ATL
grna_56		U11/U12 small nucleolar RNA	34	34	1	3400				DRR&ATL	DRR&ATL	2318	2318	DRR&ATL
grna_57		U11/U12 small nucleolar RNA	34	34	1	3400				DRR&ATL	DRR&ATL	2319	2319	DRR&ATL
grna_58		U11/U12 small nucleolar RNA	34	34	1	3400				DRR&ATL	DRR&ATL	2320	2320	DRR&ATL
grna_59		U11/U12 small nucleolar RNA	34	34	1	3400				DRR&ATL	DRR&ATL	2321	2321	DRR&ATL
grna_60		U11/U12 small nucleolar RNA	34	34	1	3400				DRR&ATL	DRR&ATL	2322	2322	DRR&ATL
grna_61		U11/U12 small nucleolar RNA	34	34	1	3400				DRR&ATL	DRR&ATL	2323	2323	DRR&ATL
grna_62		U11/U12 small nucleolar RNA	34	34	1	3400				DRR&ATL	DRR&ATL	2324	2324	DRR&ATL
grna_63		U11/U12 small nucleolar RNA	34	34	1	3400				DRR&ATL	DRR&ATL	2325	2325	DRR&ATL
grna_64		U11/U12 small nucleolar RNA	34	34	1	3400				DRR&ATL	DRR&ATL	2326	2326	DRR&ATL
grna_65		U11/U12 small nucleolar RNA	34	34	1	3400				DRR&ATL	DRR&ATL	2327	2327	DRR&ATL
grna_66		U11/U12 small nucleolar RNA	34	34	1	3400				DRR&ATL	DRR&ATL	2328	2328	DRR&ATL
grna_67		U11/U12 small nucleolar RNA	34	34	1	3400				DRR&ATL	DRR&ATL	2329	2329	DRR&ATL
grna_68		U11/U12 small nucleolar RNA	34	34	1	3400				DRR&ATL	DRR&ATL	2330	2330	DRR&ATL
grna_69		U11/U12 small nucleolar RNA	34	34	1	3400				DRR&ATL	DRR&ATL	2331	2331	DRR&ATL
grna_70		U11/U12 small nucleolar RNA	34	34	1	3400				DRR&ATL	DRR&ATL	2332	2332	DRR&ATL
grna_71		U11/U12 small nucleolar RNA	34	34	1	3400				DRR&ATL	DRR&ATL	2333	2333	DRR&ATL
grna_72		U11/U12 small nucleolar RNA	34	34	1	3400				DRR&ATL	DRR&ATL	2334	2334	DRR&ATL
grna_73		U11/U12 small nucleolar RNA	34	34	1	3400				DRR&ATL	DRR&ATL	2335	2335	DRR&ATL
grna_74		U11/U12 small nucleolar RNA	34	34	1	3400				DRR&ATL	DRR&ATL	2336	2336	DRR&ATL
grna_75		U11/U12 small nucleolar RNA	34	34	1	3400				DRR&ATL	DRR&ATL	2337	2337	DRR&ATL
grna_76		U11/U12 small nucleolar RNA	34	34	1	3400				DRR&ATL	DRR&ATL	2338	2338	DRR&ATL
grna_77		U11/U12 small nucleolar RNA	34	34	1	3400				DRR&ATL	DRR&ATL	2339	2339	DRR&ATL
grna_78		U11/U12 small nucleolar RNA	34	34	1	3400				DRR&ATL	DRR&ATL	2340	2340	DRR&ATL
grna_79		U11/U12 small nucleolar RNA	34	34	1	3400				DRR&ATL	DRR&ATL	2341	2341	DRR&ATL
grna_80		U11/U12 small nucleolar RNA	34	34	1	3400				DRR&ATL	DRR&ATL	2342	2342	DRR&ATL
grna_81		U11/U12 small nucleolar RNA	34	34	1	3400				DRR&ATL	DRR&ATL	2343	2343	DRR&ATL
grna_82		U11/U12 small nucleolar RNA	34	34	1	3400				DRR&ATL	DRR&ATL	2344	2344	DRR&ATL
grna_83		U11/U12 small nucleolar RNA	34	34	1	3400				DRR&ATL	DRR&ATL	2345	2345	DRR&ATL
grna_84		U11/U12 small nucleolar RNA	34	34	1	3400				DRR&ATL	DRR&ATL	2346	2346	DRR&ATL
grna_85		U11/U12 small nucleolar RNA	34	34	1	3400				DRR&ATL	DRR&ATL	2347	2347	DRR&ATL
grna_86		U11/U12 small nucleolar RNA	34	34	1	3400				DRR&ATL	DRR&ATL	2348	2348	DRR&ATL
grna_87		U11/U12 small nucleolar RNA	34	34	1	3400				DRR&ATL	DRR&ATL	2349	2349	DRR&ATL
grna_88		U11/U12 small nucleolar RNA	34	34	1	3400				DRR&ATL	DRR&ATL	2350	2350	DRR&ATL
grna_89		U11/U12 small nucleolar RNA	34	34	1	3400				DRR&ATL	DRR&ATL	2351	2351	DRR&ATL
grna_90		U11/U12 small nucleolar RNA	34	34	1	3400				DRR&ATL	DRR&ATL	2352	2352	DRR&ATL
grna_91		U11/U12 small nucleolar RNA	34	34	1	3400				DRR&ATL	DRR&ATL	2353	2353	DRR&ATL
grna_92		U11/U12 small nucleolar RNA	34	34	1	3400				DRR&ATL	DRR&ATL	2354	2354	DRR&ATL
grna_93		U11/U12 small nucleolar RNA	34	34	1	3400				DRR&ATL	DRR&ATL	2355	2355	DRR&ATL
grna_94		U11/U12 small nucleolar RNA	34	34	1	3400				DRR&ATL	DRR&ATL	2356	2356	DRR&ATL
grna_95		U11/U12 small nucleolar RNA	34	34	1	3400			</td					

// gene_presence_absence.Rtab

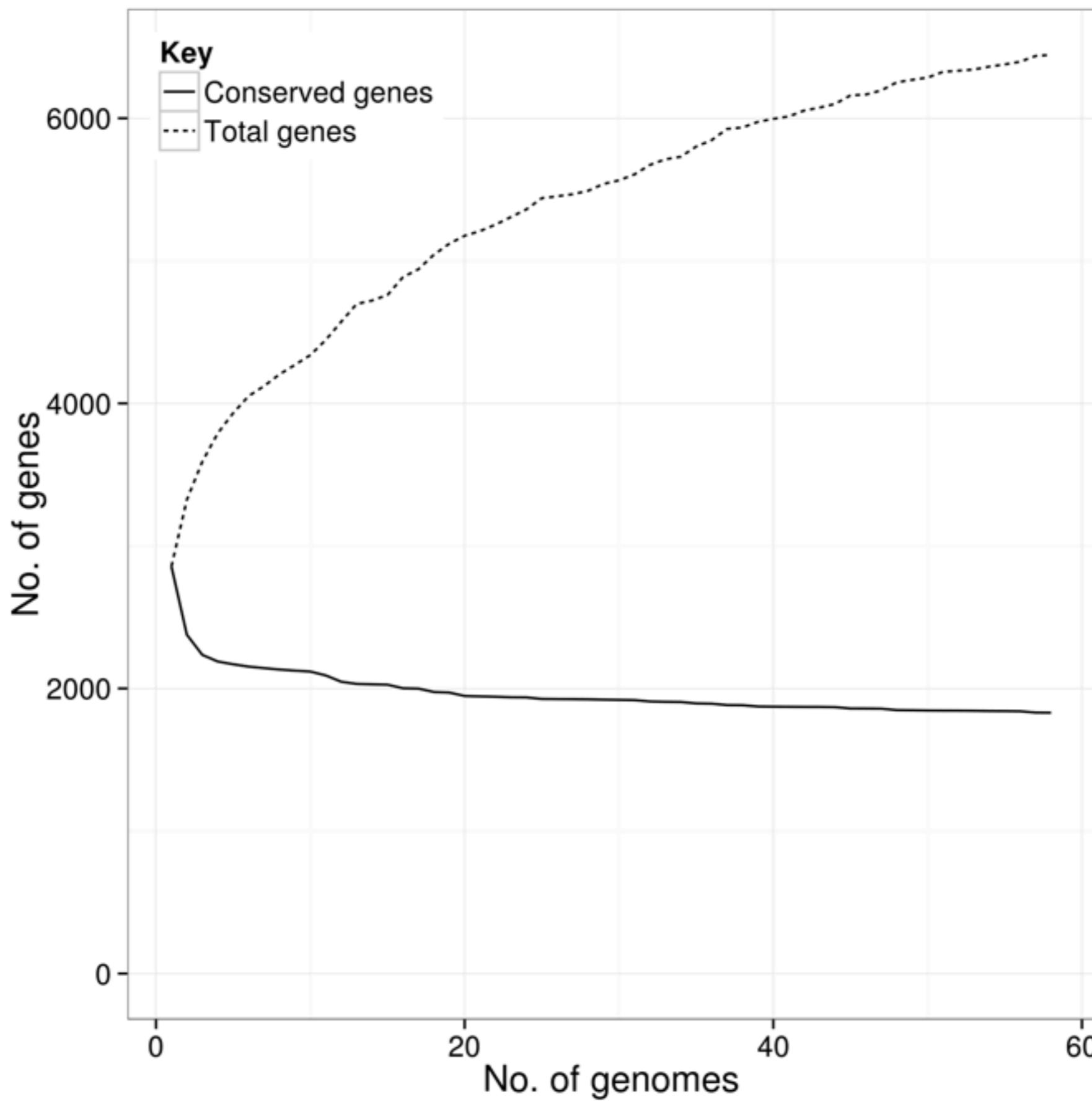
This file is similar to the gene_presence_absence.csv file, however it just contains a simple tab delimited binary matrix with the presence and absence of each gene in each sample. It can be easily loaded into R using the read.table function, which opens up a world of powerful tools. The first row is the header containing the name of each sample, and the first column contains the gene name. A 1 indicates the gene is present in the sample, a 0 indicates it is absent.

// pan_genome_reference.fa

This is a FASTA file which contains a single representative nucleotide sequence from each of the clusters in the pan genome (core and accessory). The name of each sequence is the source sequence ID followed by the cluster it came from. This file can be of use for reference guided assembly, whole genome MLST or for mapping raw reads to it.

// *.Rtab

There is an additional script called *create_pan_genome_plots.R* which requires R and the ggplot2 library. It takes in the *.Rtab files and produces graphs on how the pan genome varies as genomes are added (in random orders).



// accessory_binary_genes.fa.newick

This is a tree created using the binary presence and absence of accessory genes. It is in Newick format and can be viewed in FigTree. It is only a quick and dirty tree to roughly group isolates together based on their accessory genome and is in no way reliable other than to give a quick insight into the data. If you want a more accurate tree you need to use the core gene alignment as your starting point.

// accessory_graph.dot

A graph in DOT format of how genes are linked together at the contig level in the accessory genome. You can view it using [Gephi](#). Nodes correspond to genes. An edge represents where two genes are next to each other on any of the input assemblies. To generate the weights on the edges, a count is kept of the frequency of the genes being beside each other, they are then weighted by the overall clustering of the accessory binary presence and absense (to reduce bias by over represented genotypes), then inverted.

// core_accessory_graph.dot

A graph in DOT format of how genes are linked together at the contig level in the pan genome. You can view it using [Gephi](#).

// clustered_proteins

Groups file where each line lists the sequences in a cluster.

// core_gene_alignment.aln

If you pass in the `-e` parameter to roary, a multi-FASTA alignment of all of the core genes is created. By default it uses PRANK (Löytynoja, 2014) which performs a codon aware alignment. It is slow but accurate. If you pass in `-e --mafft` it will use MAFFT which performs a nucleotide alignment. It is very fast but less accurate. This can then be used as input to build a phylogenetic tree. To reduce the memory and run time, you can pre filter the alignment using [snp_sites](#). Just be aware that recombination is not taken care of with this method.

/// query_pan_genome

Perform set operations on the pan genome to see the gene differences between groups of isolates.

Options:

-g STR	groups filename [clustered_proteins]
-a STR	action (union/intersection/complement/gene_multifasta/difference) [
-c FLOAT	percentage of isolates a gene must be in to be core [99]
-o STR	output filename [pan_genome_results]
-n STR	comma separated list of gene names for use with gene_multifasta act
-i STR	comma separated list of filenames, comparison set one
-t STR	comma separated list of filenames, comparison set two
-v	verbose output to STDOUT
-h	this help message

Union of genes found in isolates

```
query_pan_genome -a union *.gff
```

Intersection of genes found in isolates (core genes)

```
query_pan_genome -a intersection *.gff
```

Complement of genes found in isolates (accessory genes)

```
query_pan_genome -a complement *.gff
```

Extract the sequence of each gene listed and create multi-FASTA files

```
query_pan_genome -a gene_multifasta -n gryA,mecA,abc *.gff
```

Gene differences between sets of isolates

```
query_pan_genome -a difference --input_set_one 1.gff,2.gff --input_set_two 3.gff,4.gff
```

Union of genes found in isolates

```
query_pan_genome -a union *.gff
```

Intersection of genes found in isolates (core genes)

```
query_pan_genome -a intersection *.gff
```

Complement of genes found in isolates (accessory genes)

```
query_pan_genome -a complement *.gff
```

Extract the sequence of each gene listed and create multi-FASTA files

```
query_pan_genome -a gene_multifasta -n gryA,mecA,abc *.gff
```

Gene differences between sets of isolates

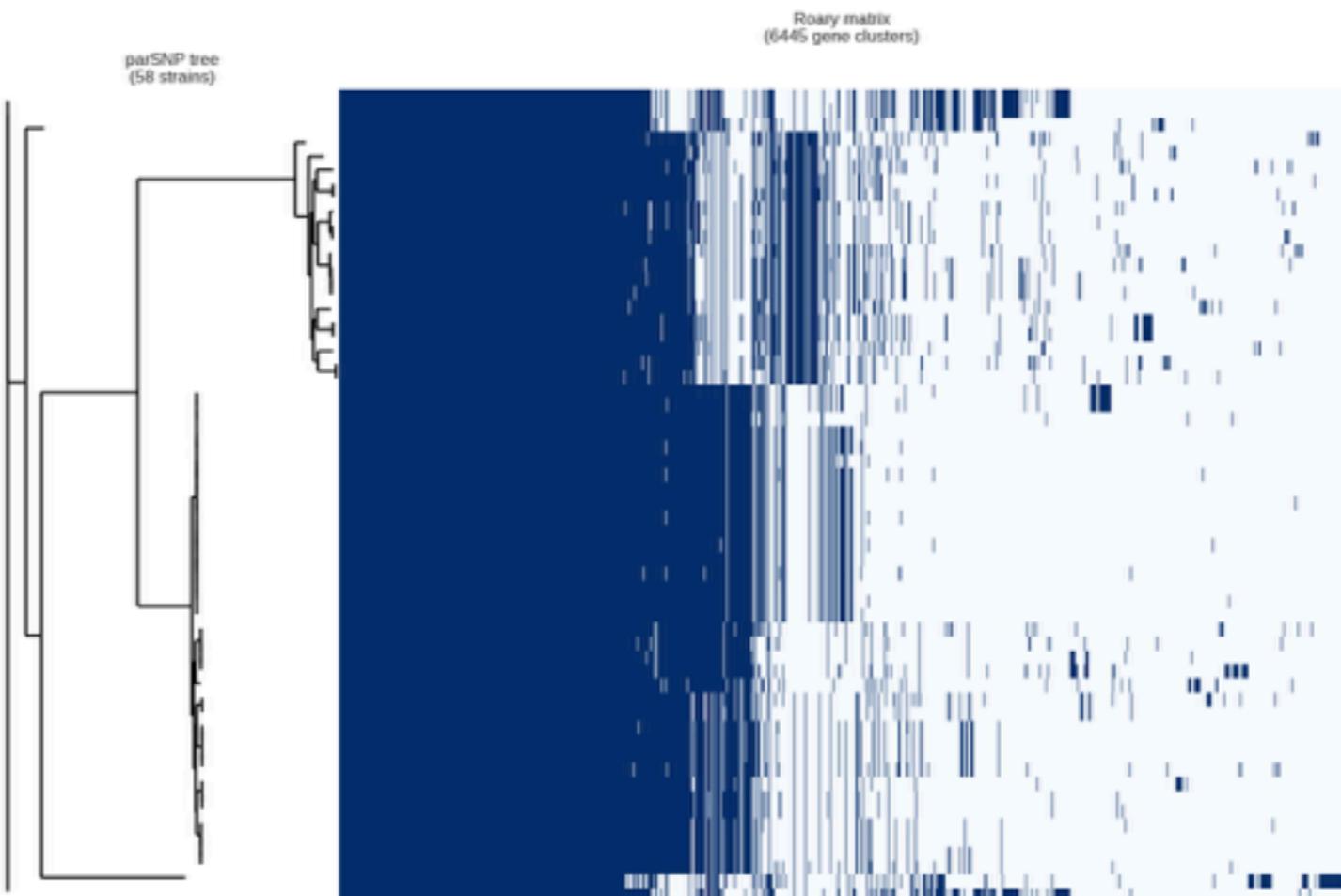
```
query_pan_genome -a difference --input_set_one 1.gff,2.gff --input_set_two 3.gff,4.gff
```

Visualising results

/// roary_plots.py

This contributed script by Marco Galardini is not installed by default but can be very useful. [Additional details can be found here in the repository](#). It provides 3 figures, showing the tree compared to a matrix with the presence and absence of core and accessory genes. The next is an pie chart of the breakdown of genes and the number of isolate they are present in. And finally there is a graph with the frequency of genes versus the number of genomes.

```
roary_plots.py name_of_your_newick_tree_file.tre gene_presence_absence.csv
```



Cross-Border Transmission of *Salmonella*

Choleraesuis var. Kunzendorf in European Pigs and Wild Boar: Infection, Genetics, and Evolution

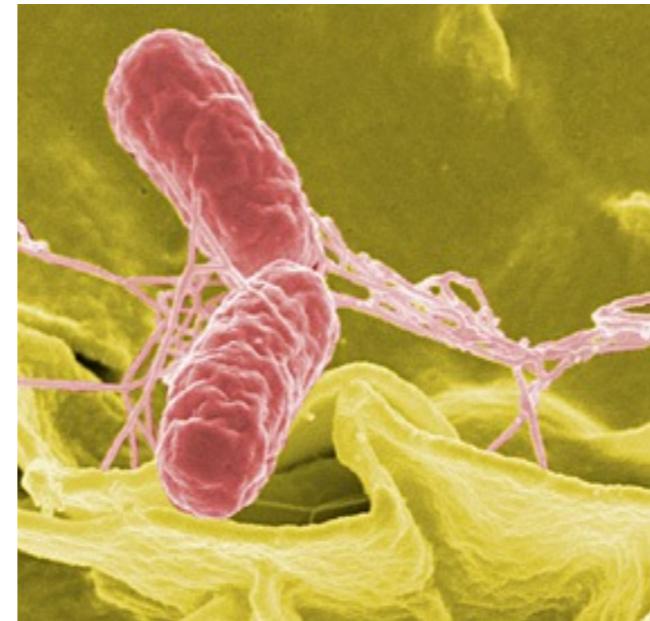
“Shinny” Pimplapas Leekitcharoenphon
Research Group for Genomic Epidemiology
DTU Food
Technical University of Denmark (DTU)

$$f(x+\Delta x) = \sum_{i=0}^{\infty} \frac{(\Delta x)^i}{i!} f^{(i)}(x)$$
$$\int_a^b \Theta^{\sqrt{17}} + \Omega \int \delta e^{i\pi} =$$
$$\infty = \frac{1}{\chi^2} \sum \Sigma !,$$

Salmonella

- *Salmonella*

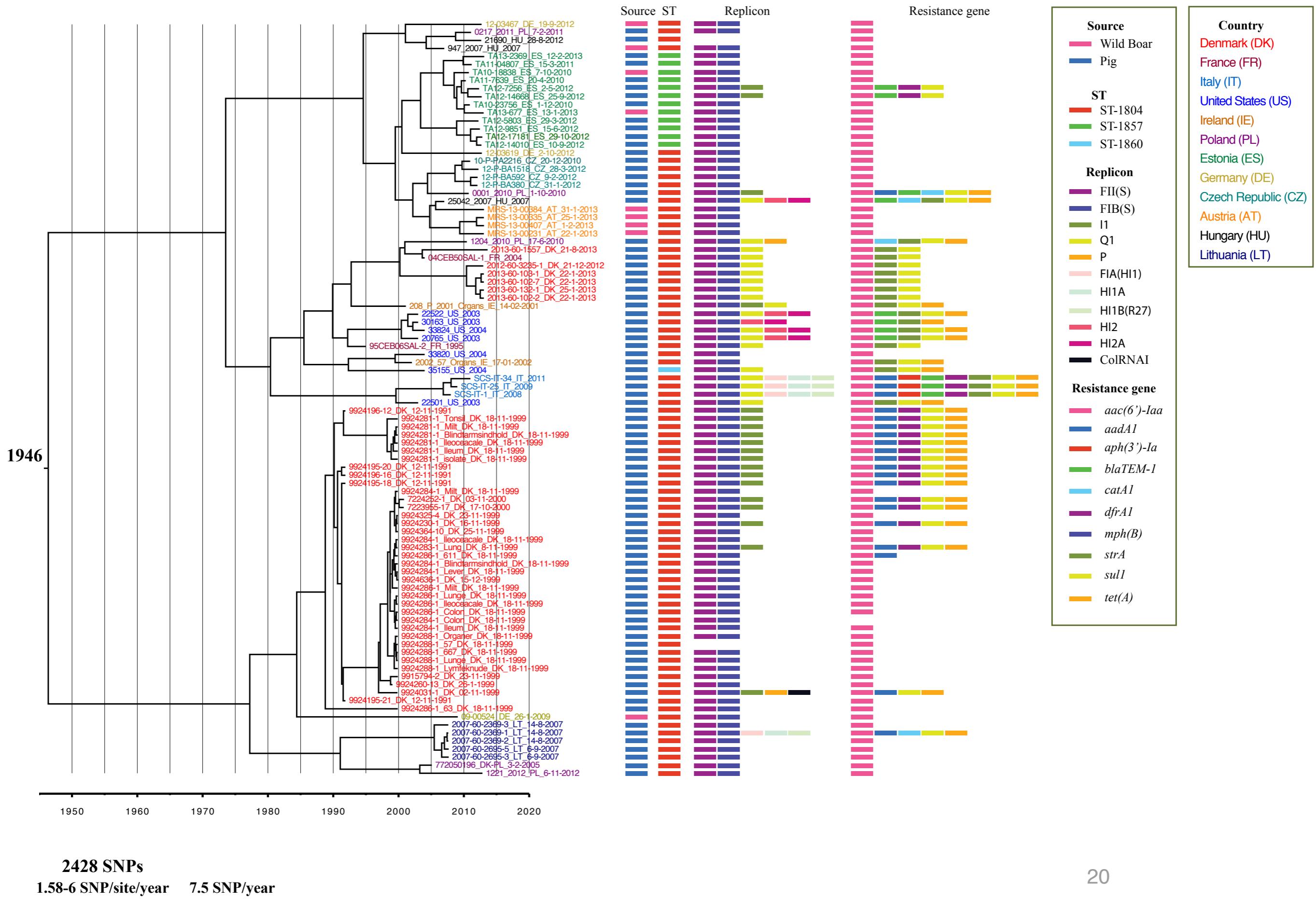
- *S.bongori*
- *S.enterica*
 - Salamae
 - arizonae
 - diarizone
 - houtenae
 - Indica
 - enterica
 - more than 2500 serotypes
 - Typhimurium, Typhi, paratyphi, etc.
 - phagetypes, e.g. DT104

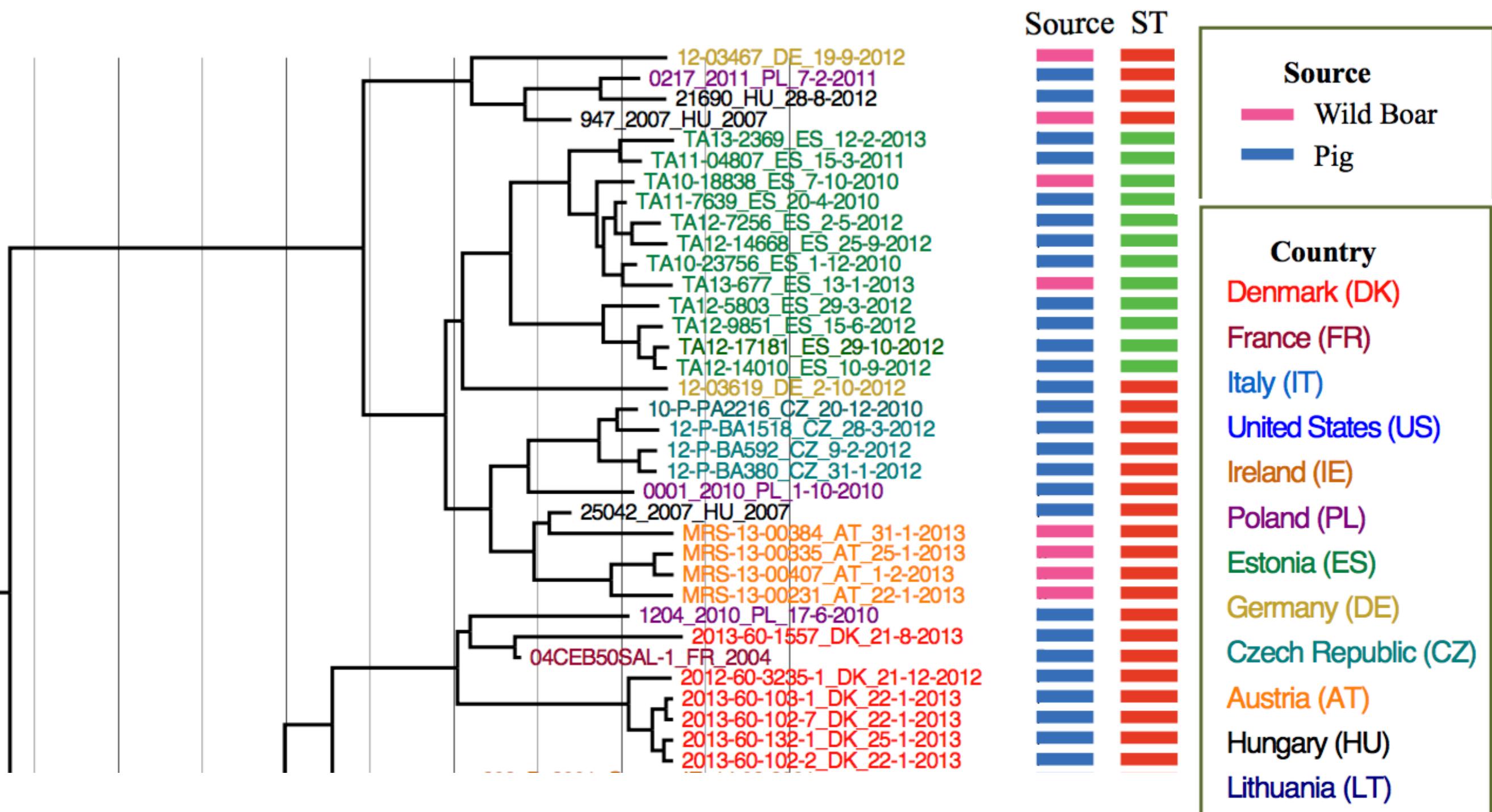


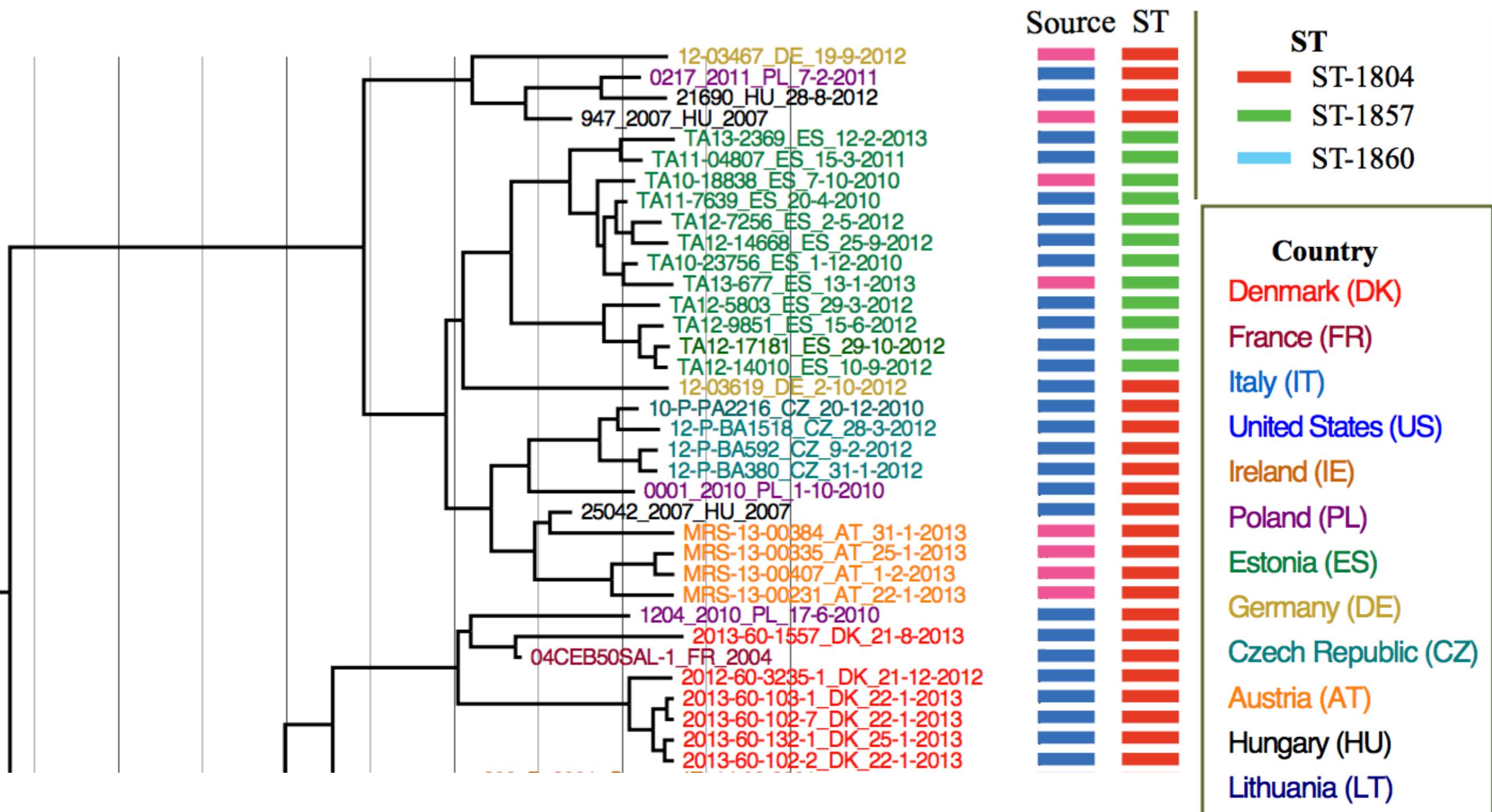
© Martin Myers, MD

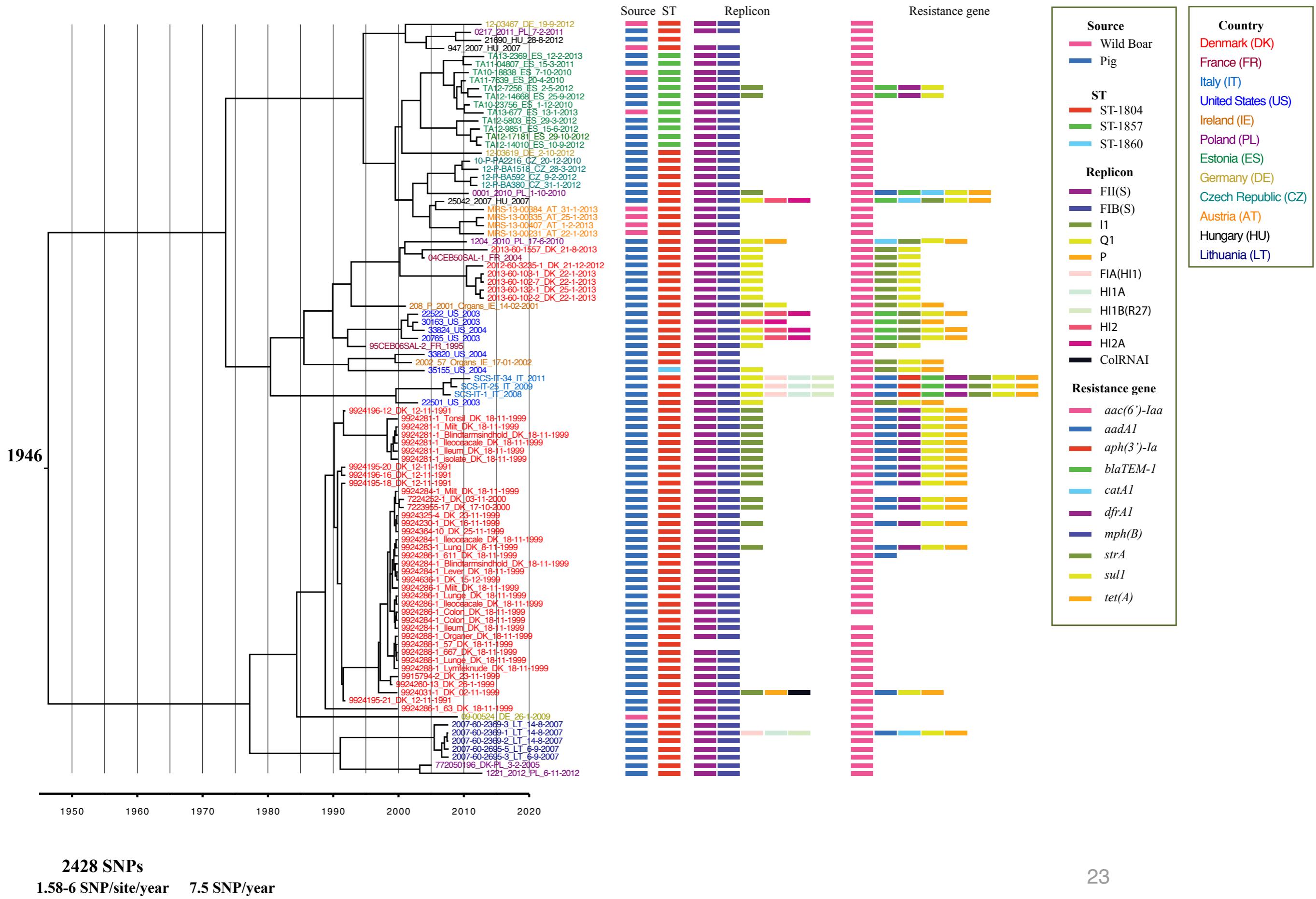
S. Choleraesuis* var. *Kunzendorf

- *S. Choleraesuis* is a porcine adapted serovar which may cause serious outbreak in pigs.
- The majority of the *S. Choleraesuis* outbreaks in pigs are caused by var. *Kunzendorf*
- In Europe, *S. Choleraesuis* is a relatively rare serovar but it has been reported with low frequency in a number of countries
- In Denmark, *S. Choleraesuis* was last found in pigs at an outbreak in 1999, but in 2012 and 2013 it reappeared with outbreak of severe salmonellosis in four farms.

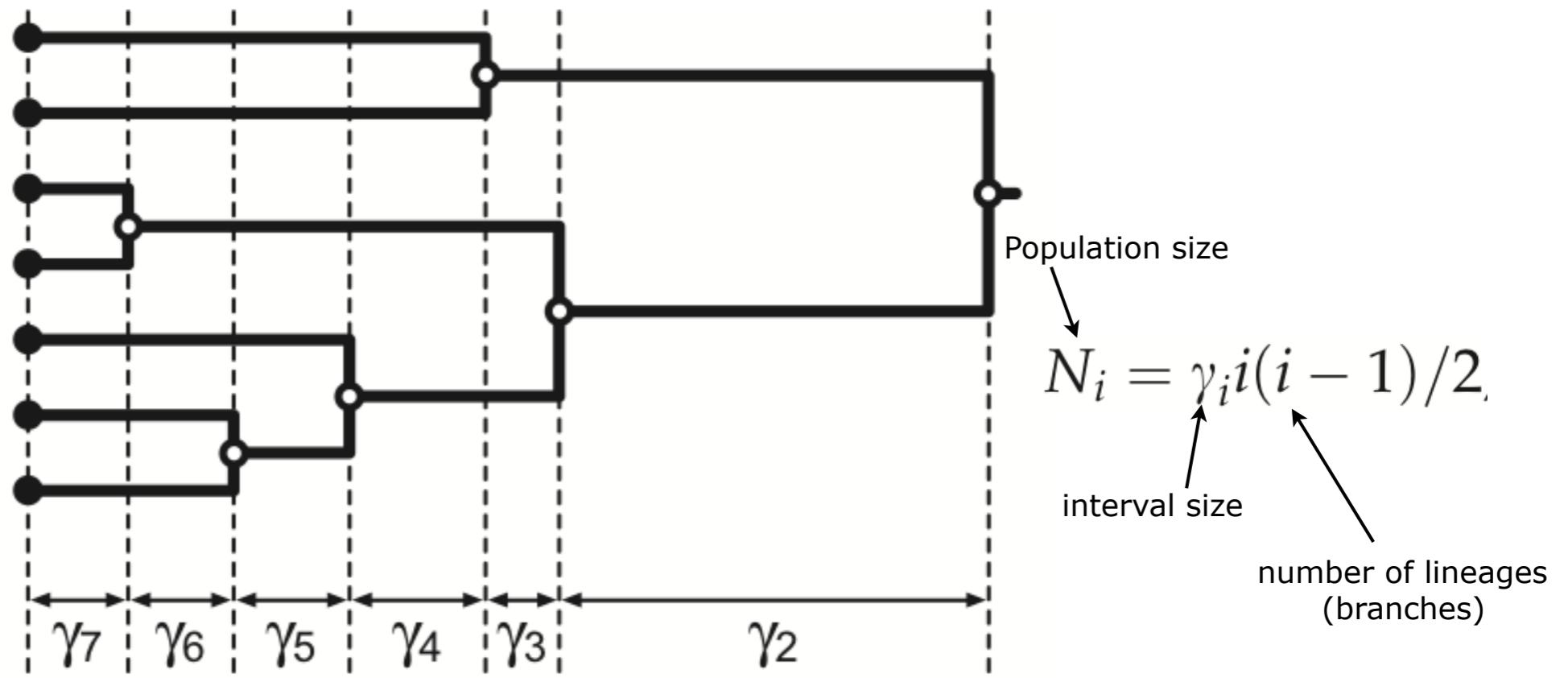


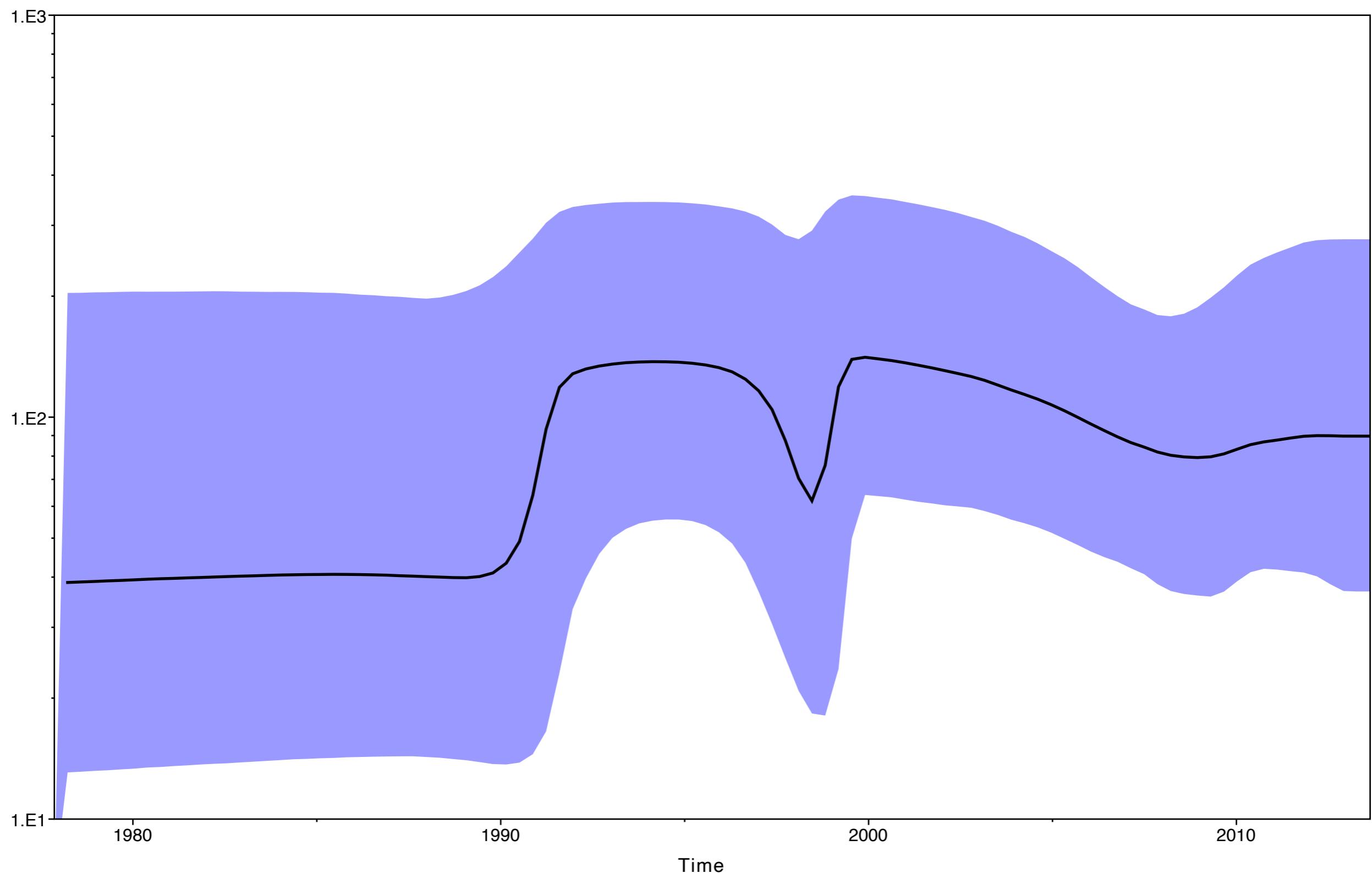






Demographic history using Bayesian skyline plot

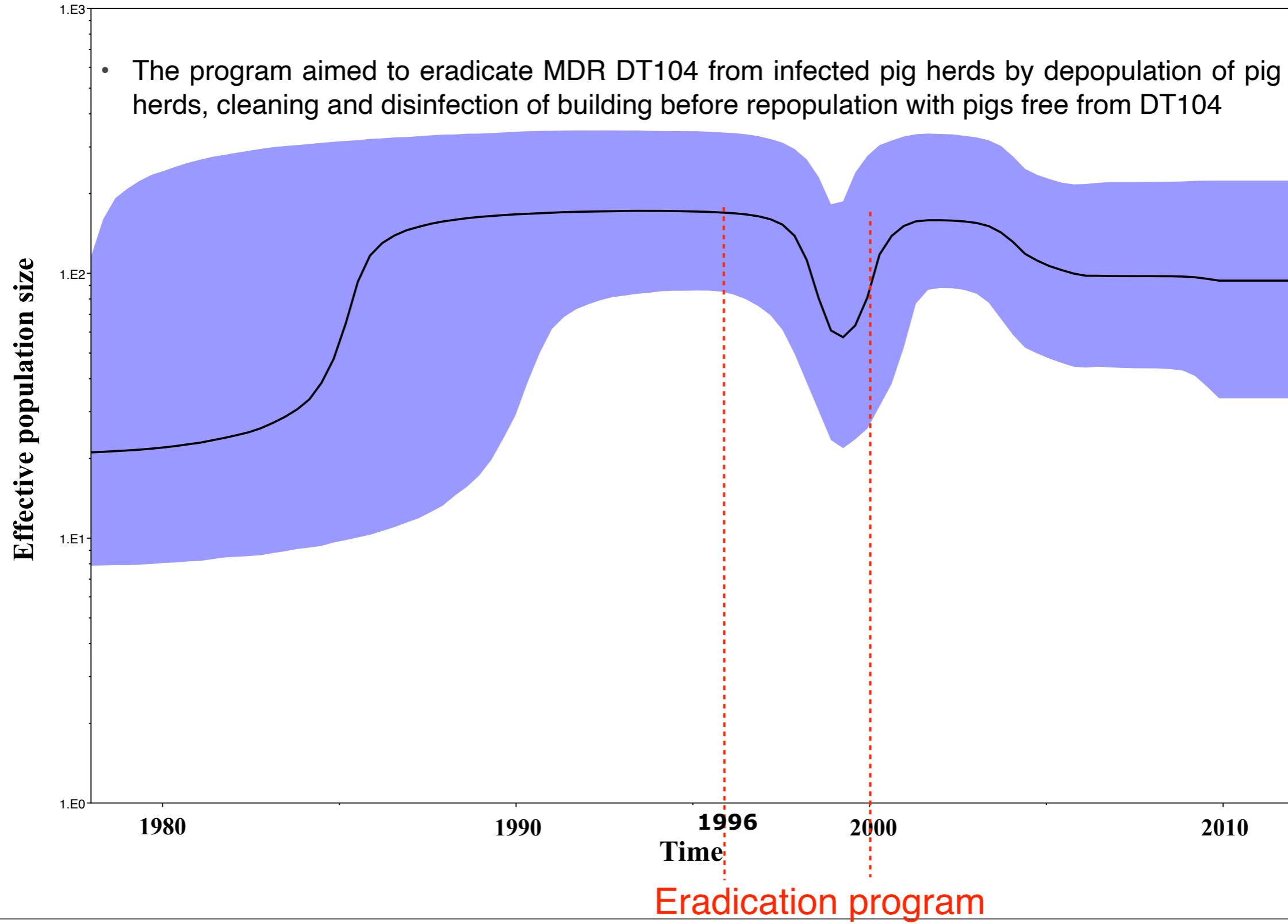


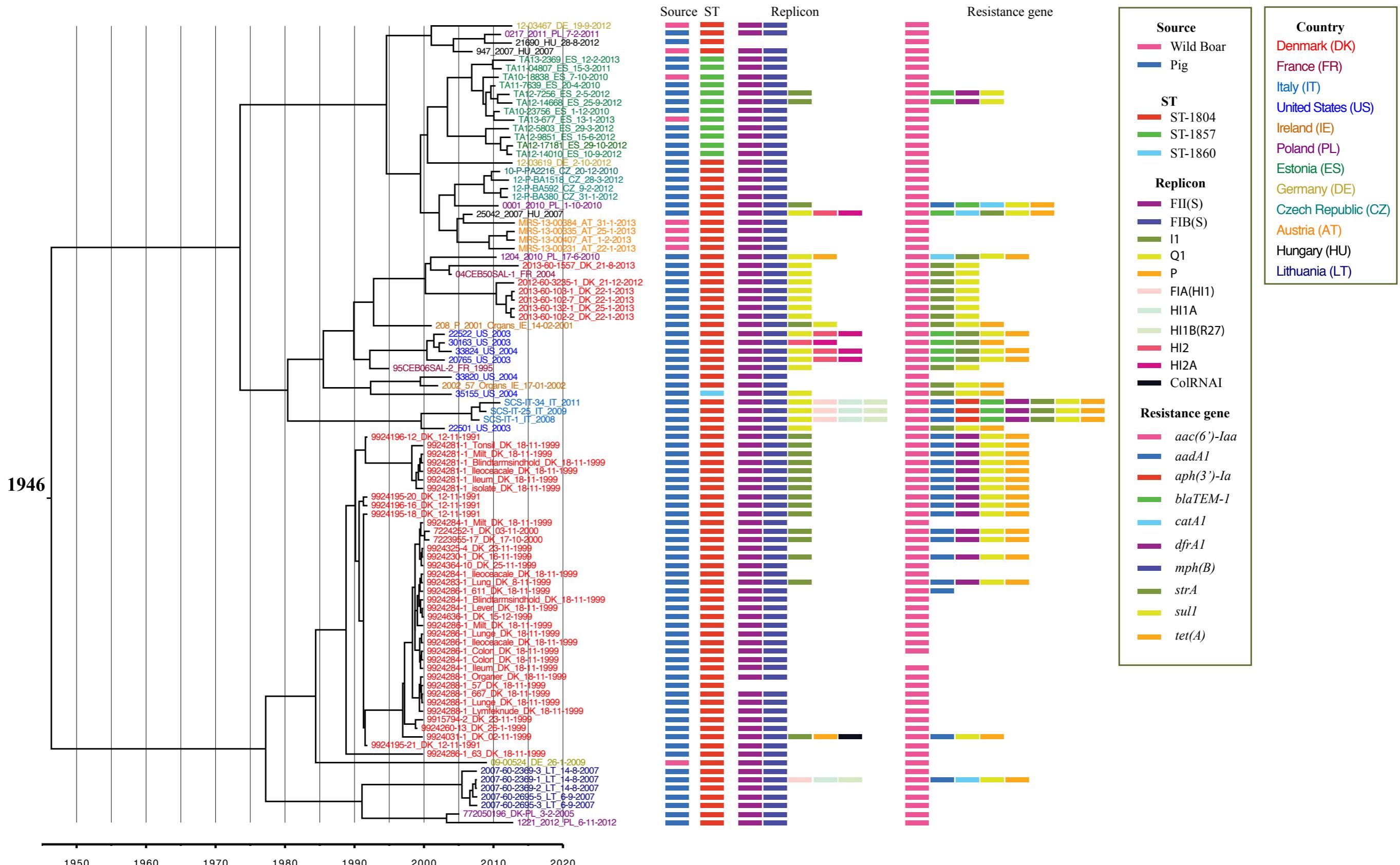


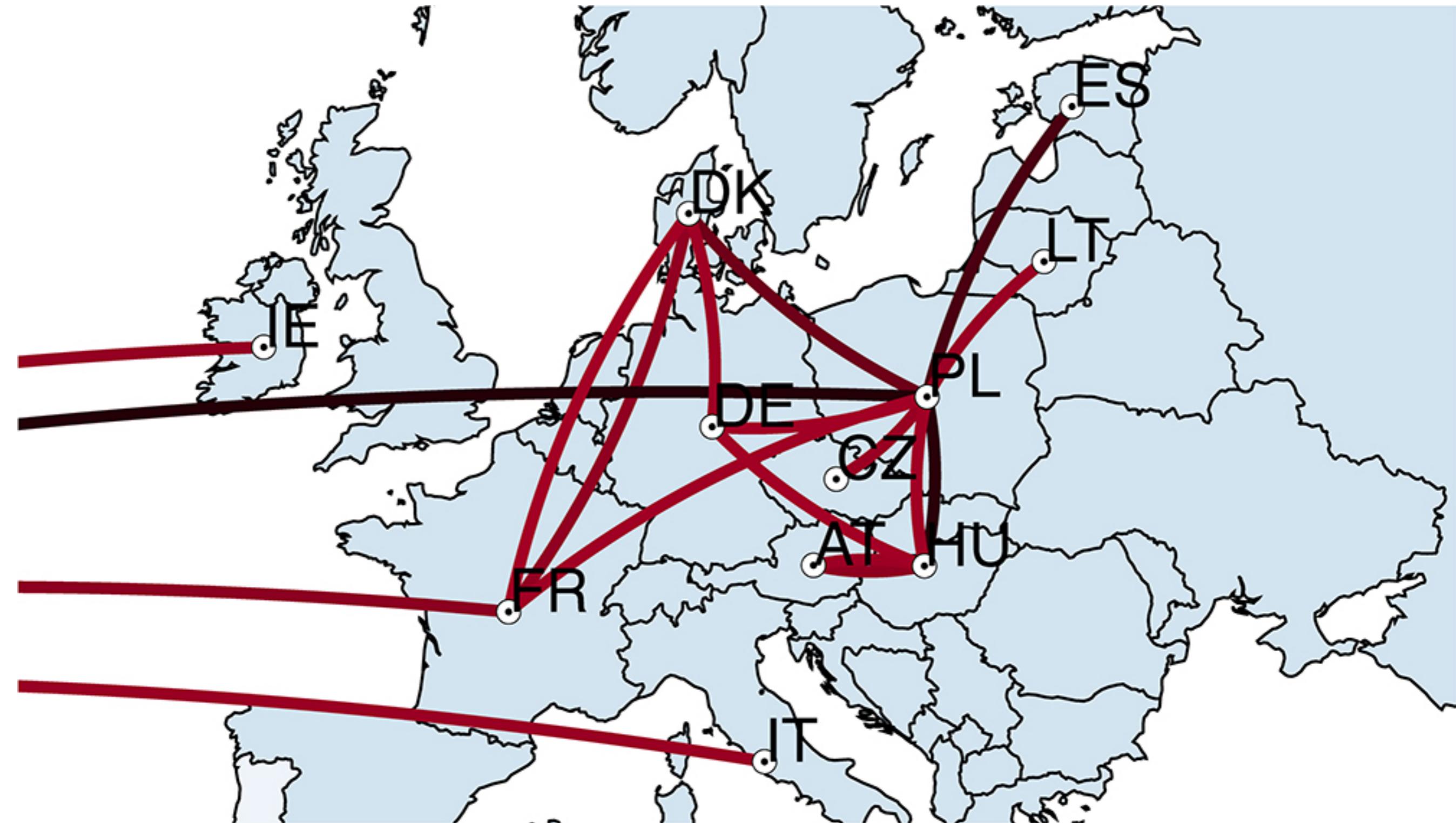
Demographic history of Danish MDR DT104

- The evidence of the success of the eradication program implemented by the Federation of Danish Pig Producers and Slaughterhouse in collaboration with the Danish Veterinary Service

- The program aimed to eradicate MDR DT104 from infected pig herds by depopulation of pig herds, cleaning and disinfection of building before repopulation with pigs free from DT104





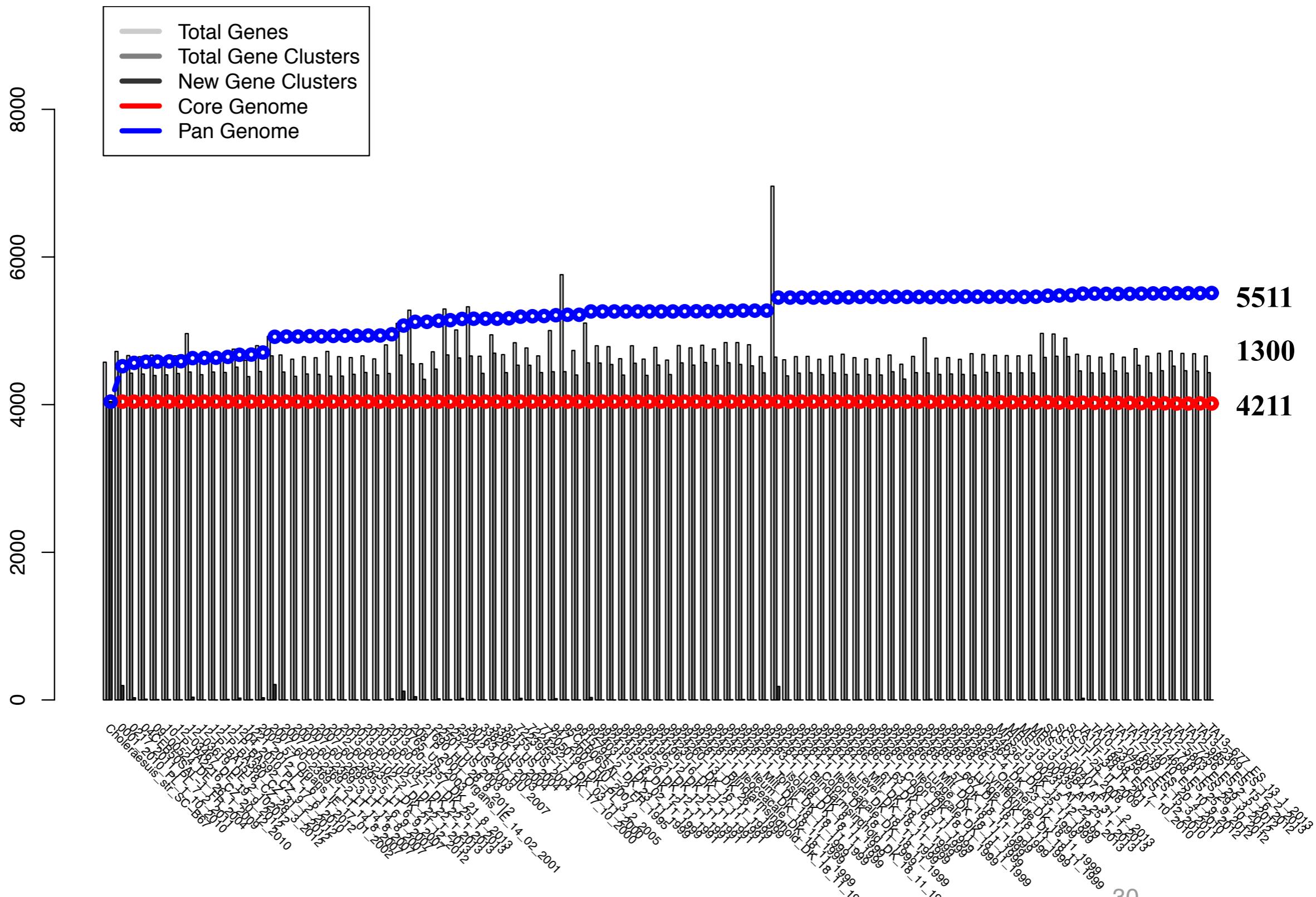


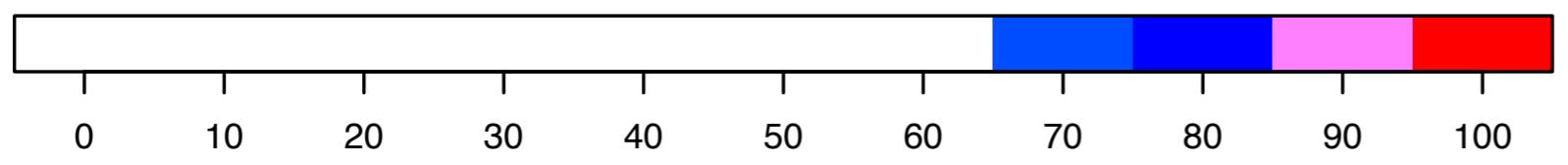
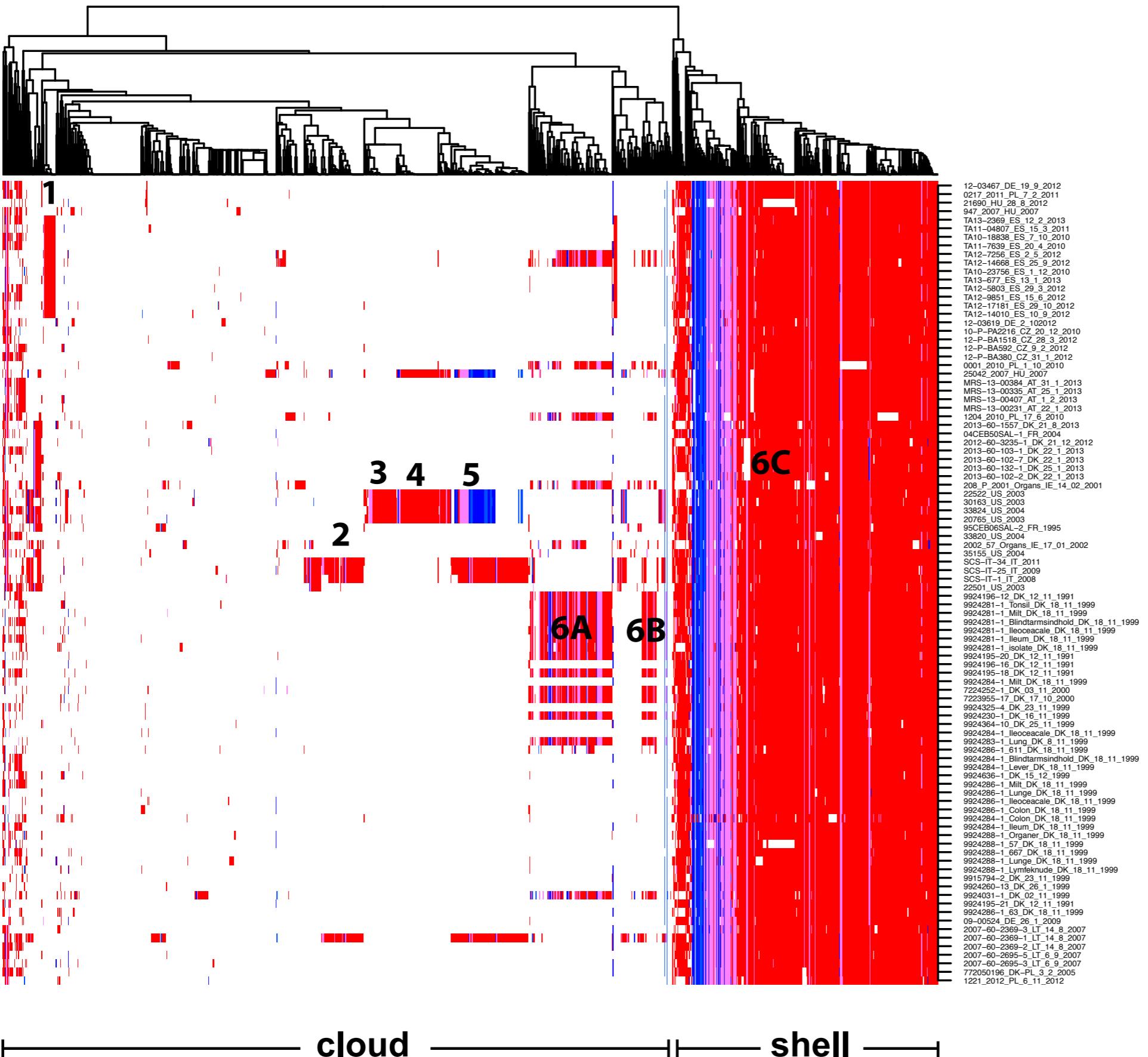
A 1999

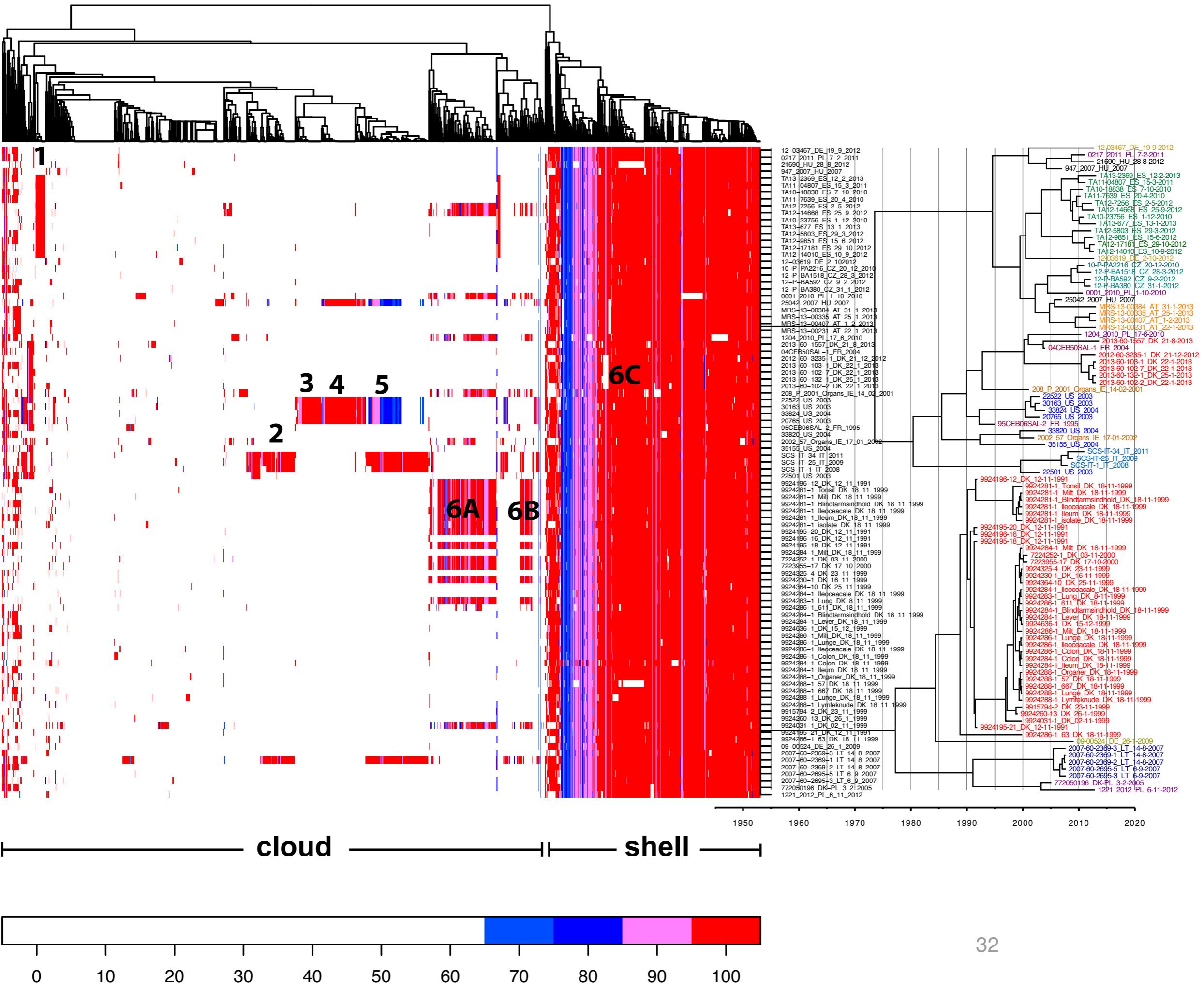


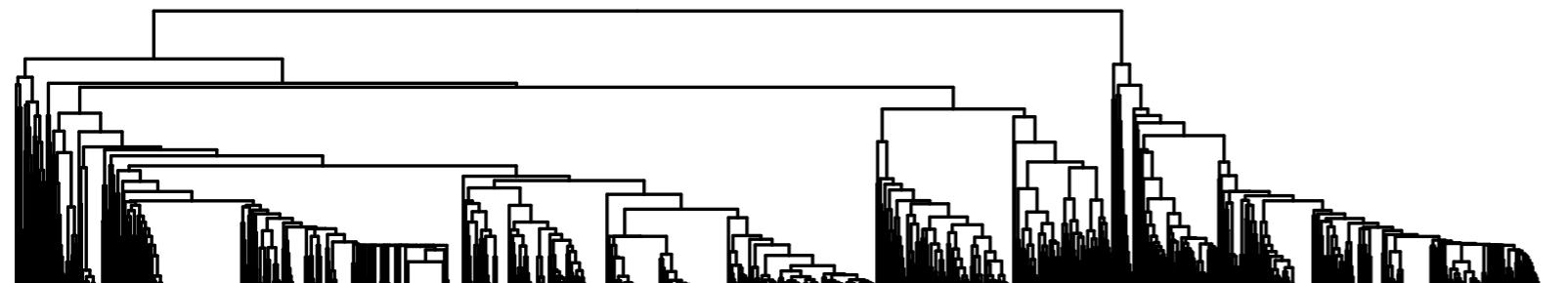
B 2013







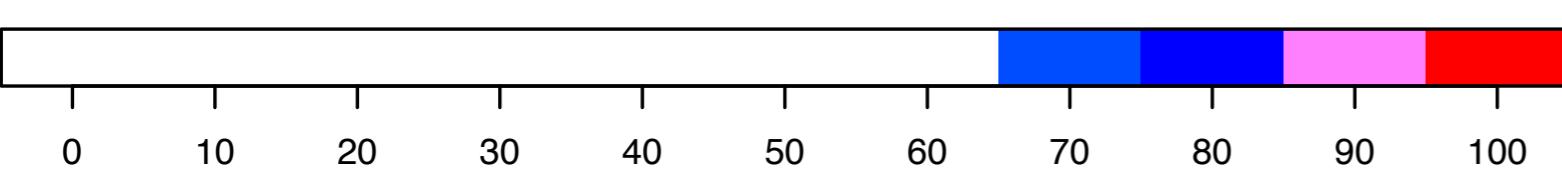


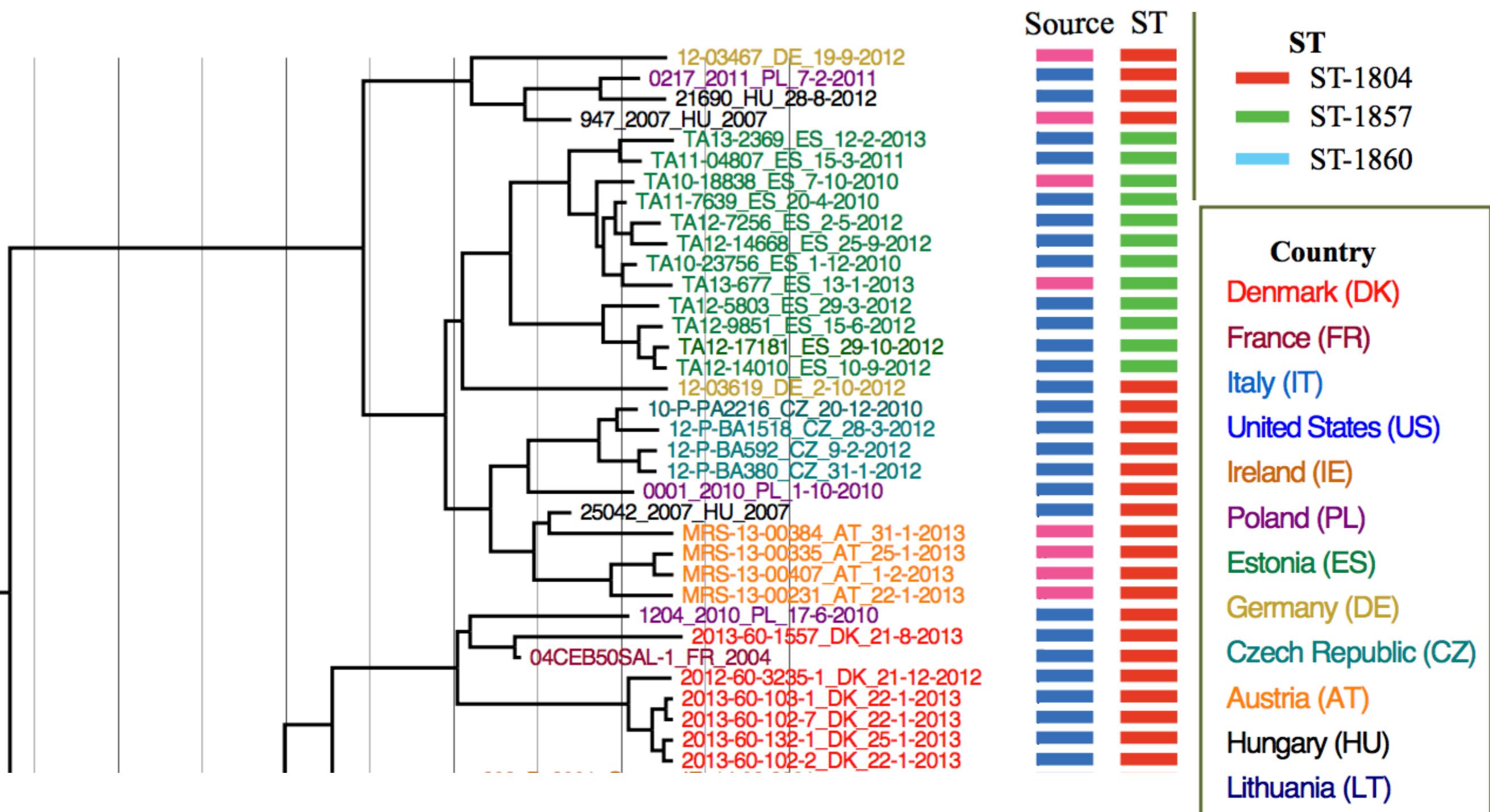


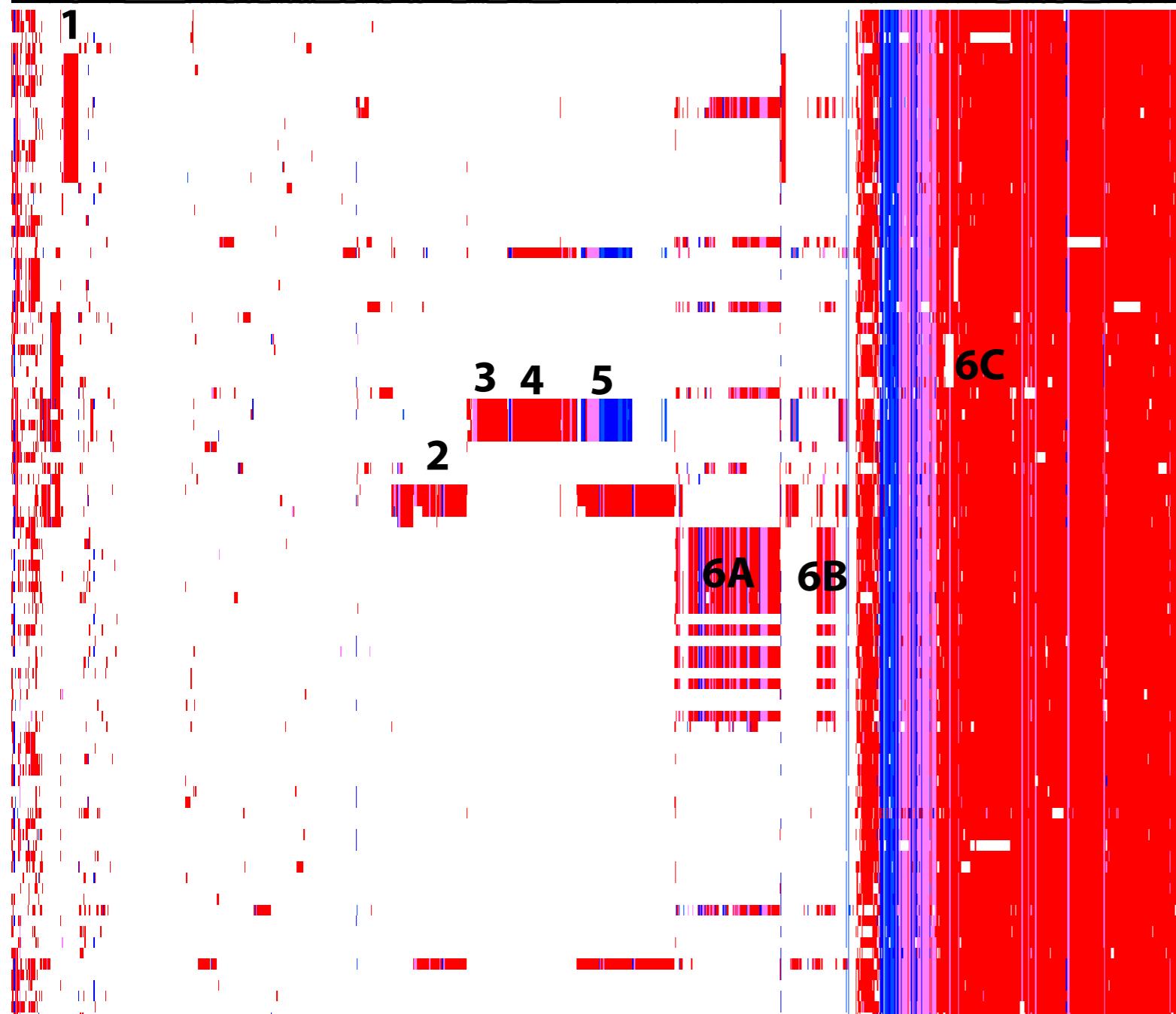
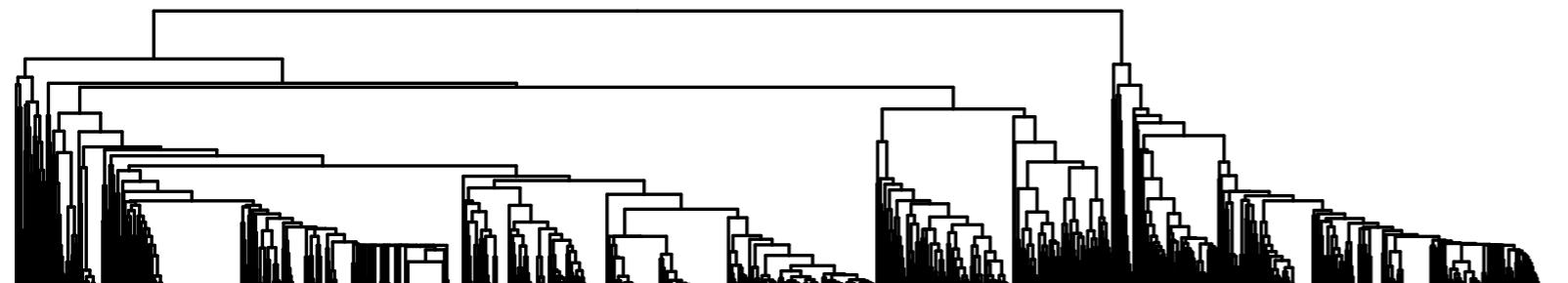
12-03467_DE_19_9_2012
 0217_2011_PL_7_2_2011
 21690_HU_28_8_2012
 947_2007_HU_2007
 TA13-2369_ES_12_2_2013
 TA11-04807_ES_15_3_2011
 TA10-18838_ES_7_10_2010
 TA11-7639_ES_20_4_2010
 TA12-23756_ES_2_5_2012
 TA12-14668_ES_25_9_2012
 TA10-23756_ES_1_12_2010
 TA13-677_ES_13_1_2013
 TA12-5803_ES_29_3_2012
 TA12-9851_ES_15_6_2012
 TA12-17181_ES_29_10_2012
 TA12-14010_ES_10_9_2012
 12-03619_DE_2_102012
 10-P-PA2216_CZ_20_12_2010
 12-P-BA1516_CZ_28_3_2012
 12-P-BA592_CZ_9_2_2012
 12-P-BA380_CZ_31_1_2012
 0001_2010_PL_1_10_2010
 25042_2007_HU_2007
 MRS-13-00384_AT_31_1_2013
 MRS-13-00335_AT_25_1_2013
 MRS-13-00407_AT_1_2_2013
 MRS-13-00231_AT_22_1_2013
 1204_2010_PL_17_6_2010
 2013-60-1557_DK_21_8_2013
 04CEB50SAL-1_FR_2004
 2012-60-3235-1_DK_21_12_2012
 2013-60-103-1_DK_22_1_2013
 2013-60-102-7_DK_22_1_2013
 2013-60-132-1_DK_25_1_2013
 2013-60-102-2_DK_22_1_2013
 208_P_2001_Organs_IE_14_02_2001
 22522_US_2003
 30163_US_2003
 38824_US_2004
 20765_US_2003
 95CEB06SAL-2_FR_1995
 38820_US_2004
 2002_57_Organs_IE_17_01_2002
 39155_US_2004
 SCS-IT-34_IT_2011
 SCS-IT-25_IT_2009
 SCS-IT-1_IT_2008
 22501_US_2004
 9924196-12_DK_12_11_1991
 9924281-1_Ionsit_DK_18_11_1999
 9924281-1_Mitt_DK_18_11_1999
 9924281-1_Blindarmsmehold_DK_18_11_1999
 9924281-1_Ileocecale_DK_18_11_1999
 9924281-1_Ileum_DK_18_11_1999
 9924281-1_isolate_DK_18_11_1999
 9924195-20_DK_12_11_1991
 9924196-16_DK_12_11_1991
 9924195-18_DK_12_11_1991
 9924284-1_Mitt_DK_18_11_1999
 7224252-1_DK_03_11_2000
 7223955-17_DK_17_10_2000
 9924325-4_DK_23_11_1999
 9924230-1_DK_16_11_1999
 9924364-10_DK_25_11_1999
 9924284-1_Ileocecale_DK_18_11_1999
 9924283-1_Lunge_DK_8_11_1999
 9924286-1_611_DK_18_11_1999
 9924284-1_Blindarmsmehold_DK_18_11_1999
 9924284-1_Lever_DK_18_11_1999
 9924263-1_DK_15_12_1999
 9924286-1_Mitt_DK_18_11_1999
 9924286-1_Lunge_DK_18_11_1999
 9924286-1_Ileocecale_DK_18_11_1999
 9924286-1_Colon_DK_18_11_1999
 9924284-1_Colon_DK_18_11_1999
 9924284-1_Ileum_DK_18_11_1999
 9924288-1_Organer_DK_18_11_1999
 9924288-1_57_DK_18_11_1999
 9924288-1_667_DK_18_11_1999
 9924288-1_Lunge_DK_18_11_1999
 9924288-1_Lymfeknude_DK_18_11_1999
 9915794-2_DK_23_11_1999
 9924260-13_DK_26_1_1999
 9924031-1_DK_02_11_1999
 9924195-21_DK_12_11_1991
 9924286-1_63_DK_18_11_1999
 09-00524_DE_26_1_2009
 2007-60-2369-3_LT_14_8_2007
 2007-60-2369-1_LT_14_8_2007
 2007-60-2369-2_LT_14_8_2007
 2007-60-2695-5_LT_6_9_2007
 2007-60-2695-3_LT_6_9_2007
 772050196_DK-PL_3_2_2005
 1221_2012_PL_6_11_2012

- **Group 1 (17 genes)**
- None plasmid related genes

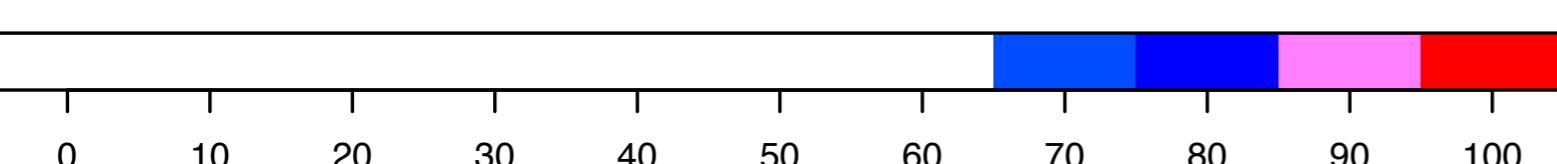
cloud shell



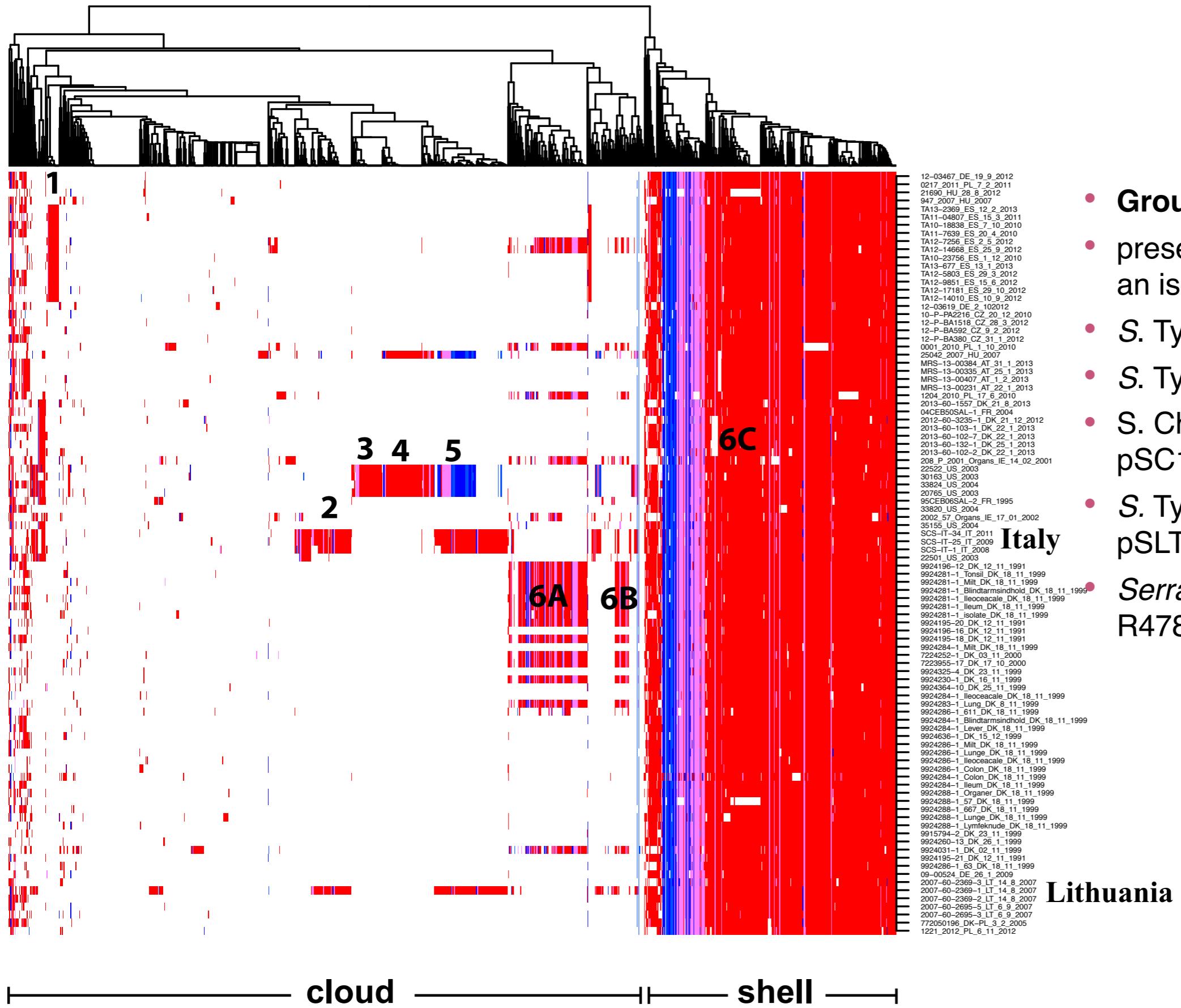




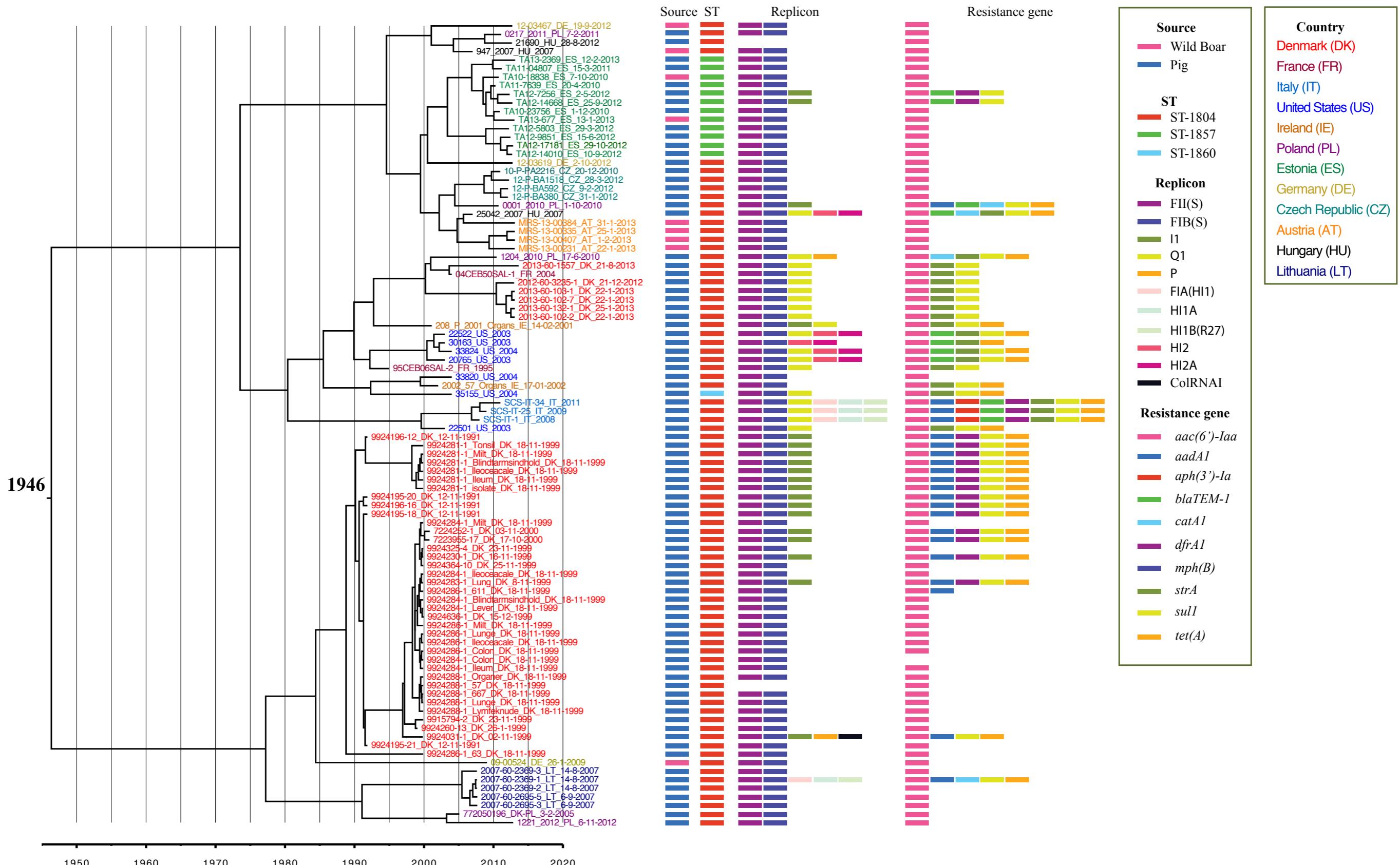
— cloud — shell —



- **Group 1 (17 genes)**
- None plasmid related genes
- Potentially connect to ST-1857

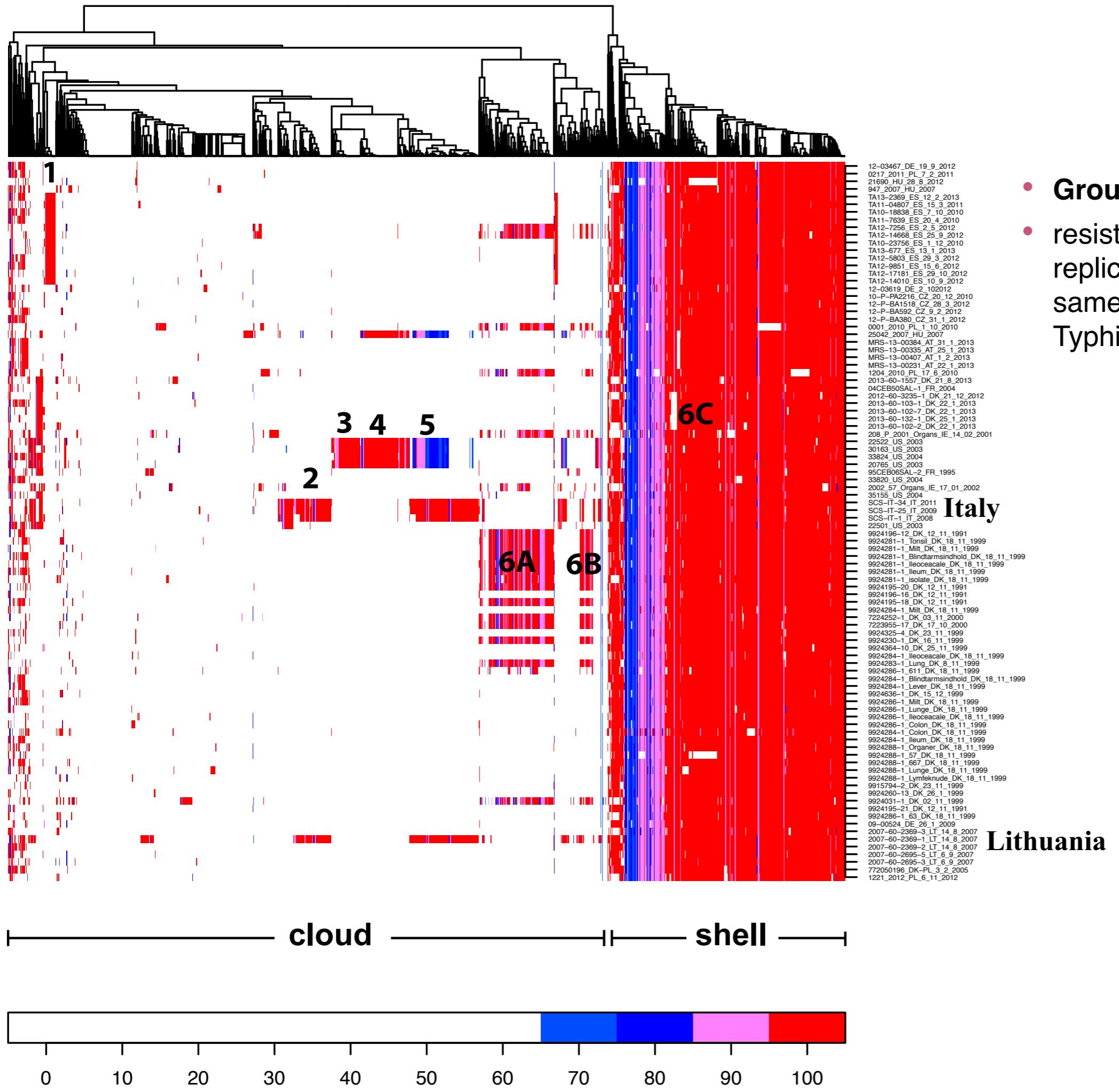


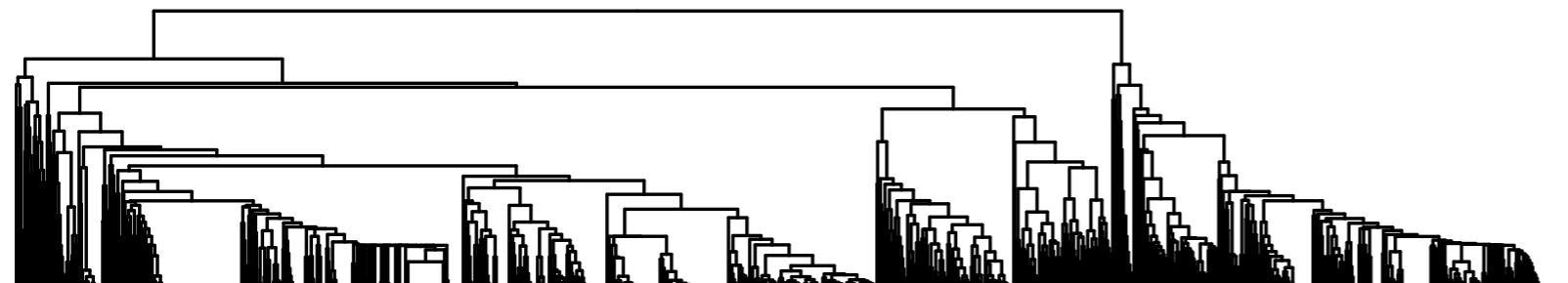
- **Group 2 (88 genes)**
- present in Italian strains and an isolate from Lithuania
- *S. Typhi* R27 plasmid
- *S. Typhimurium* plasmid R64
- *S. Choleraesuis* plasmid pSC138
- *S. Typhimurium* plasmid pSLT-BT
- *Serratia marcescens* plasmid R478



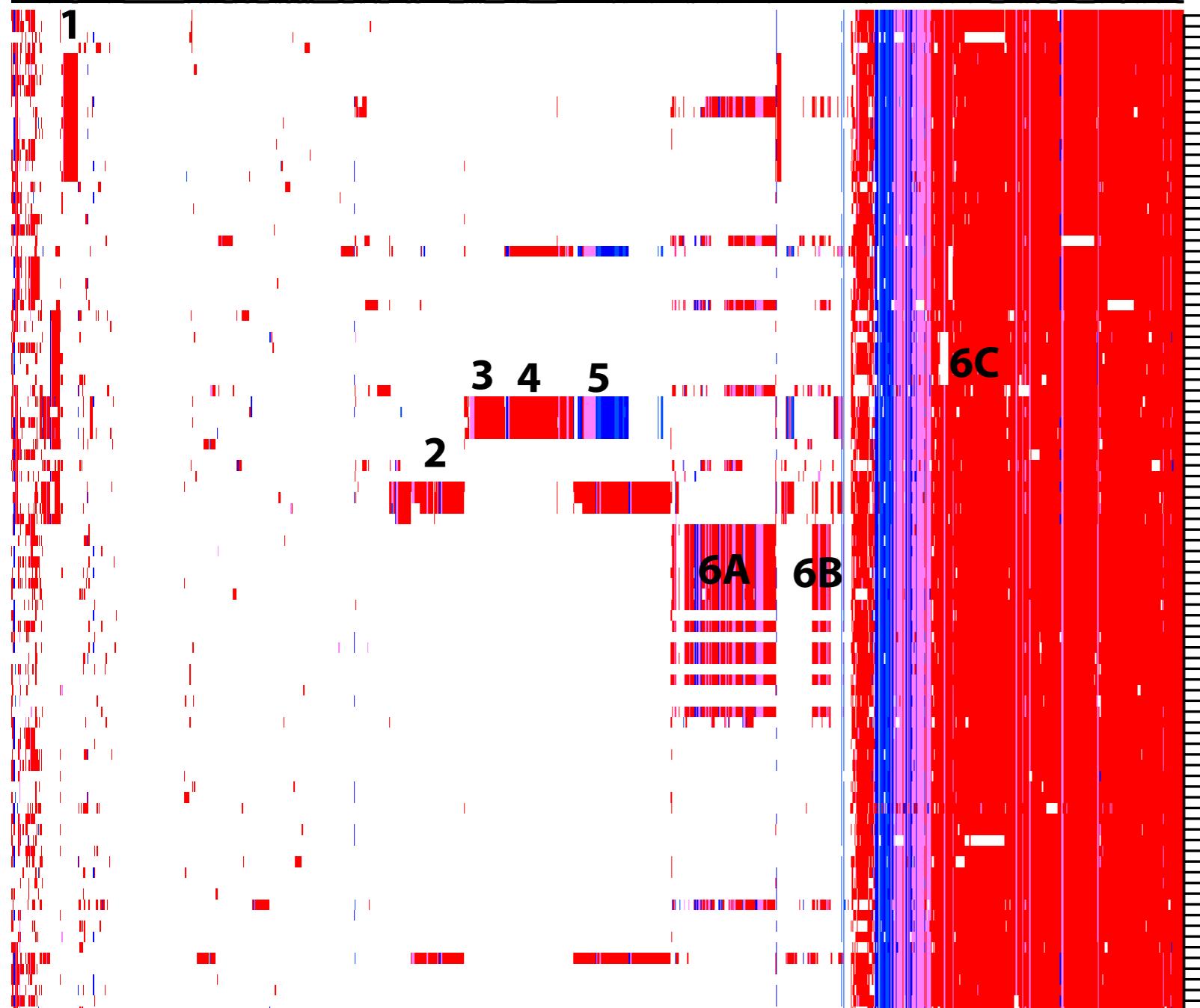
2428 SNPs

1.58-6 SNP/site/year 7.5 SNP/year

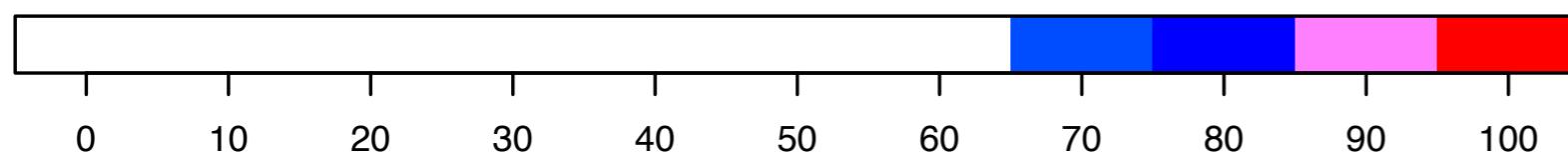


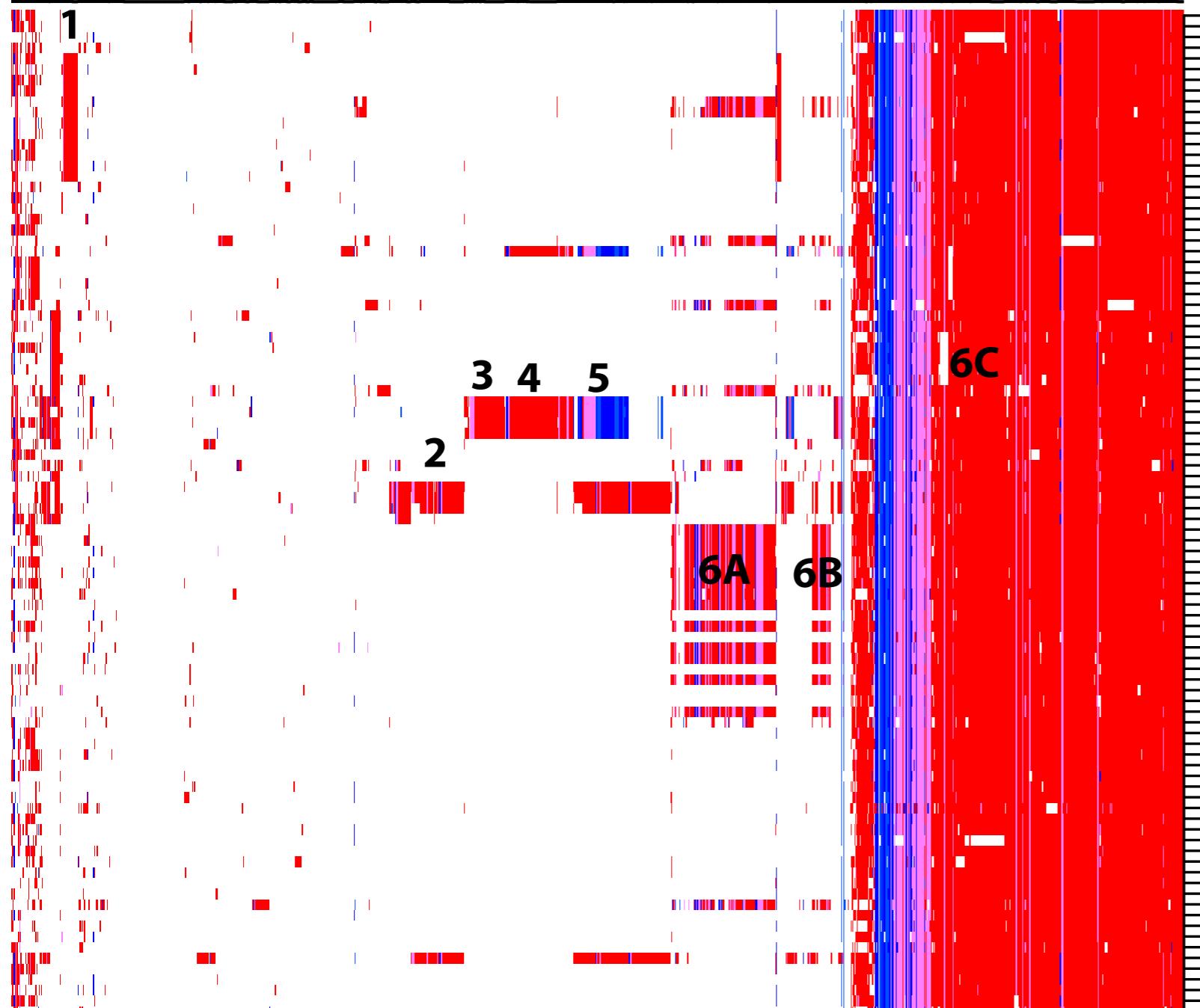
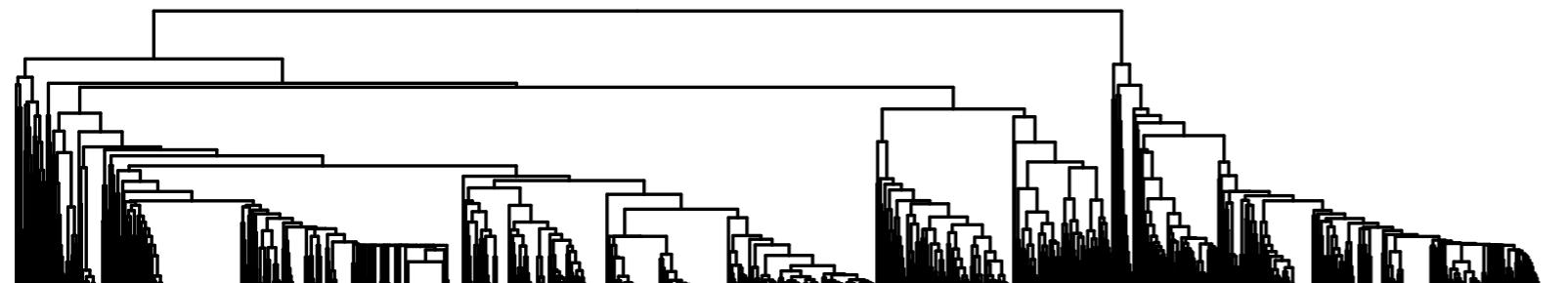


- **Group 3 (34 genes)**
- None plasmid related genes



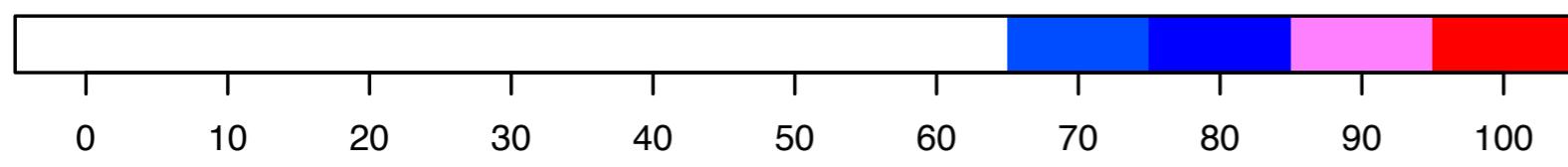
— cloud — shell —

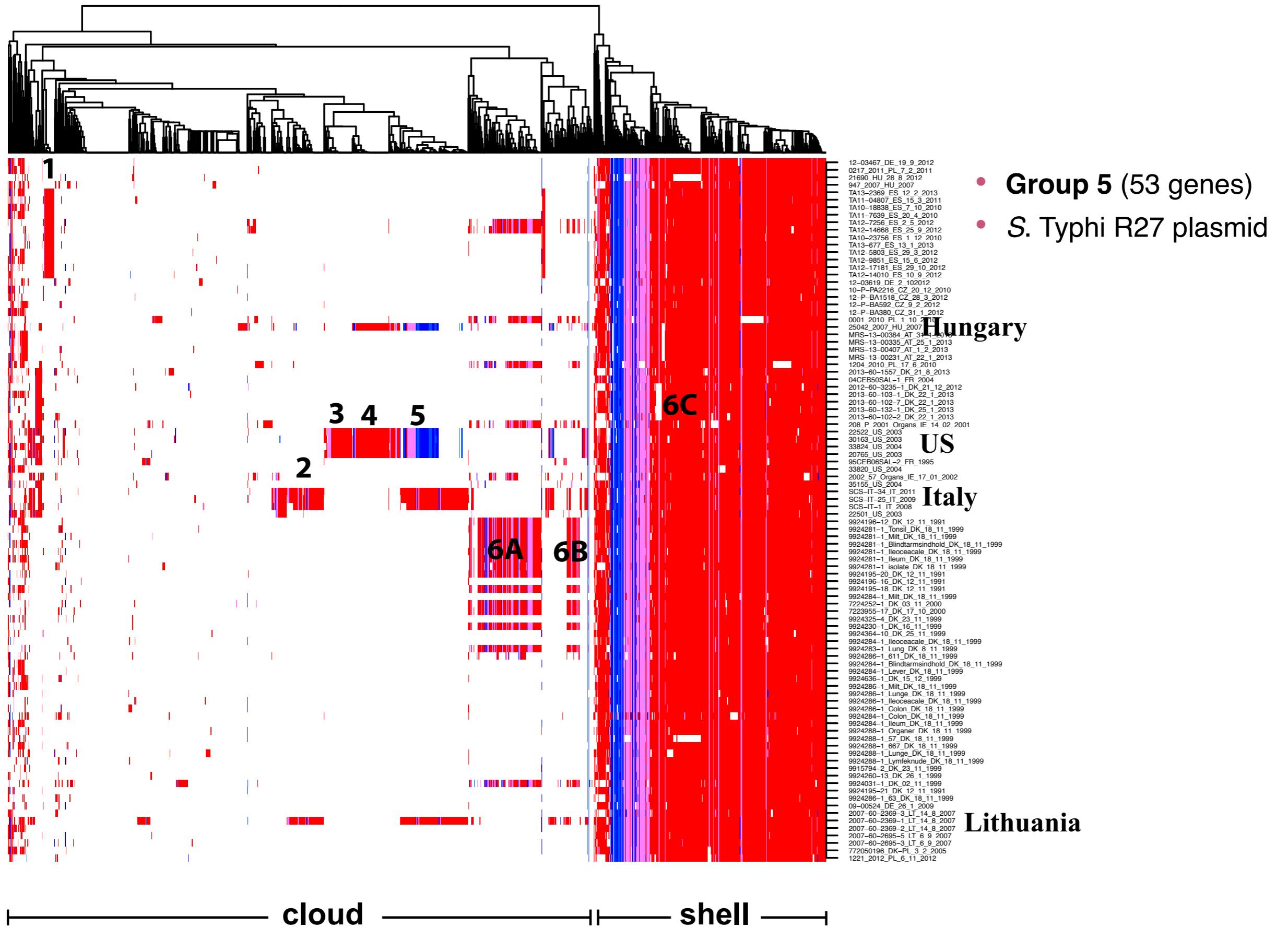


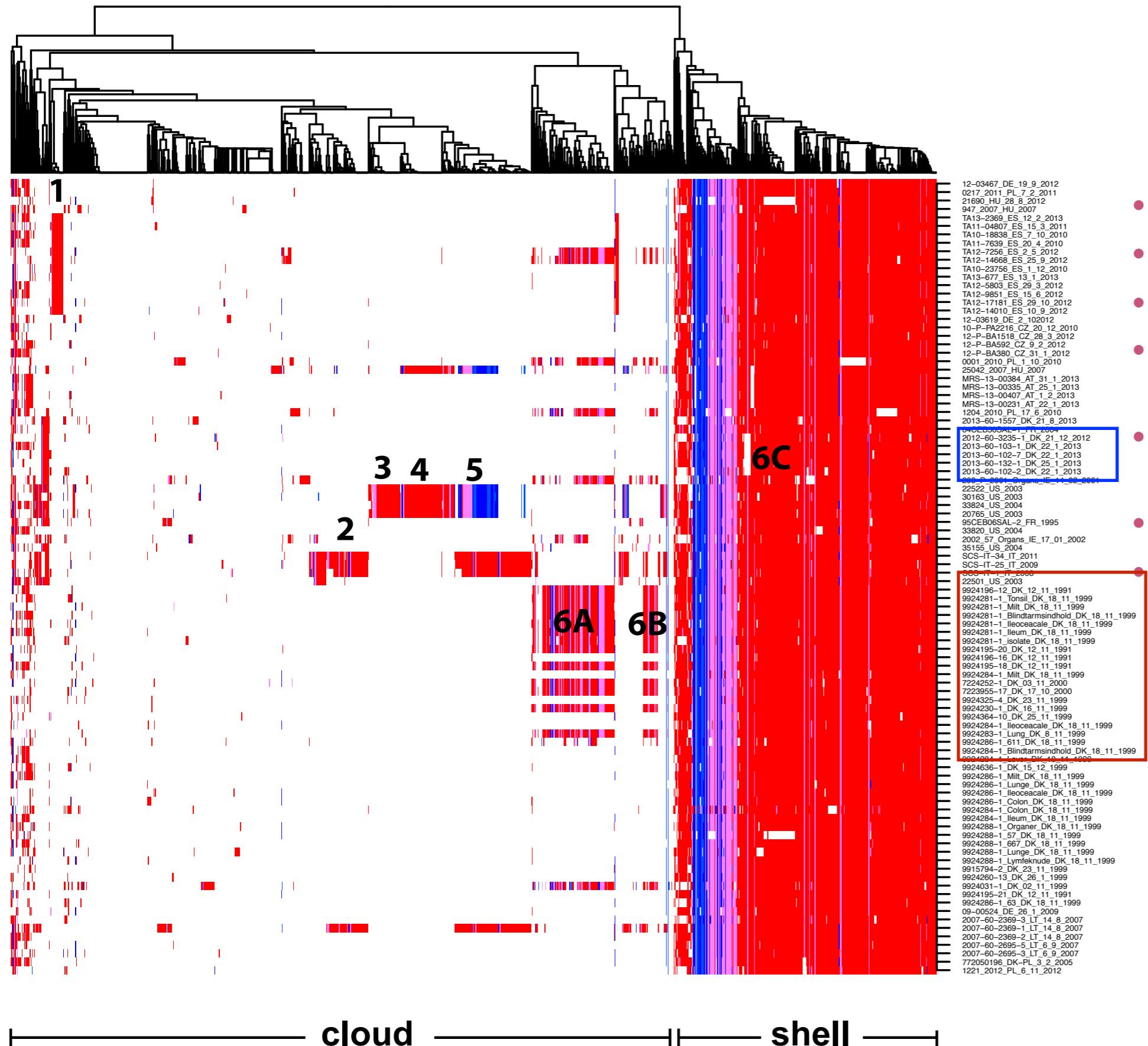


12-03467_DE_19_9_2012
 0217_2011_PL_7_2_2011
 21690_HU_28_8_2012
 947_2007_HU_2007
 TA13-2369_ES_12_2_2013
 TA11-04807_ES_15_3_2011
 TA10-18838_ES_7_10_2010
 TA11-7639_ES_20_4_2010
 TA12-7256_ES_2_5_2012
 TA12-14668_ES_25_9_2012
 TA10-23756_ES_1_12_2010
 TA13-677_ES_13_1_2013
 TA12-5803_ES_29_3_2012
 TA12-9851_ES_15_6_2012
 TA12-17181_ES_29_10_2012
 TA12-14010_ES_10_9_2012
 12-03619_DE_2_102012
 10-P-PA2216_CZ_20_12_2010
 12-P-BA1516_CZ_28_3_2012
 12-P-BA592_CZ_9_2_2012
 12-P-BA380_CZ_31_1_2012
 0001_2010_PL_1_10_2010
 25042_2007_HU_2007
 MRS-13-00384_AT_31_1_2013
 MRS-13-00335_AT_25_1_2013
 MRS-13-00407_AT_1_2_2013
 MRS-13-00231_AT_22_1_2013
 1204_2010_PL_17_6_2010
 2013-60-1557_DK_21_8_2013
 04CEB50SAL-1_FR_2004
 2012-60-3235-1_DK_21_12_2012
 2013-60-103-1_DK_22_1_2013
 2013-60-102-7_DK_22_1_2013
 2013-60-132-1_DK_25_1_2013
 2013-60-102-2_DK_22_1_2013
 208_P_2001_Organs_IE_14_02_2001
 2522_US_2003
 30163_US_2003
 33824_US_2004
 95CEB06SAL-2_FR_1995
 3820_US_2004
 2002_57_Organs_IE_17_01_2002
 35155_US_2004
 SCS-IT-34_IT_2011
 SCS-IT-25_IT_2009
 SCS-IT-1_IT_2008
 22501_US_2004
 9924196-12_DK_12_11_1991
 9924281-1_Ionsit_DK_18_11_1999
 9924281-1_Mitt_DK_18_11_1999
 9924281-1_Blindarmsmehold_DK_18_11_1999
 9924281-1_Ileoceacale_DK_18_11_1999
 9924281-1_Ileum_DK_18_11_1999
 9924281-1_isolate_DK_18_11_1999
 9924195-20_DK_12_11_1991
 9924196-16_DK_12_11_1991
 9924195-18_DK_12_11_1991
 9924284-1_Mitt_DK_18_11_1999
 7224252-1_DK_03_11_2000
 7223955-17_DK_17_10_2000
 9924325-4_DK_23_11_1999
 9924230-1_DK_16_11_1999
 9924364-10_DK_25_11_1999
 9924284-1_Ileoceacale_DK_18_11_1999
 9924283-1_Lunge_DK_8_11_1999
 9924286-1_611_DK_18_11_1999
 9924284-1_Blindarmsmehold_DK_18_11_1999
 9924284-1_Lever_DK_18_11_1999
 9924636-1_DK_15_12_1999
 9924286-1_Mitt_DK_18_11_1999
 9924286-1_Lunge_DK_18_11_1999
 9924286-1_Ileoceacale_DK_18_11_1999
 9924286-1_Colon_DK_18_11_1999
 9924284-1_Colon_DK_18_11_1999
 9924284-1_Ileum_DK_18_11_1999
 9924288-1_Organer_DK_18_11_1999
 9924288-1_57_DK_18_11_1999
 9924288-1_667_DK_18_11_1999
 9924288-1_Lunge_DK_18_11_1999
 9924288-1_Lymfeknude_DK_18_11_1999
 9915794-2_DK_23_11_1999
 9924260-13_DK_26_1_1999
 9924031-1_DK_02_11_1999
 9924195-21_DK_12_11_1991
 9924286-1_63_DK_18_11_1999
 09-00524_DE_26_1_2009
 2007-60-2369-3_LT_14_8_2007
 2007-60-2369-1_LT_14_8_2007
 2007-60-2369-2_LT_14_8_2007
 2007-60-2695-5_LT_6_9_2007
 2007-60-2695-3_LT_6_9_2007
 772050196_DK-PL_3_2_2005
 1221_2012_PL_6_11_2012

— cloud — shell —



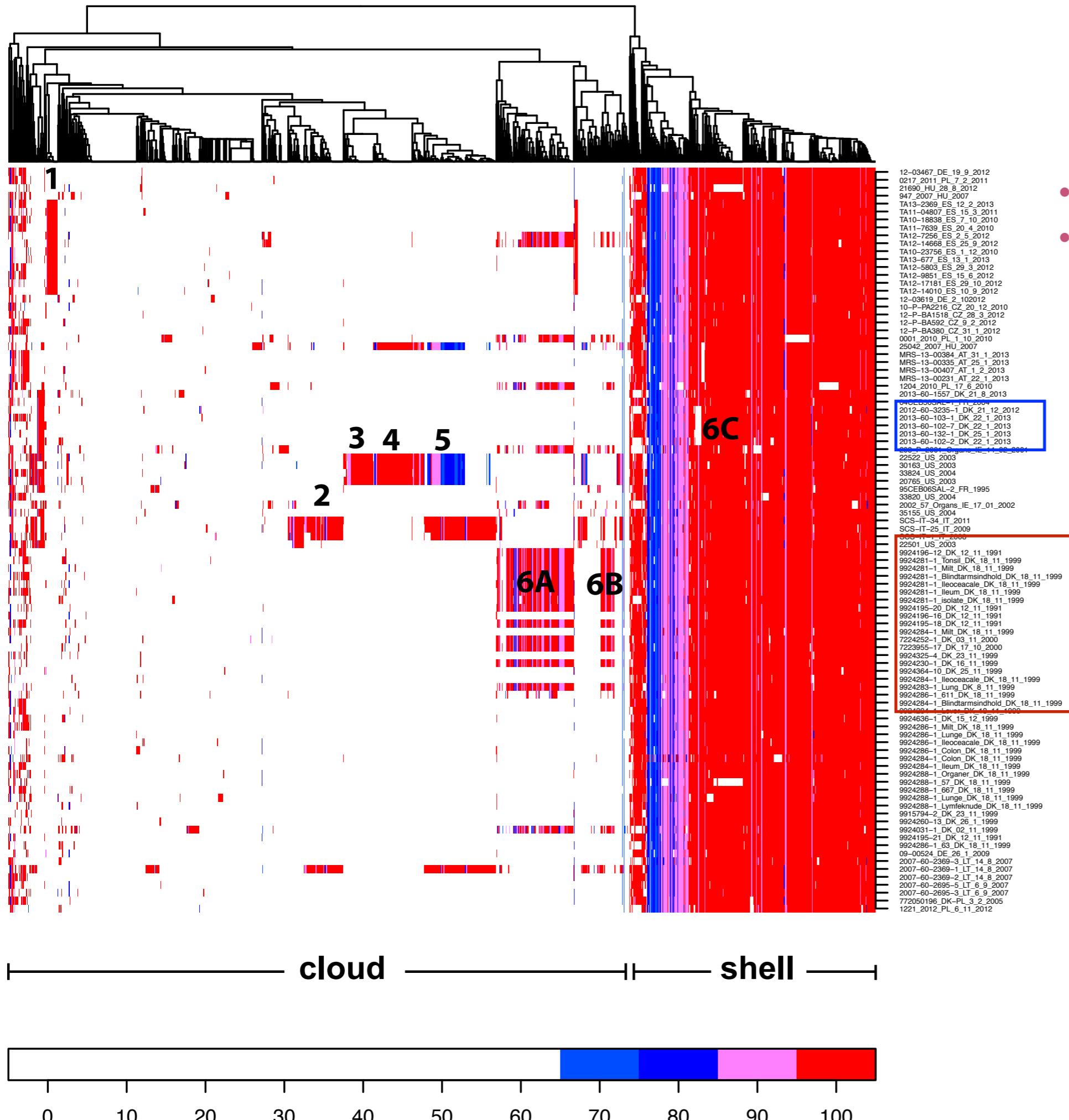


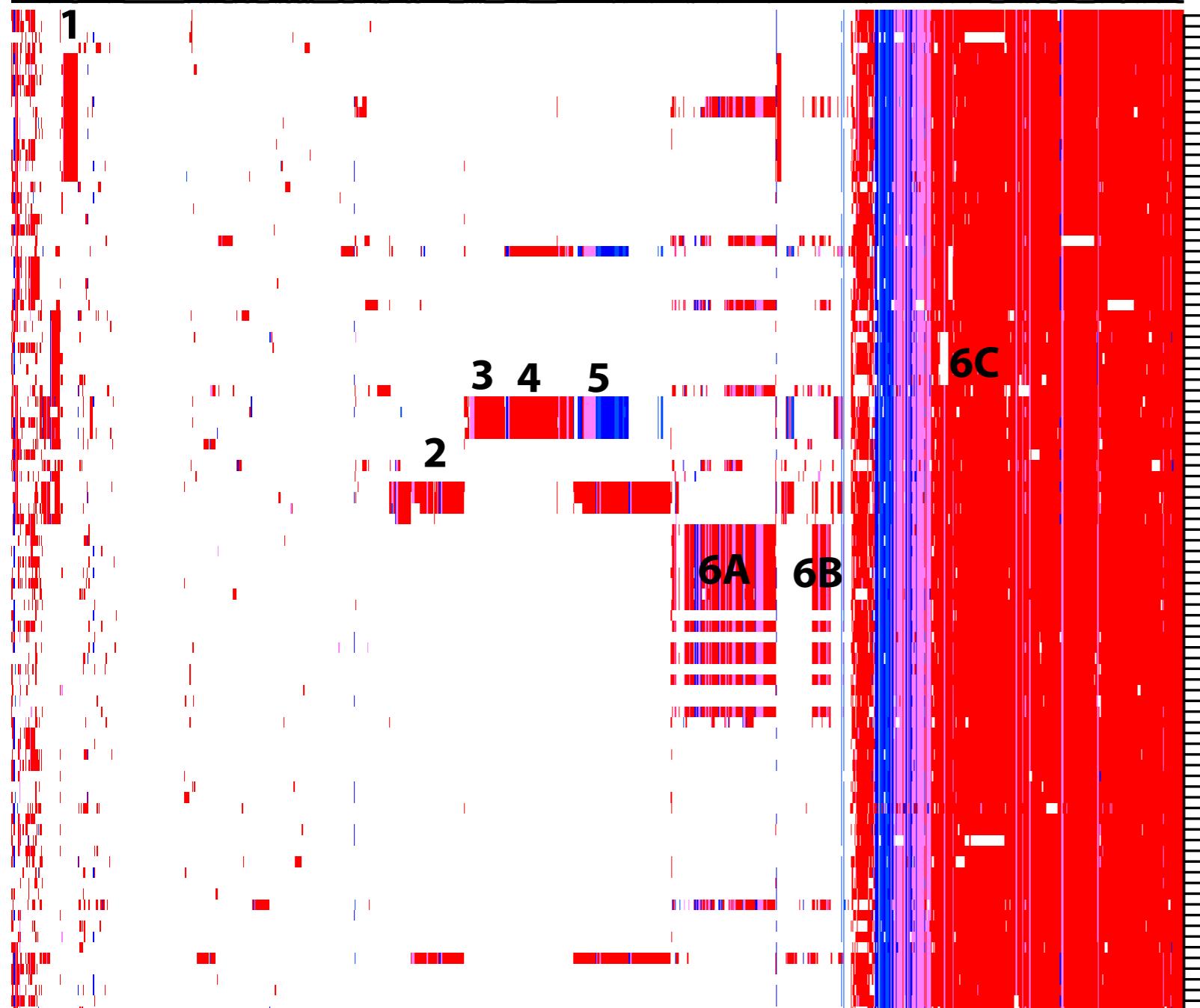
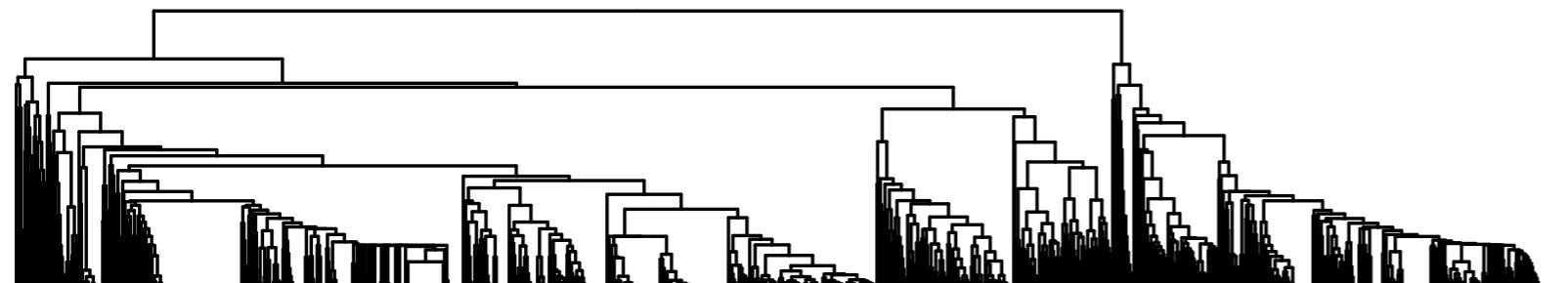


- **Group 6A** (109 genes)
- **Group 6B** (25 genes)
- *S. Typhimurium* plasmid R64
- *S. Choleraesuis* plasmid pSC138
- *Serratia marcescens* plasmid R478
- *S. Typhi* R27 plasmid

- *S. Typhi* R27 plasmid

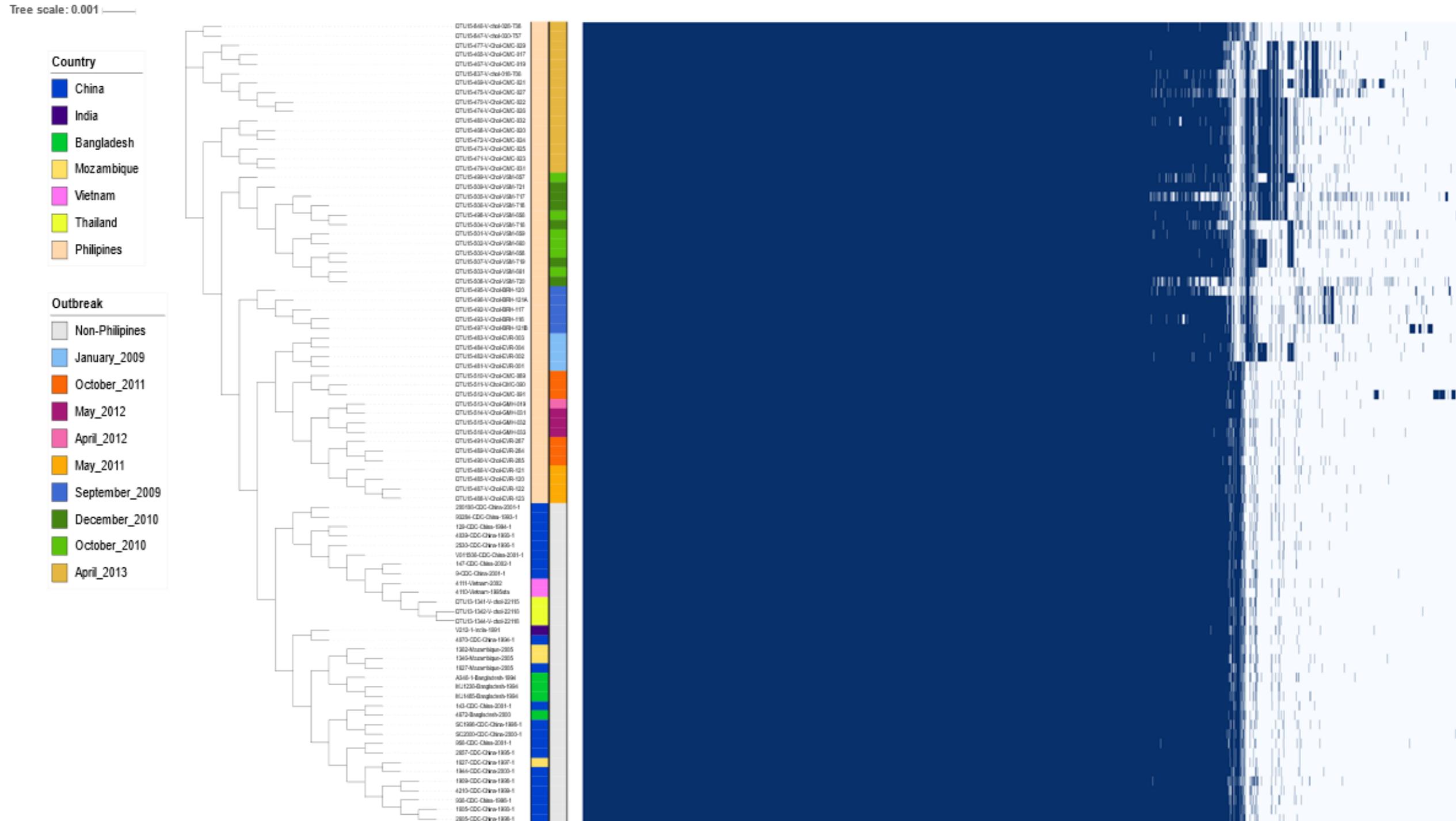
- cloud —————— shell ——————





- The presence/absence of the accessories genes contribute to the phylogeny
- Plasmids might be one of important roles in virulence of *S. Choleraesuis*

Presence/absence of genes (Vibrio)



Pan-genome analysis

