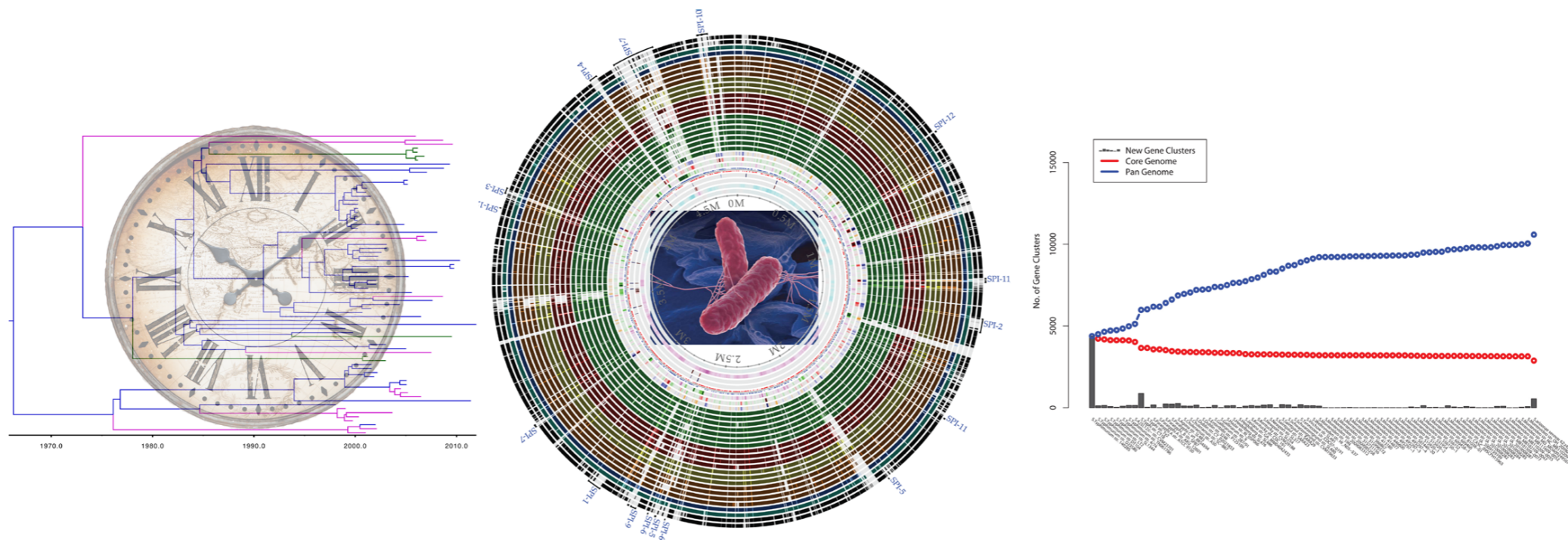


# Genomic Epidemiology



Pimlapas Leekitcharoenphon (Shinny)  
Research Group for Genomic Epidemiologies  
National Food Institute (DTU Food)

[pile@food.dtu.dk](mailto:pile@food.dtu.dk)



@ShinnyPimlapas



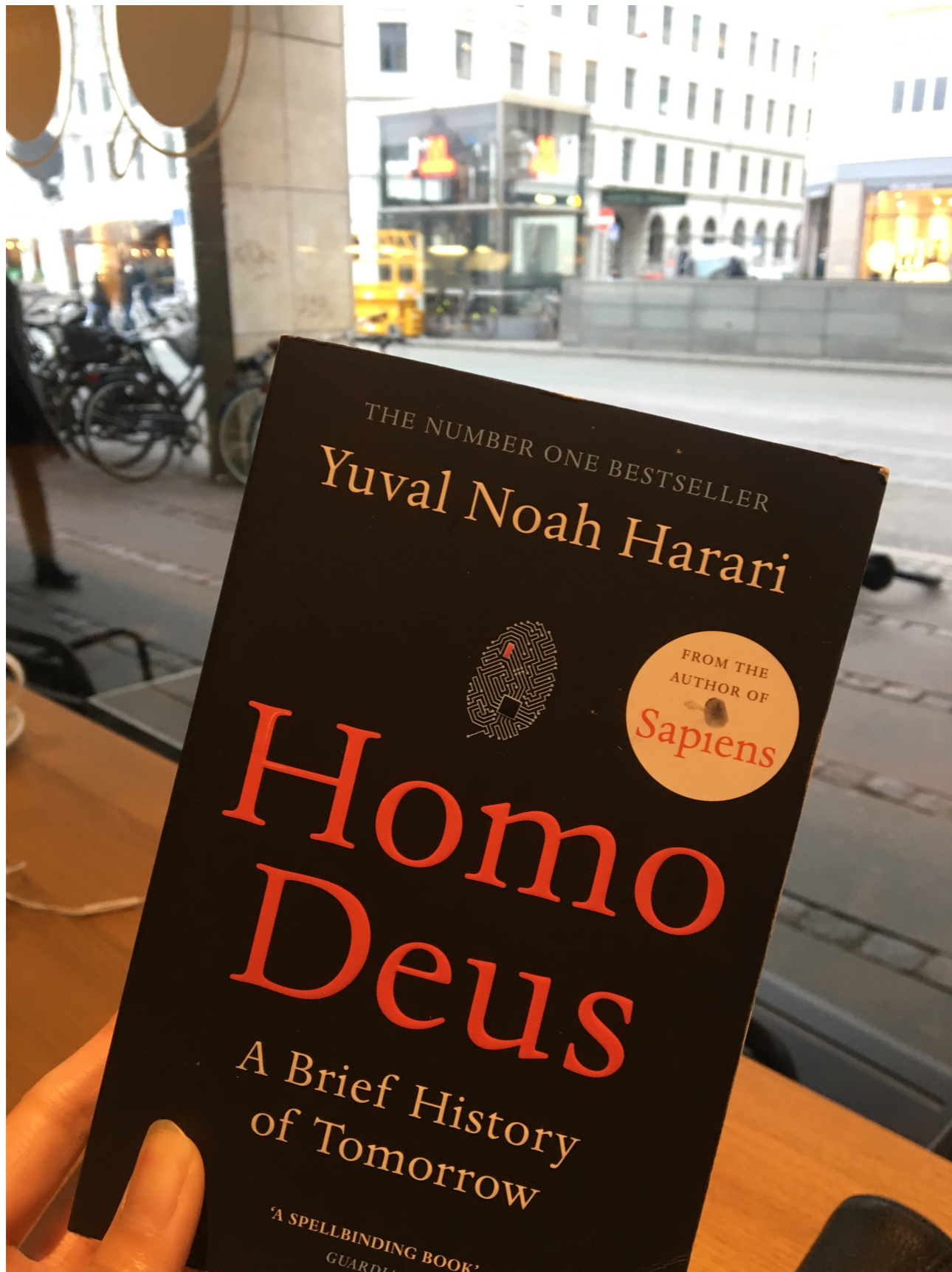
10th January 2023

# About myself

- Biotechnology and Bioinformatics background
- WGS analysis of food borne pathogens and metagenomics for surveillance of AMR
- Senior Researcher at DTU Food
- DTU course
  - 23205 Fighting Infectious Diseases
  - 23262 Infectious Disease Bioinformatics
- Online courses on COURSERA on AMR, WGS and Metagenomics

# Topics

- Epidemiology
- Application of WGS in routine typing and surveillance of infectious diseases
- Genomic epidemiology for global surveillance AMR



- Famine
- War
- Infectious diseases

# Epidemiology

- The science that studies the patterns, causes, and effects of health and disease conditions in defined populations
- Questions;
  - What is it ?
  - Has it been seen before ?
  - How can we fight it ?
  - Is it an outbreak ?

# Identification and Typing

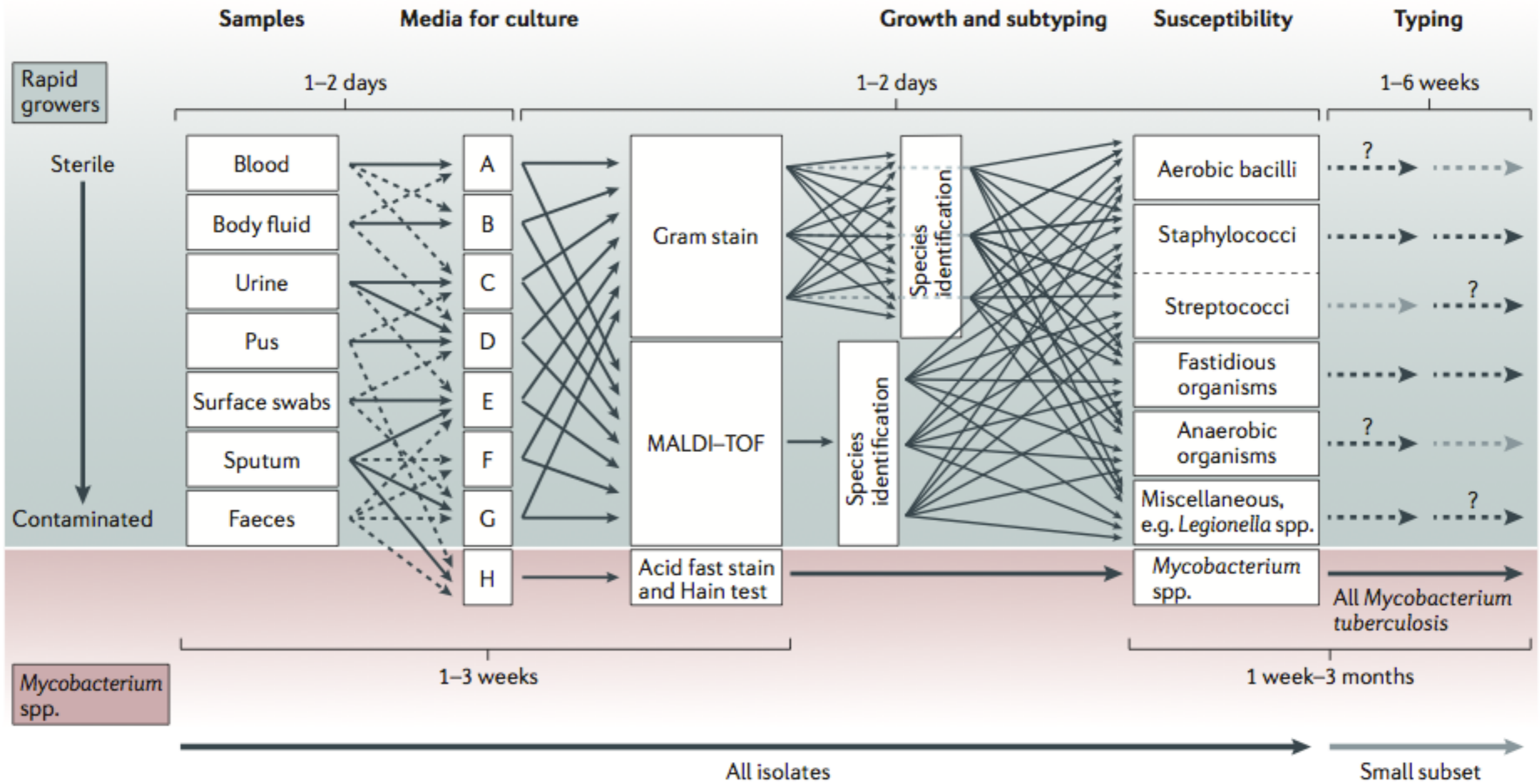
- Any characterization below the (sub-) species level is termed “typing”
- Methods used for this characterization are per definition “typing methods”

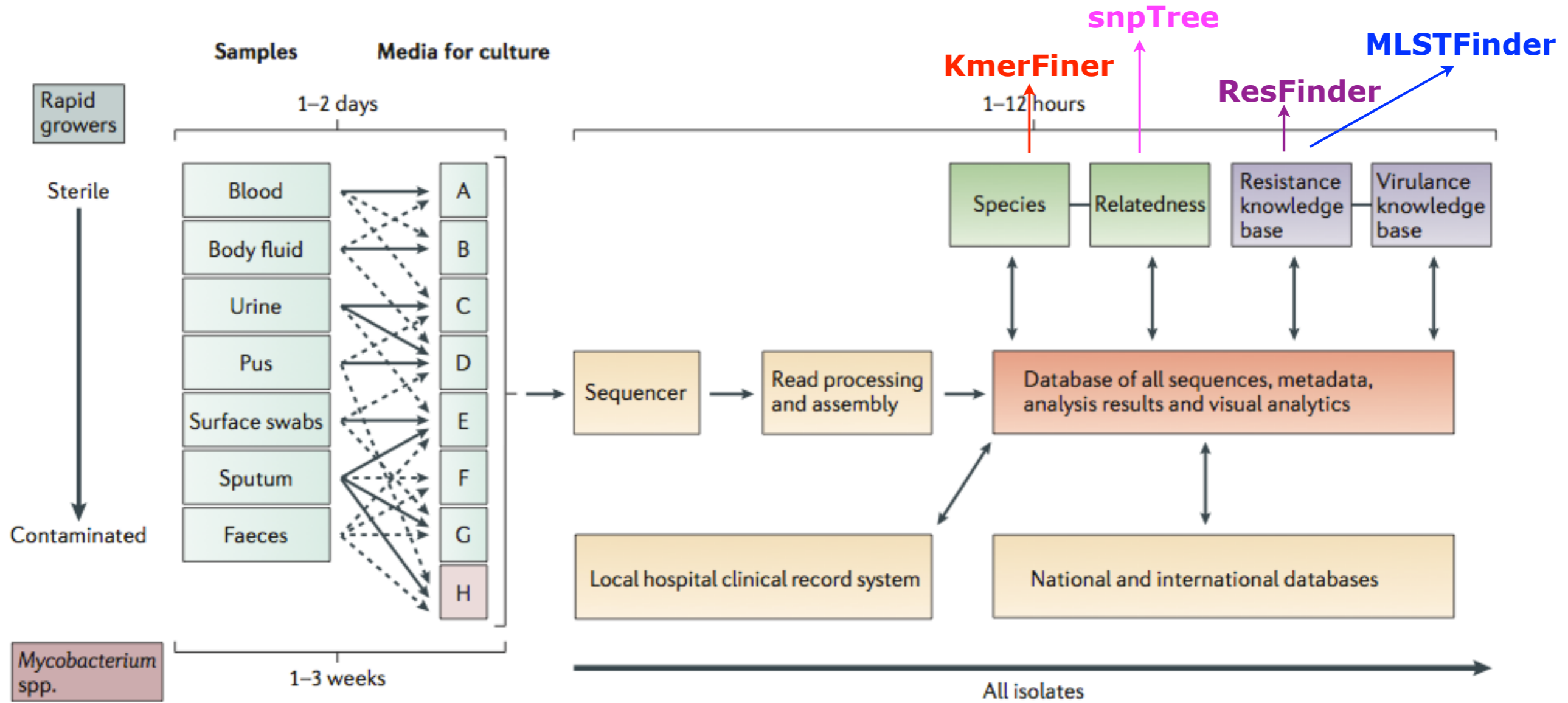
Family  
Genus  
Species  
(Subspecies)

**Identification**

Serovar  
Phagetype  
Ribotype  
PFGE type  
MLVA type  
MLST type  
DNA Microarray analysis  
Whole genomic sequence

**Typing**







# Epidemiology

- The science that studies the patterns, causes, and effects of health and disease conditions in defined populations
- Questions;
  - **What is it ?**
  - Has it been seen before ?
  - How can we fight it ?
  - Is it an outbreak ?

# Species Identification

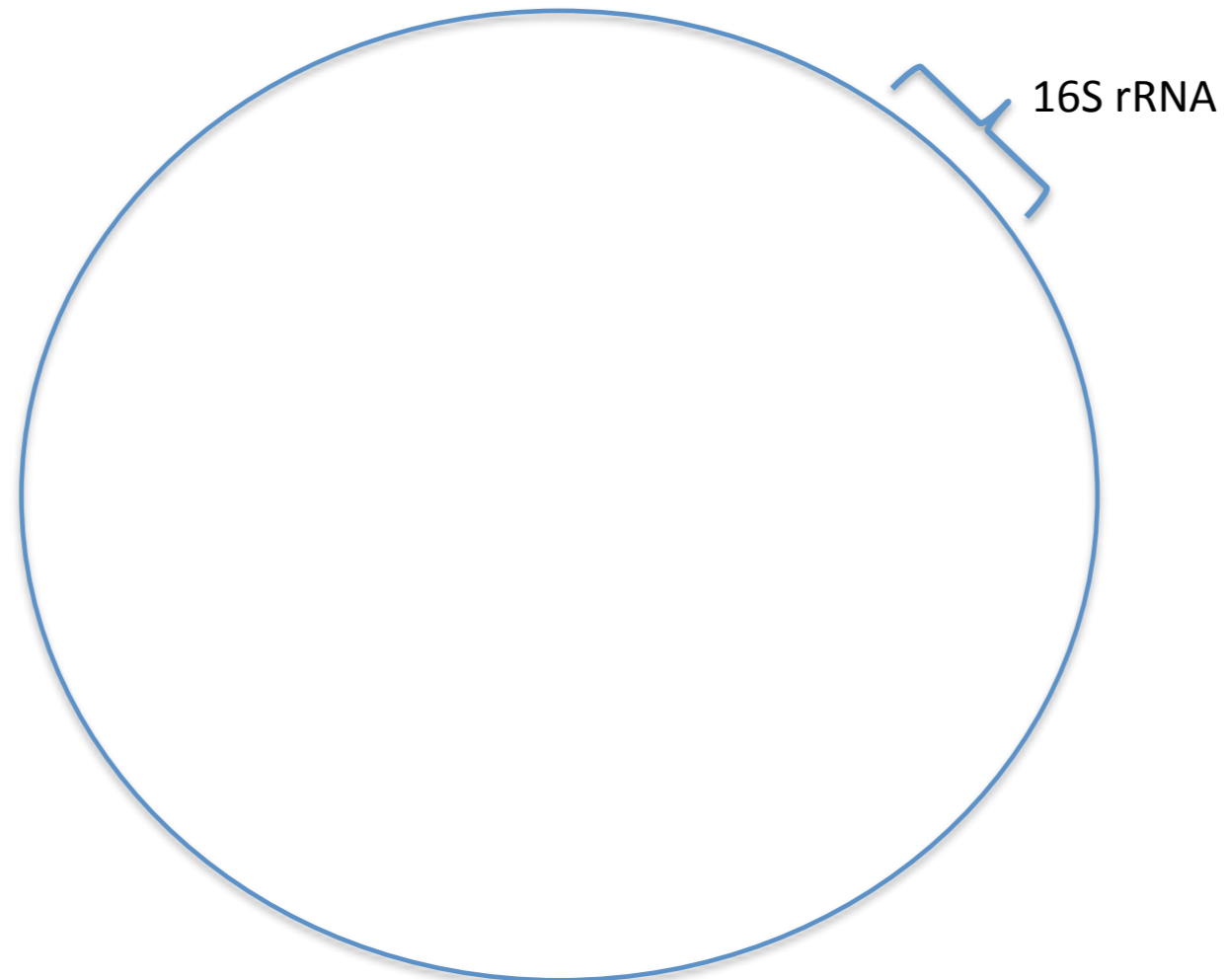
# CGE implementation of 16S rRNA species identification - SpeciesFinder

## Reference database

- 16S rRNA genes are isolated from genomes in NCBI

Sequence	Isolate in ref. db.	Species
ACGCCG.....CACG	CP32523	<i>K. pneumonia</i>
GATGAG....CGGG	CP64333	<i>E. coli</i>
TGAGGT...TGTTT	CP11212	<i>S. aureus</i>
TGAGGT...TTTTT	CP87878	<i>S. aureus</i>
AAATAG...TGTTT	CP11122	<i>S. enterica</i>
TATAAA....AAAA	CP12121	<i>L. lactis</i>
GATGAG....CGGG	CP86533	<i>E. coli</i>
GTTTAG....CGGG	CP12333	<i>E. coli</i>
GTATTA....AAAA	CP99888	<i>S. pyogenes</i>

The 16s rRNA gene represents only a small fraction of the entire genome



# K-mer ?

- A k-mer is a contiguous sequence of k bases
- k is any positive integer
- Sequences with high similarity must share k-mers

sequence

**ATGGAAGTCGCGGAATC**

7 mers

ATGGAAG  
TGGAAAGT  
GGAAGTC  
GAAGTCG  
AAGTCGC  
AGTCGCG  
GTCGCGG  
TCGCGGA  
CGCGGAA  
GCGGAAT  
CGGAATC

# Species identification by K-mer

Known species **ATGGAAGTCGCGGAATC**

**ATGGAAGTCGCGGAATC** Unknown species

k-mers

k-mers

ATGGAAG  
TGGAAGT  
GGAAGTC  
GAAGTCG  
AAGTCGC  
AGTCGCG  
GTCGCGG  
TCGCGGA  
CGCGGAA  
GCGGAAT  
CGGAATC

ATGGAAG  
TGGAAGT  
GGAAGTC  
GAAGTCG  
AAGTCGC  
AGTCGCG  
GTCGCGG  
TCGCGGA  
CGCGGAA  
GCGGAAT  
CGGAATC



Species

## Training data (database)

- ✧ 1,647 completed / almost completed genomes downloaded from NCBI in 2011 (1,009 different species)

# Evaluating the methods

## Evaluation data (testing data)

### ✧ NCBI draft genomes

- 695 isolates from species that overlap with training set (151 species)

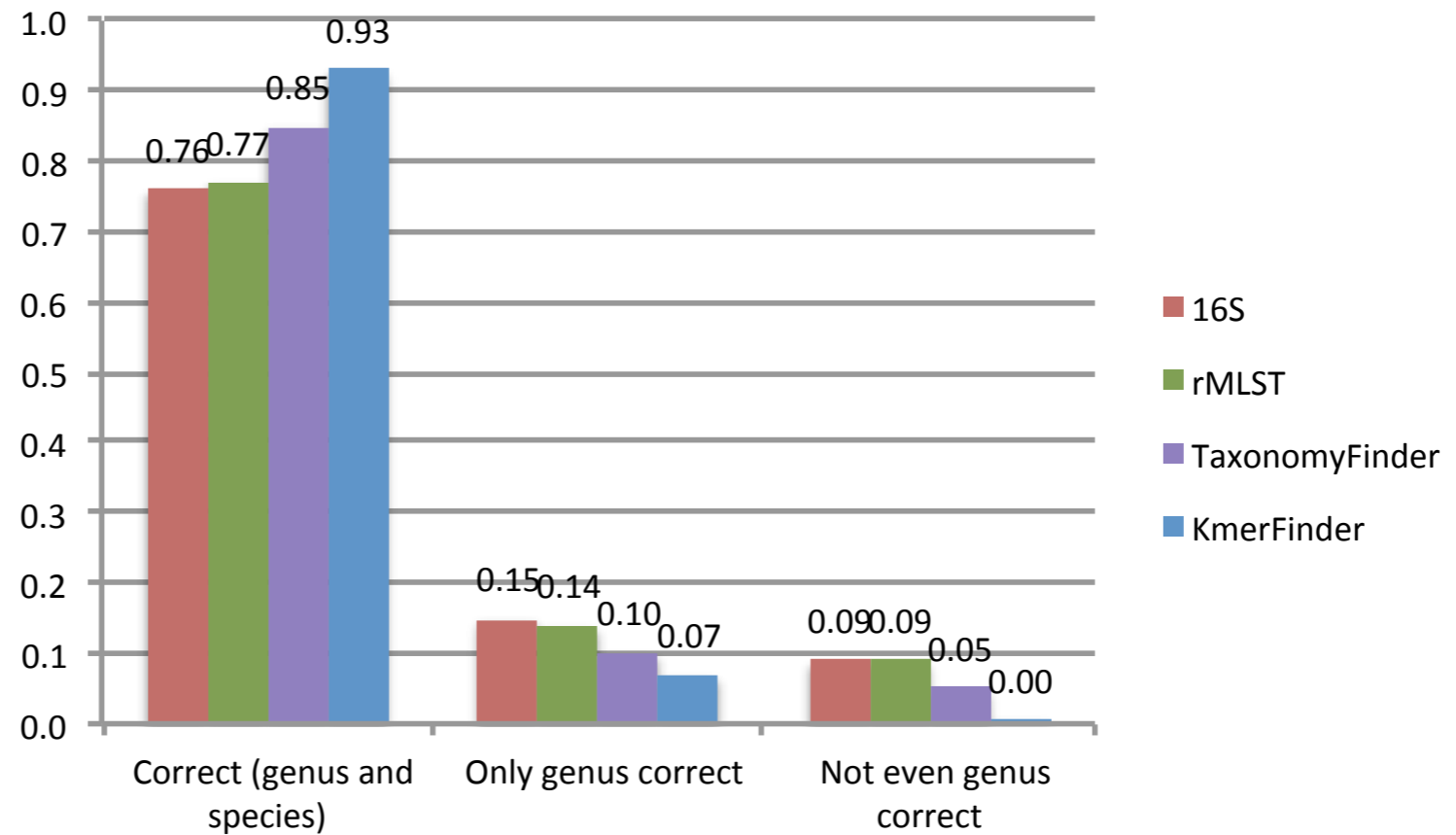
### ✧ SRA draft genomes

- 10,407 draft genomes from Illumina data (168 species)

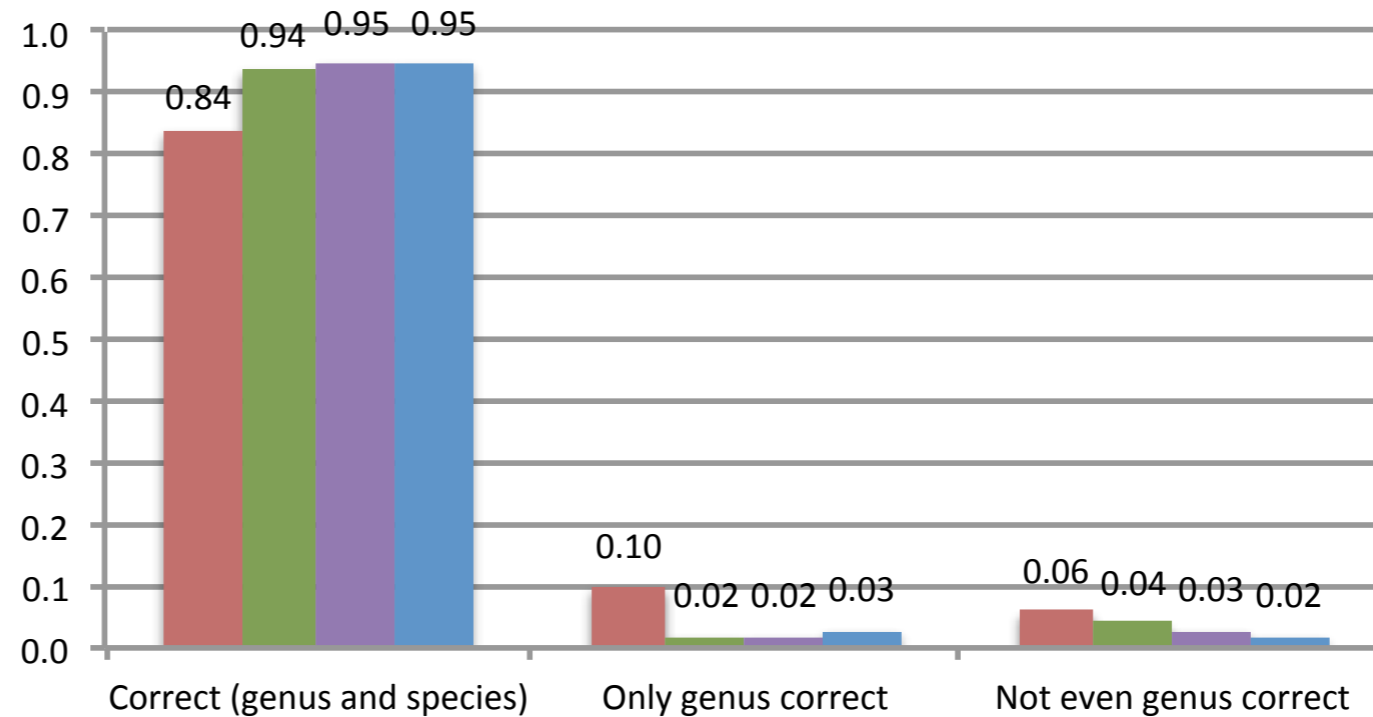


# Performance

## NCBI draft genomes



## SRA draft genomes

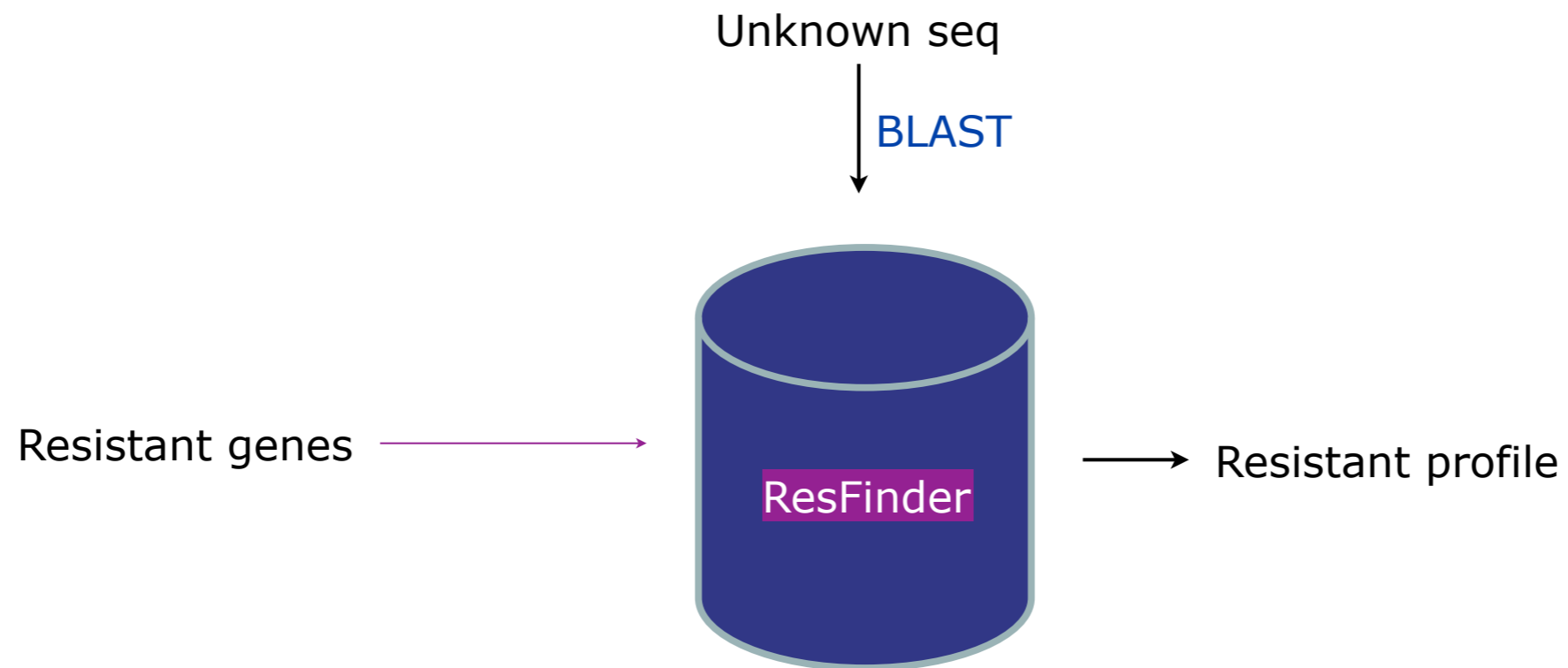


# Epidemiology

- The science that studies the patterns, causes, and effects of health and disease conditions in defined populations
- Questions;
  - What is it ?
  - Has it been seen before ?
  - **How can we fight it ?**
  - Is it an outbreak ?

# ResFinder

# Resistant finding



# ResFinder

- ResFinder is based on curated database, public databases as well as on scientific papers
- The ResFinder is a web-friendly interface and freely accessible tool
- ResFinder will detect the presence of resistance genes and point mutation causing resistance in WGS data (raw reads or assembled genomes)
- High concordance (99.74%) between phenotypic and predicted antimicrobial susceptibility was observed

# Epidemiology

- The science that studies the patterns, causes, and effects of health and disease conditions in defined populations
- Questions;
  - What is it ?
  - **Has it been seen before ?**
  - How can we fight it ?
  - **Is it an outbreak ?**

# What is phylogeny used for

- Classify taxonomy – The classic use
- Outbreak detection – Increasing with WGS data

# What is phylogeny used for

- Cholera outbreak in Haiti 2010
- Listeria outbreak 2014

Whole-genome Sequencing Used to Investigate a Nationwide Outbreak of Listeriosis Caused by Ready-to-eat Delicatessen Meat, Denmark, 2014.

Kvistholm Jensen et al. Clin Infect Dis. (2016) 63 (1): 64-70. doi: 10.1093/cid/ciw192



## Case story

- *Vibrio Cholerae* outbreak in Haiti followed the 2010 earthquake
- Rumors said that the outbreak may have come from Nepal, travelling along with UN soldiers from Nepal
- No proof had been given of this until the Hendriksen *et al.* paper in 2011

Population Genetics of *Vibrio cholerae* from Nepal in 2010: Evidence on the Origin of the Haitian Outbreak. Hendriksen et al. 23 August 2011 mBio vol. 2 no. 4 e00157-11. doi: 10.1128/mBio.00157-11

## Case story

- Data
  - 24 recent *V. cholerae* strains from Nepal
  - 10 previously sequenced *V. cholerae* isolates, including 3 from the Haitian outbreak
- Analysis
  - Antimicrobial susceptibility testing
  - PFGE (pulsed-field gel electrophoresis) to analyze for genetic relatedness
  - Whole genome sequencing, SNP identification and phylogenetic analysis

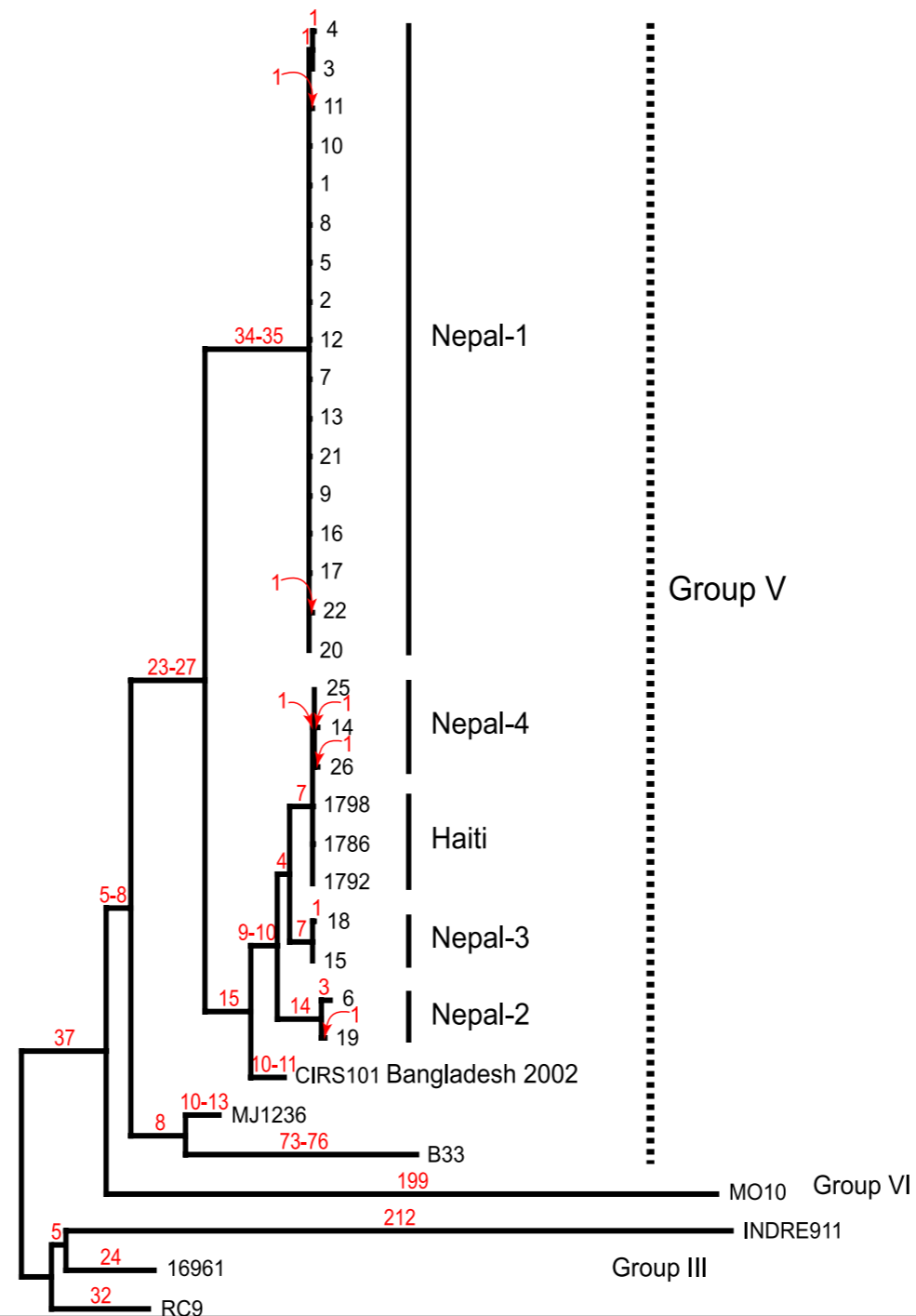
# Case story - Results

Resistance profile	Susceptible	Decreased susceptibility	Resistant
Nepalese strains <i>Hendriksen et al. 2011</i>	Tetracycline	Ciprofloxacin	Trimethoprim, Sulfamethoxazole Nalidixic
Haitian outbreak strains Centers for Disease Control and Prevention, 2010	Tetracycline	Ciprofloxacin	Trimethoprim, Sulfamethoxazole Nalidixic

## Case story - Results

- Pulsed-field gel electrophoresis (PFGE)
  - Nepalese isolates divided in 4 groups
  - Most common Haitian type in same group as four Nepalese strains

# Case story - Results



# SNPs detection

...ATCGAATTCCGGGTTTTTAACCGGATCGTACGATCGGGGAAAAA..

TTCCAGG

TTCCAGG

TTCCAGG

TTCCAGG

TTCCAGG

TTCCAGG

# SNPs detection



# Variant calling format (VCF)

Genome 1	position	ref	change
Ref_genome	10	T	C
Ref_genome	20	C	T
Ref_genome	30	A	C
Ref_genome	40	A	C
Ref_genome	50	G	A

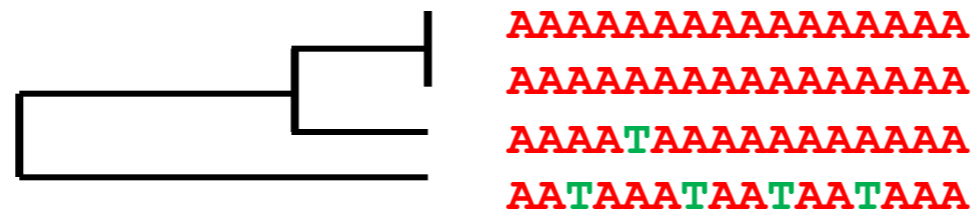
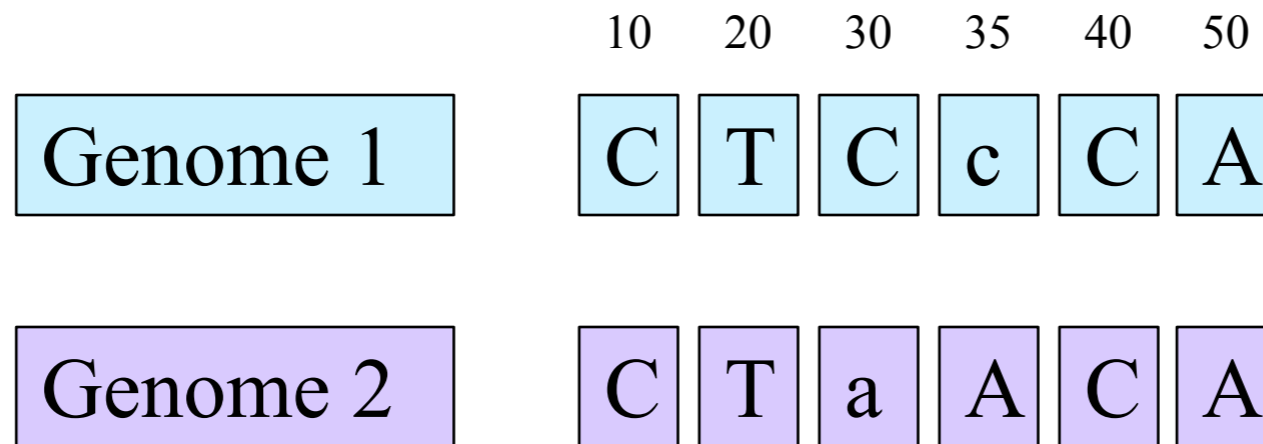
Genome 2	position	ref	change
Ref_genome	10	T	C
Ref_genome	20	C	T
Ref_genome	35	C	A
Ref_genome	40	A	C
Ref_genome	50	G	A



# Concatenated SNPs

Genome 1	position	ref	change
Ref_genome	10	T	C
Ref_genome	20	C	T
Ref_genome	30	A	C
Ref_genome	40	A	C
Ref_genome	50	G	A

Genome 2	position	ref	change
Ref_genome	10	T	C
Ref_genome	20	C	T
Ref_genome	35	C	A
Ref_genome	40	A	C
Ref_genome	50	G	A



# Nextstrain

## Real-time tracking of pathogen evolution

Nextstrain is an open-source project to harness the scientific and public health potential of pathogen genome data. We provide a continually-updated view of publicly available data alongside powerful analytic and visualization tools for use by the community. Our goal is to aid epidemiological understanding and improve outbreak response. If you have any questions, or simply want to say hi, please give us a shout at [hello@nextstrain.org](mailto:hello@nextstrain.org).

READ MORE

## SARS-CoV-2 (COVID-19)

We are incorporating SARS-CoV-2 genomes as soon as they are shared and providing analyses and situation reports. In addition we have developed a number of resources and tools, and are facilitating independent groups to run their own analyses. Please see the [SARS-CoV-2 resources page](#) for more information.

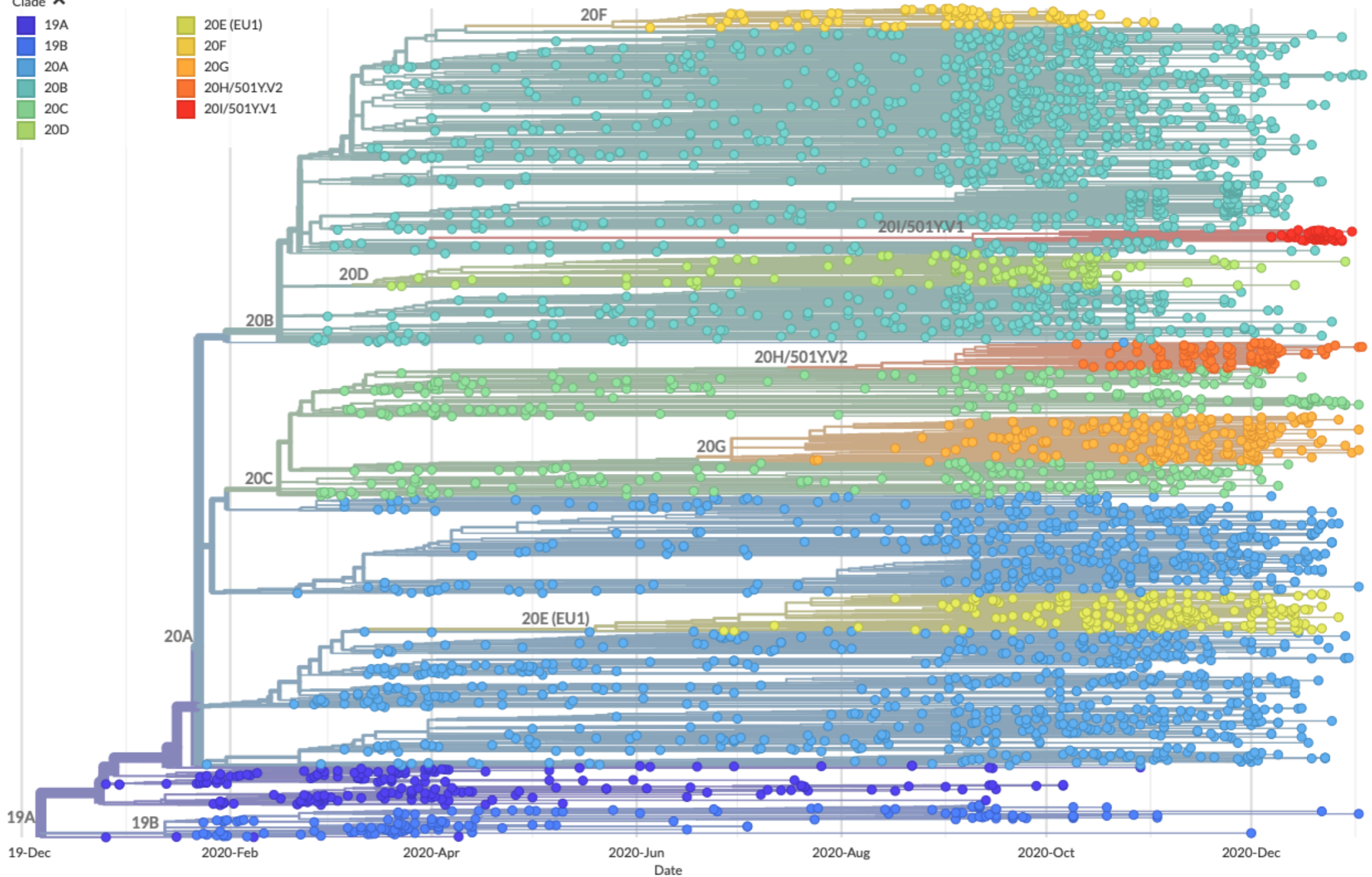
# Genomic epidemiology of novel coronavirus - Global subsampling

Maintained by the [Nextstrain](#) team. Enabled by data from [GISAID](#)  
Showing 3917 of 3917 genomes sampled between Dec 2019 and Jan 2021.

## Phylogeny

- Clade ^
- 19A
  - 19B
  - 20A
  - 20B
  - 20C
  - 20D

- 20E (EU1)
- 20F
- 20G
- 20H/501Y.V2
- 20I/501Y.V1



ZOOM TO SELECTED RESET LAYOUT

### Dataset

ncov  
global

### Date Range

2019-12-03 2021-01-07

### Color By

Clade

### Filter Data

Type filter query here...

### Tree Options

Layout  
 RECTANGULAR  
 RADIAL  
 UNROOTED  
 CLOCK

### Branch Length

TIME  DIVERGENCE

Show confidence intervals

### Branch Labels

clade

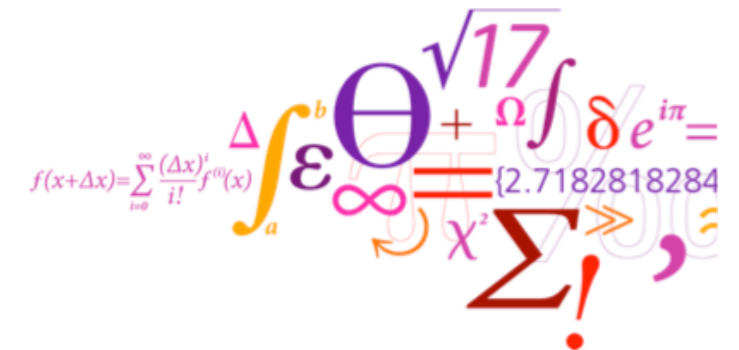
### Tip Labels

Sample Name

### Second Tree

Select...

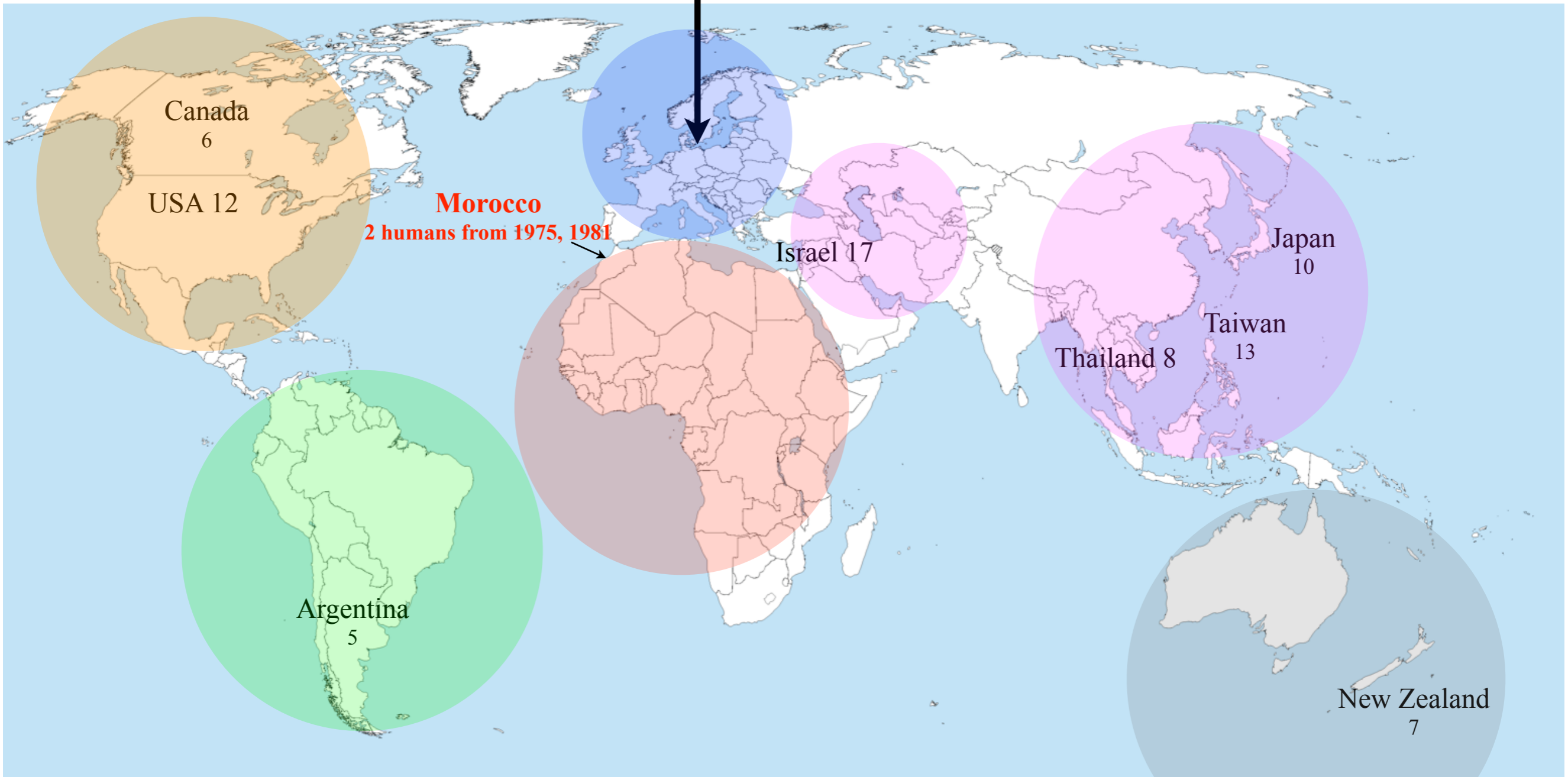
# Genomic epidemiology of the global occurrence *S. Typhimurium* DT104



## **S. Typhimurium DT104**

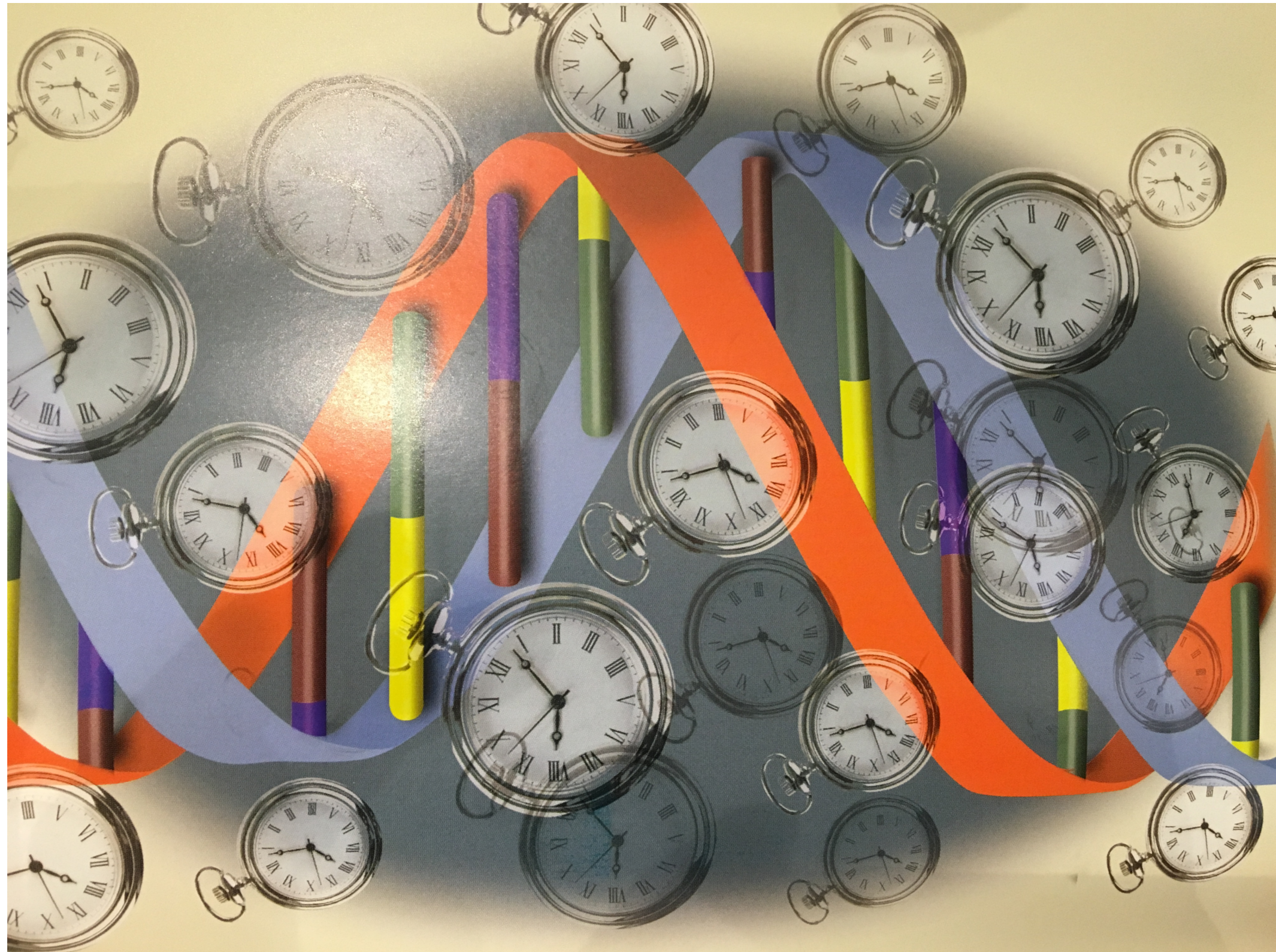
- During the last three decades, *S. Typhimurium* phage type DT104 emerged as the most important phage type and one of the best-studied because of its rapid global dissemination [Lan R, *et al.* Infect Genet Evol. 2009] [Helms M, *et al.* Emerg Infect Dis. 2005]
- DT104 has a multiple antimicrobial resistance pattern to ampicillin, chloramphenicol, streptomycin, sulphonamide and tetracycline (ACSSuT) [Mulvey MR, *et al.* Microbes Infect. 2006]
- Previous epidemics with MDR phage types of *S. Typhimurium*, such as DTs 29, 204, 193 and 204c, were mostly restricted to cattle [Threlfall EJ. J Antimicrob Chemother. 2000]
- DT104 spread among all domestic animals including cattle, poultry, pigs and sheep [Threlfall EJ. J Antimicrob Chemother. 2000]

Ireland 10	Scotland 14	Denmark 79
Spain	Austria 30	Netherlands 22
1 human from 1976	Germany 27	
France 9	Luxembourg 13	Switzerland 8
1 human from 1969	Poland 13	Czech Republic 9

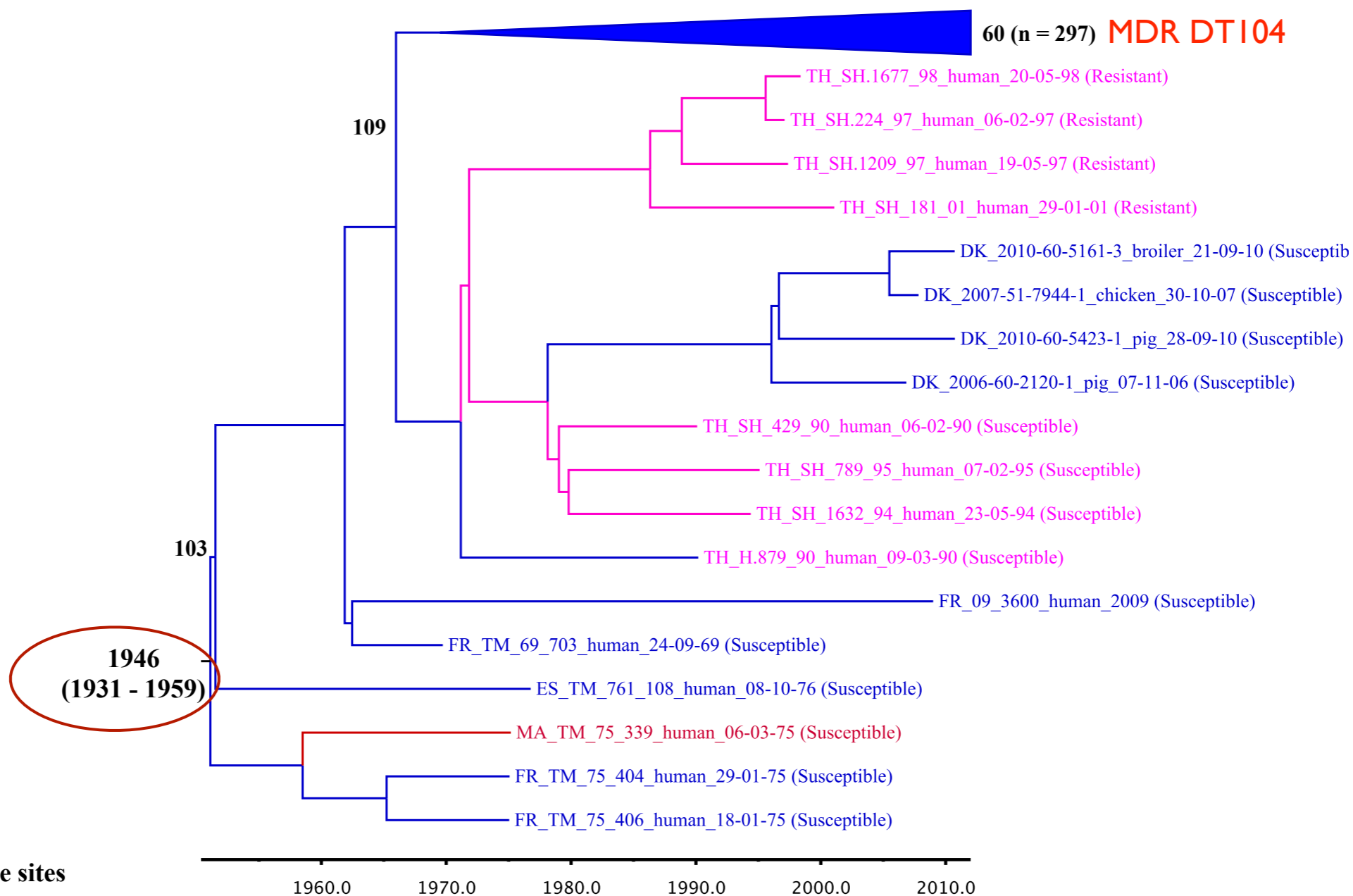


**315 genomes**  
*197 animal isolates*  
*118 human isolates*

# BEAST



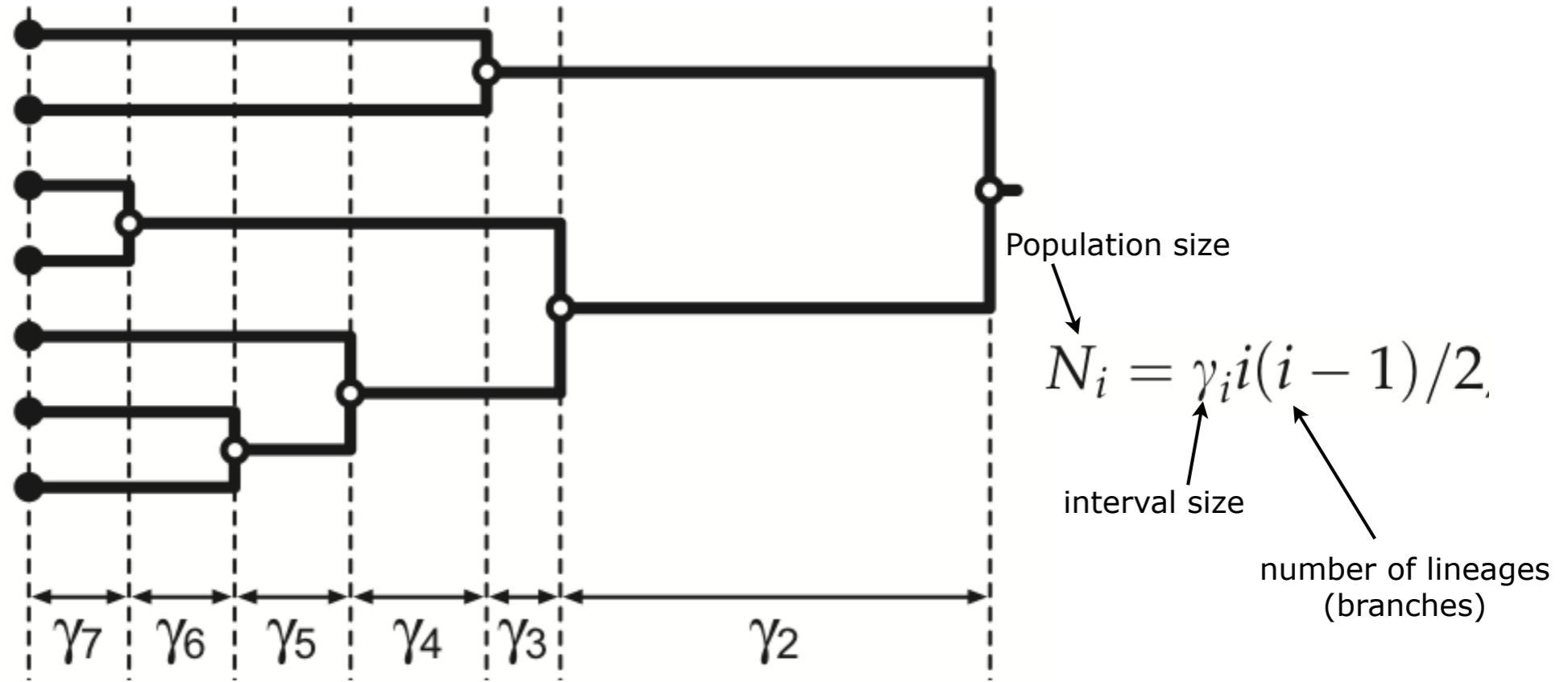
- Europe
- North America
- South America
- Asia
- Australia

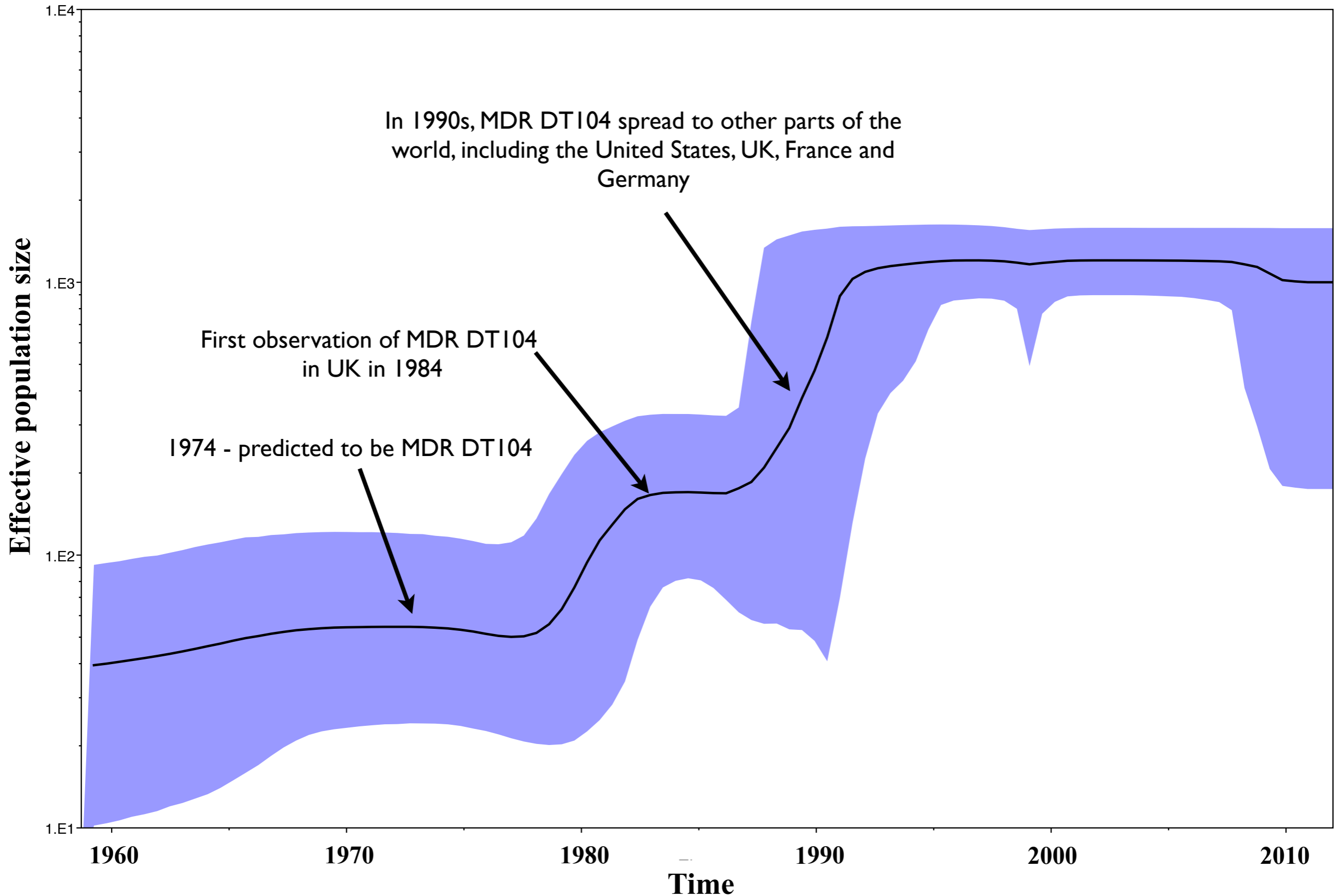


4619 variable sites

$2.97 \times 10^{-7}$  SNP/site/year ~ 1.47 SNP/year







Search

Get Directions History

our Buildings ed es able\_fasta\_danish...

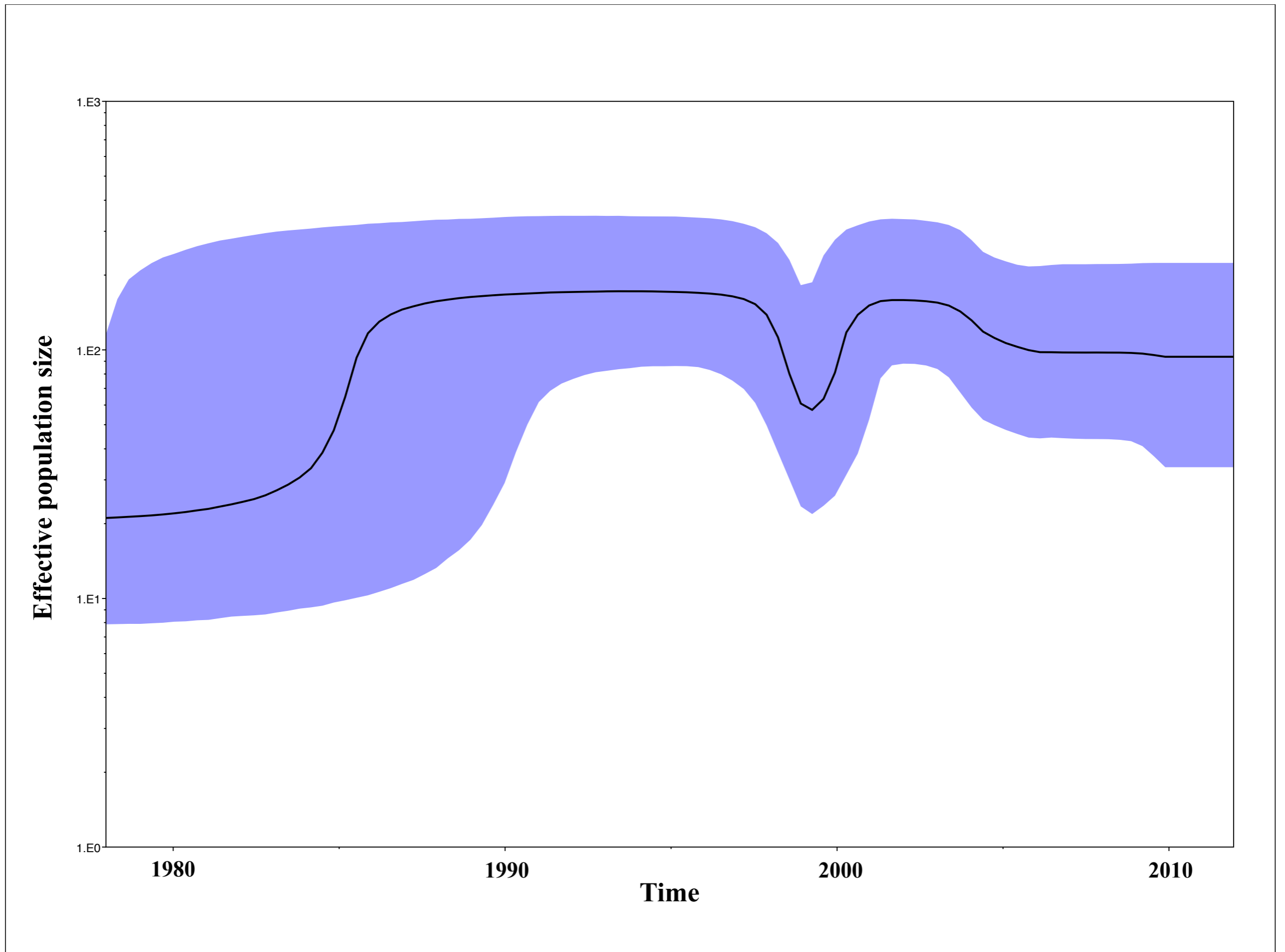
3/1960

1960 2011



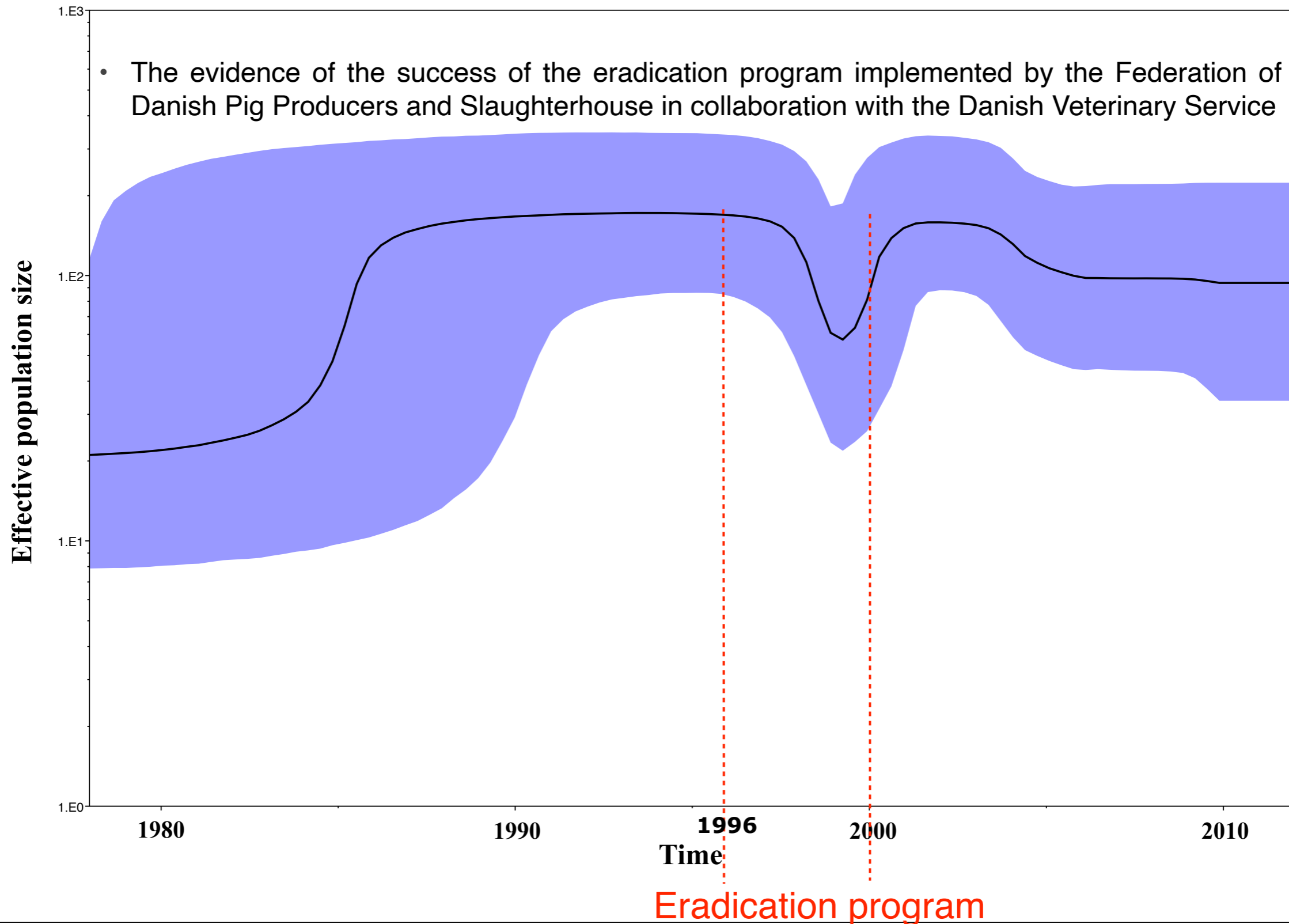
© 2009 GeoBasis-DE/BKG  
 Data SIO, NOAA, U.S. Navy, NGA, GEBCO  
 © 2013 Google  
 Image Landsat

# Demographic history of Danish MDR DT104

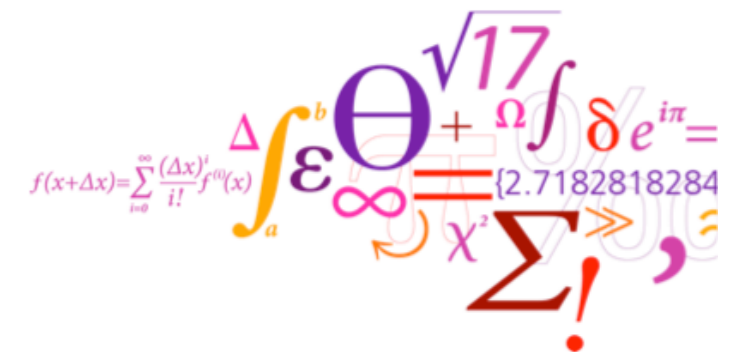


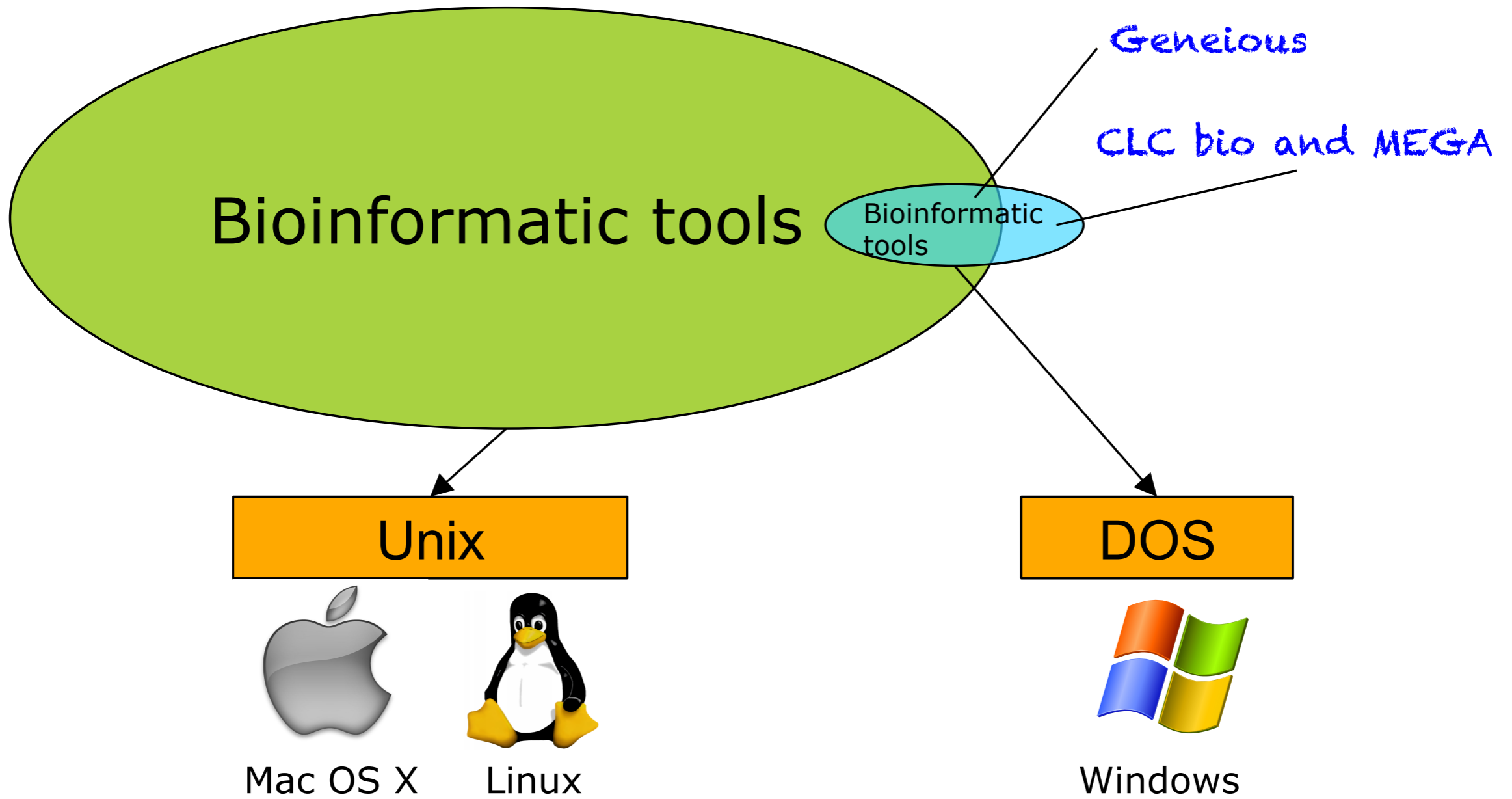
- The program aimed to eradicate MDR DT104 from infected pig herds by depopulation of pig herds, cleaning and disinfection of building before repopulation with pigs free from DT104

- 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



# Application of WGS in routine typing and surveillance of infectious diseases



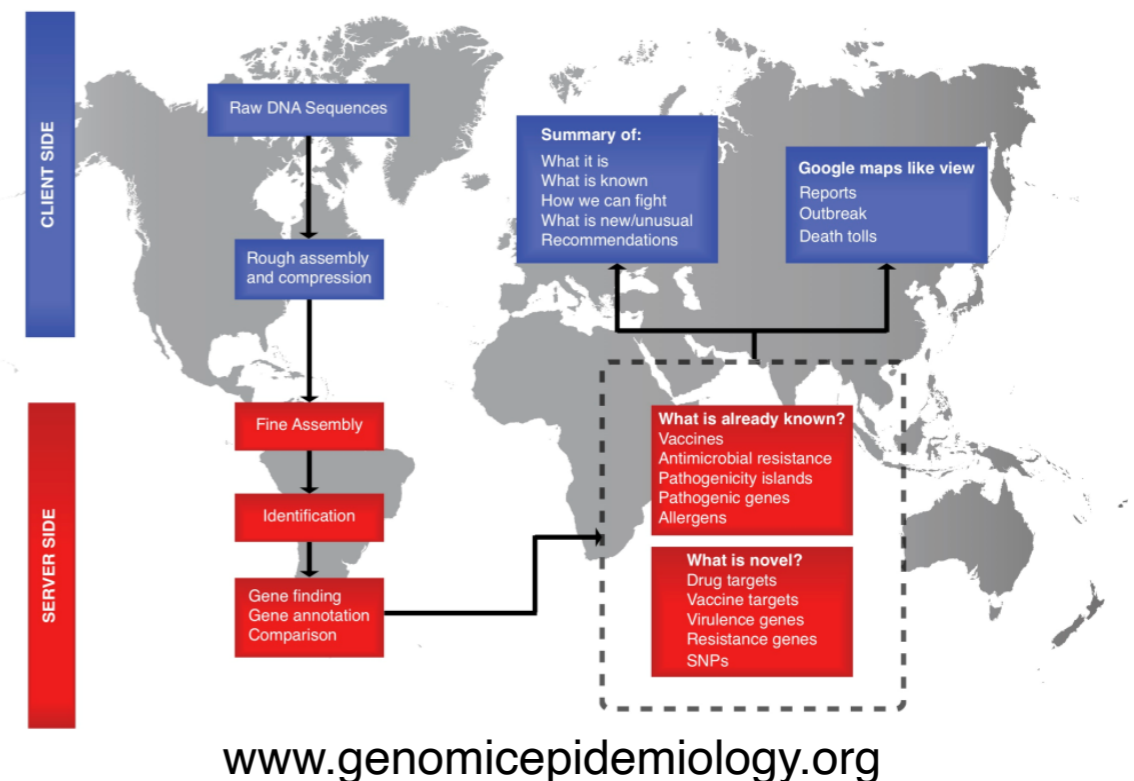


# Preparing for Global Surveillance - Center for Genomic Epidemiology

- Provide a proof of concept of combining bioinformatics with global epidemiology in real-time
- Provide foundation for web based solutions (plug and play tool)

- What is it
- How dangerous is it
- Have we seen it before
- With what can it be treated

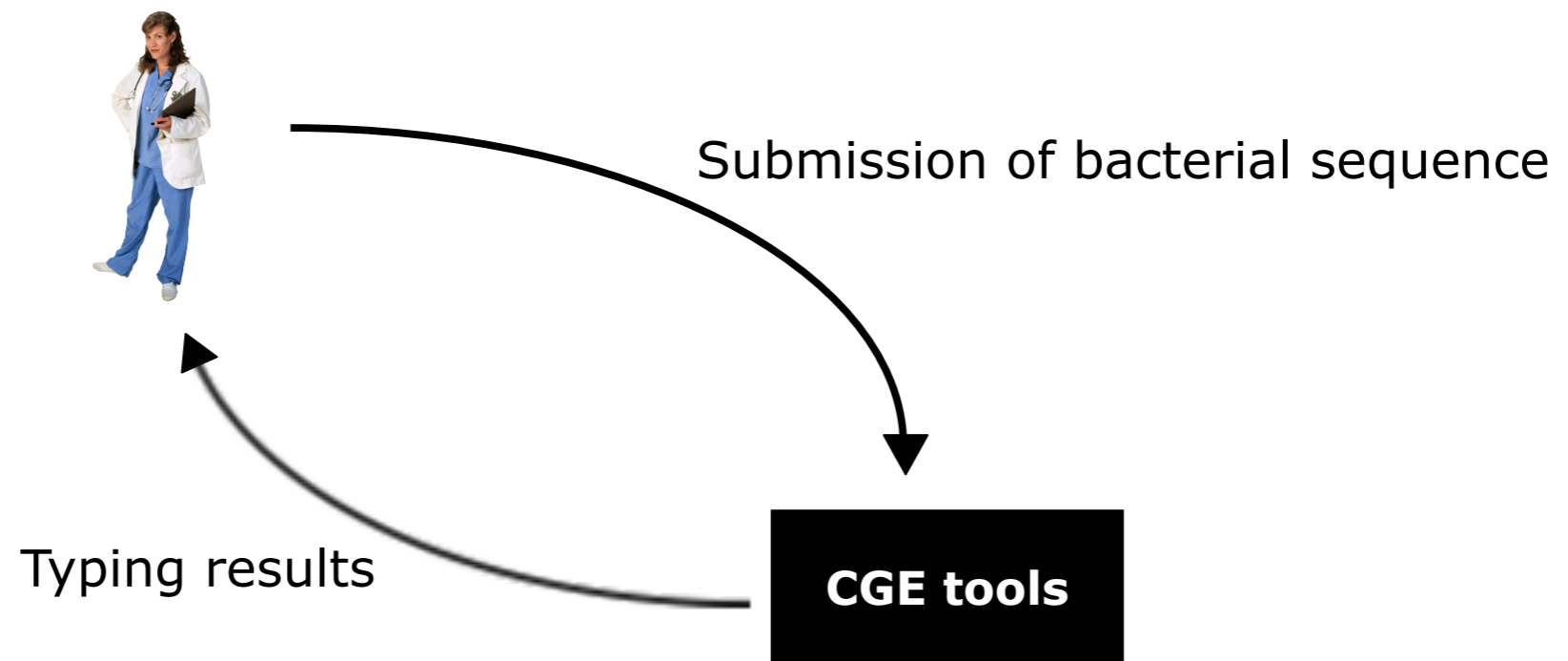
- + Platform independent
- + Requires little computer resources
- + Can be done everywhere
- Requires patience







# CGE tools



# Center for Genomic Epidemiology

Home

Organization

Project

Services

Contact

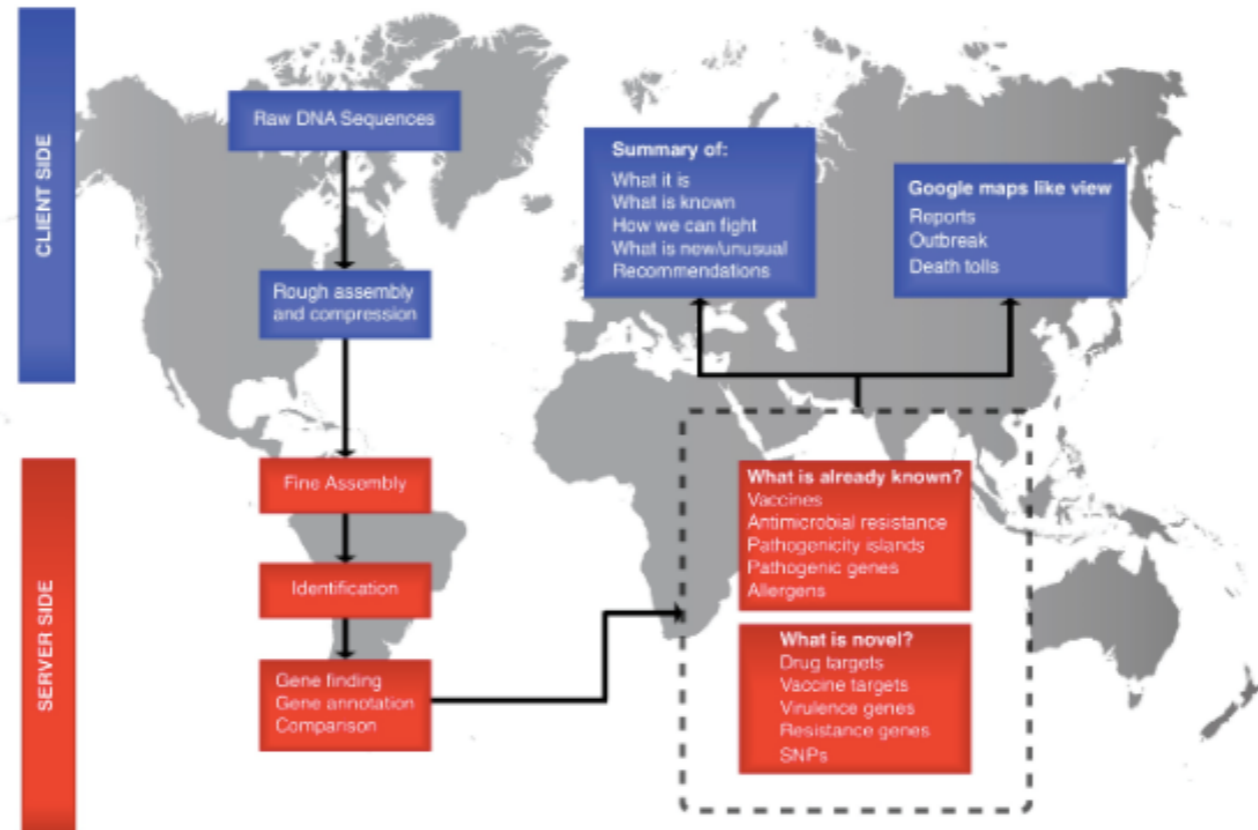
## Services

### Phenotyping:

- Identification of acquired antibiotic resistance genes. [ResFinder](#)
- Prediction of a bacteria's pathogenicity towards human hosts. [PathogenFinder](#)
- Identification of acquired virulence genes. [VirulenceFinder](#)

### Typing:

- Multi Locus Sequence Typing (MLST) from an assembled genome or from a set of reads [MLST](#)
- PlasmidFinder identifies plasmids in total or partial sequenced isolates of bacteria. [PlasmidFinder](#)
- Multi Locus Sequence Typing (MLST) from an assembled plasmid or from a set of reads [pMLST](#)
- Prediction of bacterial species using a fast K-mer algorithm. [KmerFinder](#)
- Prediction of bacterial species using the S16 ribosomal DNA sequence. [SpeciesFinder](#)



## Welcome to the Center for Genomic Epidemiology

The cost of sequencing a bacterial genome is \$50 and is expected to decrease further in the near future and the equipment needed cost less than \$150 000. Thus, within a few years all clinical microbiological laboratories will have a sequencer in use on a daily basis. The price of genome sequencing is already so low that whole genome sequencing will also find worldwide application in human and veterinary practices as well as many other places where bacteria are handled. In Denmark alone this equals more than 1 million isolates annually in 15-20 laboratories and globally up to 1-2 billion isolates per year. The limiting factor will therefore in the future not be the cost of the sequencing, but how to assemble, process and handle the large amount of data in a standardized way that will make the information useful, especially for diagnostic and surveillance.

## News

### Course on the use of the CGE tools in November 2014

September 2014  
The course is for clinical microbiologists to learn how to use the CGE tools. The course will be taught in English and take place at the Technical University of Denmark [Course flyer \(pdf\)](#)

### Benchmarking of Methods for Genomic Taxonomy

April 2014  
How to optimally determine taxonomy from whole genome sequences. [Link to article...](#)

### CGE tools applied for bacteriophage characterization

March 2014  
Applying the ResFinder and VirulenceFinder web-services for easy identification of acquired antibiotic resistance and E. coli virulence genes in bacteriophage and prophage nucleotide sequences. [Link to article...](#)

### Evaluation of Whole Genome Sequencing for Outbreak Detection of Salmonella enterica

March 2014  
We evaluated WGS for outbreak detection of Salmonella enterica including different approaches for analyzing and comparing with a traditional typing, PFGE. [Link to article...](#)

### Low-bandwidth and non-compute intensive remote identification of microbes from raw sequencing reads

January 2014  
Cheap dna sequencing may soon become routine not only for human

# Center for Genomic Epidemiology

Username   
Password

Home

Services

Datasets

User Home

## Overview of Services

### Workflows

[Bacterial Analysis Batch Upload Pipeline](#) (Works)

### Phenotyping

[ResFinder](#) (Works)  
[PathogenFinder](#) (Works)  
[VirulenceFinder](#) (Works)  
[Restriction-ModificationFinder](#) (Works)

### Typing

[SeqSero](#) (Works)  
[SerotypeFinder](#) (Works)  
[PAst](#) (in development)  
[VirusFinder](#) (in development)  
[spaTyper](#) (Works)  
[MLST](#) (Works)  
[pMLST](#) (Works)  
[PlasmidFinder](#) (Works)  
[KmerFinder](#) (Works)  
[SpeciesFinder](#) (Works)  
[Read2Type](#) (This service is not implemented on the new server)  
[TaxonomyFinder](#) (This program is in development)  
[Tapir](#) (This service is not implemented on the new server)

### Phylogeny

[snpTree](#) (Works)  
[NDtree](#) (Works)  
[CSIPhylogeny](#) (Works)  
[TreeViewer](#) (Works)

### Other

[Assembler](#) (Works)  
[ENAUUploader](#) (in development)  
[PanFunPro](#) (Works)  
[MGmapper](#) (Works)  
[MyDbFinder](#) (Works)  
[SPIFinder](#) (Works)  
[HostPhinder](#) (in development)  
[GeneticDiseaseProject](#) (Not associated with CGE)  
[NetFCM](#) (Not associated with CGE)

# KmerFinder

<http://cge.food.dtu.dk/services/KmerFinder/>

# Center for Genomic Epidemiology

Username   
Password

Home

Services

Instructions

Output

Article abstract

## KmerFinder 3.1

View the [version history](#) of this server.

### Select the database

bacteria organisms (K: 16, P: ATG)

If you get an "Access forbidden. Error 403": Make sure the start of the web adress is https and not just http. Fix it by clicking [here](#).

Name	Size	Progress	Status
------	------	----------	--------

# Center for Genomic Epidemiology

Username   
Password

Home

Services

Instructions

Output

Article abstract

## KmerFinder 3.1

View the [version history](#) of this server.

### Select the database

- ✓ bacteria organisms (K: 16, P: ATG)
- bacteria plasmids (K: 16, P: T)
- bacteria type strains (K: 16, P: ATG)
- fungi (K: 16, P: ATG)
- protozoa (K: 16, P: ATG)
- archaea (K: 16, P: ATG)

start of the web address is https and not just http. Fix it by clicking [here](#).

Name

Size

Progress

Status

# Center for Genomic Epidemiology

## Your job is being processed

Wait here to watch the progress of your job, or fill in the form below to get an email message upon completion.

To get notified by email:

This page will update itself automatically.

# KmerFinder output – standard scoring method

## Center for Genomic Epidemiology

Home

Services

Instructions

Output

### KmerFinder-3.1 Server - Results

KmerFinder 3.1 results:

Template	Num	Score	Expected	Template_length	Query_Coverage	Template_Coverage	Depth
NC_016854.1 Salmonella enterica subsp. enterica serovar Typhimurium str. D23580 complete genome	7004	6094006	30	157485	96.86	99.99	38.70

EXTENDED OUTPUT

**Input Files:** *Salmonella-spp-02-03-002\_R1\_001.trim.fq* *Salmonella-spp-02-03-002\_R2\_001.trim.fq*

RESULTS as text (tab separated)





# ResFinder

<https://cge.food.dtu.dk/services/ResFinder/>

# Center for Genomic Epidemiology

Username   
Password

Home

Services

Instructions

Output

Overview of genes

Article abstract

## ResFinder 3.0

ResFinder identifies acquired antimicrobial resistance genes and/or find chromosomal mutations in total or partial sequenced isolates of bacteria.

View the [version history](#) of this server.

The database is curated by:  
**Valeria Bortolaia**  
(click to contact)

Chromosomal point mutations

Acquired antimicrobial resistance genes

Select type of your reads

Assembled Genome/Contigs\*

If you get an "Access forbidden. Error 403": Make sure the start of the web address is https and not just http. Fix it by clicking [here](#).

Name	Size	Progress	Status

# Center for Genomic Epidemiology

Username   
Password

Home

Services

Instructions

Output

Overview of genes

Article abstract

## ResFinder 3.0

ResFinder identifies acquired antimicrobial resistance genes and/or find chromosomal mutations in total or partial sequenced isolates of bacteria.

View the [version history](#) of this server.

The database is curated by:  
**Valeria Bortolaia**  
(click to contact)

Chromosomal point mutations

### Resistance caused by mutations

Select species

- Campylobacter
- E. coli
- Salmonella
- N. gonorrhoeae
- M. tuberculosis

Acquired antimicrobial resistance genes

### Select type of your reads

Assembled Genome/Contigs\*

# Center for Genomic Epidemiology

Username   
Password

Home

Services

Instructions

Output

Overview of genes

Article abstract

## ResFinder 3.0

ResFinder identifies acquired antimicrobial resistance genes and/or find chromosomal mutations in total or partial sequenced isolates of bacteria.

View the [version history](#) of this server.

The database is curated by:  
**Valeria Bortolaia**  
(click to contact)

Chromosomal point mutations

### Resistance caused by mutations

Select species

E. coli

### Show unknown mutations

Show only known mutations  
Show all mutations, known and unknown

Acquired antimicrobial resistance genes

### Select type of your reads

Assembled Genome/Contigs\*

# Center for Genomic Epidemiology

Username   
Password

Home

Services

Instructions

Output

Overview of genes

Article abstract

## ResFinder 3.0

ResFinder identifies acquired antimicrobial resistance genes and/or find chromosomal mutations in total or partial sequenced isolates of bacteria.

View the [version history](#) of this server.

The database is curated by:  
**Valeria Bortolaia**  
(click to contact)

### Chromosomal point mutations

#### Resistance caused by mutations

Select species

#### Show unknown mutations

Show only known mutations

Show all mutations, known and unknown

### Acquired antimicrobial resistance genes

#### Select Antimicrobial configuration

Select multiple items, with Ctrl-Click (or Cmd-Click on Mac) - by default all databases are selected

- Aminoglycoside
- Beta-lactam
- Colistin
- Fluoroquinolone
- Fosfomycin
- Fusidic Acid

#### Select threshold for %ID

#### Select minimum length

Chromosomal point mutations

Resistance caused by mutations

Select species

E. coli

Show unknown mutations

- Show only known mutations
- Show all mutations, known and unknown

Acquired antimicrobial resistance genes

Select type of your reads

Assembled Genome/Contigs\*

If you get an "Access forbidden. Error 403": Make sure the start of the web address is https and not just http. Fix it by clicking [here](#).

📁 Isolate File

Name	Size	Progress	Status
strain01.fasta	4.63 MB	<div style="width: 100%;"></div>	

📤 Upload 🗑 Remove



## Chromosomal point mutations - Results

Species: *e.coli*

### Known Mutations

parE				
No mutations found in parE				

parC				
No known mutations found in parC				

folP				
No mutations found in folP				

gyrA				
Mutation	Nucleotide change	Amino acid change	Resistance	PMID
gyrA p.S83L	TCG → TTG	S → L	Quinolones, Fluoroquinolones	<a href="#">15848289</a>

pmrB				
No known mutations found in pmrB				

pmrA				
No mutations found in pmrA				

16S_rrsB				
No mutations found in 16S_rrsB				

16S_rrsH				
No known mutations found in 16S_rrsH				

gyrB				
No mutations found in gyrB				

ampC				
Mutation	Nucleotide change	Amino acid change	Resistance	PMID
ampC promoter n.-42C>T	C → T	Promoter mutations	B-lactam resistance	<a href="#">21653764</a>

# Center for Genomic Epidemiology

Home Services Instructions Output Overview of genes Article abstract

## ResFinder-2.1 Server - Results

Aminoglycoside						
Resistance gene	%Identity	Query/HSP length	Contig	Position in contig	Predicted phenotype	Accession number
<i>strA</i>	100.00	804 / 804	strain_1_contig_11	3559..4362	Aminoglycoside resistance Alternate name; aph(3")-Ib	<a href="#">AF321551</a>
<i>strB</i>	100.00	837 / 837	strain_1_contig_11	4362..5198	Aminoglycoside resistance Alternate name; aph(6)-IId	<a href="#">M96392</a>

Beta-lactam						
Resistance gene	%Identity	Query/HSP length	Contig	Position in contig	Predicted phenotype	Accession number
<i>blaCTX-M-15</i>	100.00	876 / 876	strain_1_contig_14	81110..81985	Beta-lactam resistance Alternate name; UOE-1	<a href="#">DQ302097</a>
<i>blaTEM-1B</i>	100.00	861 / 861	strain_1_contig_14	84807..85667	Beta-lactam resistance Alternate name; RblaTEM-1	<a href="#">JF910132</a>

### Colistin

No resistance genes found.



RAPID COMMUNICATIONS

# Detection of *mcr-1* encoding plasmid-mediated colistin-resistant *Escherichia coli* isolates from human bloodstream infection and imported chicken meat, Denmark 2015

H Hasman<sup>1</sup>, AM Hammerum<sup>1</sup>, F Hansen<sup>1</sup>, RS Hendriksen<sup>2</sup>, B Olesen<sup>3</sup>, Y Agersø<sup>2</sup>, E Zankari<sup>2</sup>, P Leekitcharoenphon<sup>2</sup>, M Stegger<sup>1,4</sup>, RS Kaas<sup>2</sup>, LM Cavaco<sup>2</sup>, DS Hansen<sup>3</sup>, FM Aarestrup<sup>2</sup>, RL Skov<sup>1</sup>

1. Department of Microbiology and Infection Control, Statens Serum Institut, Copenhagen, Denmark

2. National Food Institute, Technical University of Denmark, Lyngby, Denmark

3. Department of Clinical Microbiology, Herlev and Gentofte Hospital, Copenhagen University Hospital, Herlev, Denmark

4. Pathogen Genomics Division, Translational Genomics Research Institute (TGen), Flagstaff, Arizona, USA

Correspondence: Henrik Hasman (henh@ssi.dk)

Citation style for this article:

Hasman H, Hammerum A, Hansen F, Hendriksen R, Olesen B, Agersø Y, Zankari E, Leekitcharoenphon P, Stegger M, Kaas R, Cavaco L, Hansen D, Aarestrup F, Skov R. Detection of *mcr-1* encoding plasmid-mediated colistin-resistant *Escherichia coli* isolates from human bloodstream infection and imported chicken meat, Denmark 2015. *Euro Surveill.* 2015;20(49):pii=30085. DOI: <http://dx.doi.org/10.2807/1560-7917.ES.2015.20.49.30085>

Article submitted on 04 December 2015 / accepted on 10 December 2015 / published on 10 December 2015

**"The approximately 3,000 Gram-negative (*E. coli* or *Salmonella*) bacteria, which have previously been mapped using whole genome sequencing, have been reexamined to see whether MCR-1 is present. Results show that MCR-1 was found in one patient, who suffered from a blood infection in 2015 and in five food samples that have been imported from 2012-2014. All the bacteria are multi-resistant ESBL bacteria containing the MCR-1 gene, which can further complicate treatment."**



# CSIPhylogeny

<https://cge.food.dtu.dk/services/CSIPhylogeny/>

# Center for Genomic Epidemiology

Username   
Password

Home

Services

Instructions

Output

Article abstract

## CSI Phylogeny 1.1 (Call SNPs & Infer Phylogeny)

CSI Phylogeny calls SNPs, filters the SNPs, does site validation and infers a phylogeny based on the concatenated alignment of the high quality\* SNPs.

**Note:** The old version of this service is still available at: [CSI Phylogeny 1.0a](#). But it is now deprecated and no longer supported.

**Service updated (14:30 10-Mar-2016 GMT+1).** Service was down for several days due to errors in the queing system. The downtime was exploited to implement a new queing method for this service. It has been tested and should work but please don't hesitate to write Scientific support if your jobs are failing. The update does not affect output results, only where the pipeline is executed on the CGE server.

### Input data

#### Upload reference genome (fasta format)

Note: Reference genome must not be compressed.

no file selected

Include reference in final phylogeny.

#### Select min. depth at SNP positions

10x

#### Select min. relative depth at SNP positions

10 %

#### Select minimum distance between SNPs (prune)

10 bp

#### Select min. SNP quality

30

#### Select min. read mapping quality

25

#### Select min. Z-score

1.96

# Center for Genomic Epidemiology

Username   
Password

Home

Services

Instructions

Output

Article abstract

## CSI Phylogeny 1.1 (Call SNPs & Infer Phylogeny)

CSI Phylogeny calls SNPs, filters the SNPs, does site validation and infers a phylogeny based on the concatenated alignment of the high quality\* SNPs.

**Note:** The old version of this service is still available at: [CSI Phylogeny 1.0a](#). But it is now deprecated and no longer supported.

**Service updated (14:30 10-Mar-2016 GMT+1).** Service was down for several days due to errors in the queing system. The downtime was exploited to implement a new queing method for this service. It has been tested and should work but please don't hesitate to write Scientific support if your jobs are failing. The update does not affect output results, only where the pipeline is executed on the CGE server.

### Input data

#### Upload reference genome (fasta format)

Note: Reference genome must not be compressed.

no file selected

Include reference in final phylogeny.

#### Select min. depth at SNP positions

10x

#### Select min. relative depth at SNP positions

10 %

#### Select minimum distance between SNPs (prune)

10 bp

#### Select min. SNP quality

30

#### Select min. read mapping quality

25

#### Select min. Z-score

1.96

# Center for Genomic Epidemiology

Username   
Password

Home

Services

Instructions

Output

Article abstract

## CSI Phylogeny 1.1 (Call SNPs & Infer Phylogeny)

CSI Phylogeny calls SNPs, filters the SNPs, does site validation and infers a phylogeny based on the concatenated alignment of the high quality\* SNPs.

**Note:** The old version of this service is still available at: [CSI Phylogeny 1.0a](#). But it is now deprecated and no longer supported.

**Service updated (14:30 10-Mar-2016 GMT+1).** Service was down for several days due to errors in the queing system. The downtime was exploited to implement a new queing method for this service. It has been tested and should work but please don't hesitate to write Scientific support if your jobs are failing. The update does not affect output results, only where the pipeline is executed on the CGE server.

### Input data

#### Upload reference genome (fasta format)

Note: Reference genome must not be compressed.

D23580.fasta

Include reference in final phylogeny.

#### Select min. depth at SNP positions

10x

#### Select min. relative depth at SNP positions

10 %

#### Select minimum distance between SNPs (prune)

10 bp

#### Select min. SNP quality

30

#### Select min. read mapping quality

25

#### Select min. Z-score

1.96


Use altered FastTree (more accurate)

Note: Read more [here](#)



**Upload read files and/or assembled genomes (fasta or fastq format)**

Note: Read files must be compressed with gzip (compressed files often ends with .gz).

If you get an "Access forbidden. Error 403": Make sure the start of the web address is https and not just http. Fix it by clicking [here](#).

 Isolate File

Name	Size	Progress	Status
<hr/>			
<hr/>			

 Upload  Remove

**\*High quality SNPs**

A high quality SNP are defined as a SNP that obeys the following rules:

**Confidentiality:**

*The sequences are kept confidential and will be deleted after 48 hours.*

Use altered FastTree (more accurate)

Note: Read more [here](#)

**Upload read files and/or assembled genomes (fasta or fastq format)**

Note: Read files must be compressed with gzip (compressed files often ends with .gz).

If you get an "Access forbidden. Error 403": Make sure the start of the web address is https and not just http. Fix it by clicking [here](#).

Name	Size	Progress	Status
Salmonella-spp-02-03-002.fna	4.80 MB	<div style="width: 100%;"></div>	
Salmonella-spp-02-03-008.fna	4.81 MB	<div style="width: 100%;"></div>	
Salmonella-spp-05-102.fna	4.81 MB	<div style="width: 100%;"></div>	
Salmonella-spp-07-022.fna	4.80 MB	<div style="width: 100%;"></div>	



**\*High quality SNPs**

A high quality SNP are defined as a SNP that obeys the following rules:

**Confidentiality:**

*The sequences are kept confidential and will be deleted after 48 hours.*

**CITATIONS**

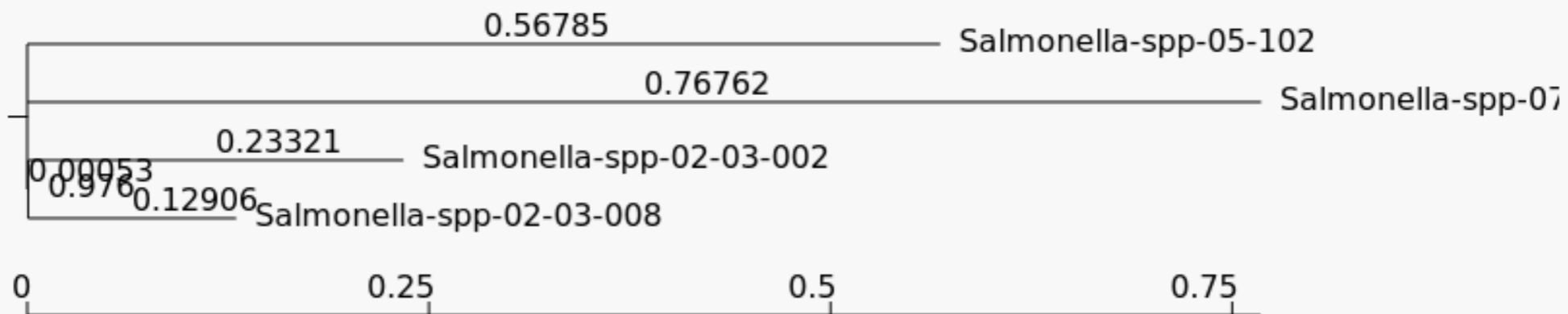
# Center for Genomic Epidemiology

- Home
- Services
- Instructions
- Output
- Article abstract

# Mapper: BWA # Submitting 7 jobs. Waiting for vcfwiz.sh to finish... 0

## CSIPhylogeny Results

The tree presented in the picture below is only meant as a preview. If the tree is meant to be shared or published, we strongly recommend that the 'Newick' file is downloaded and processed using software created for this purpose. We suggest [FigTree](#).



Download phylogeny as:

Download the filtered SNP calls in Variant Calling Format (VCF):

Note: VCF files are compressed with gzip.

Download matrix of SNP pair counts:

Download matrix as:

Download SNP alignment:



**Percentage of reference genome covered by all isolates: 98.3155029799234**

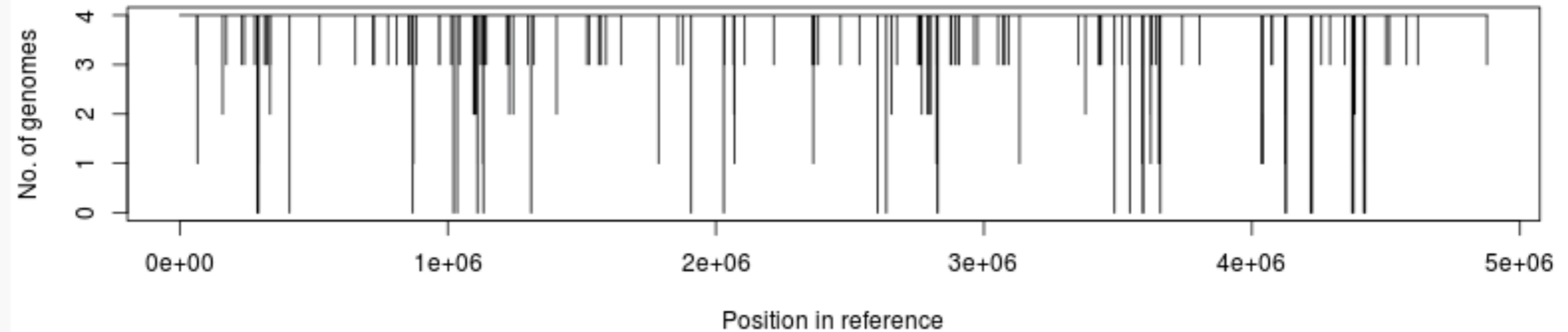
4797128 positions was found in all analyzed genomes.

Size of reference genome: 4879320

Below is listed the number of positions that are shared and trusted between each isolate and the reference genome.

File	Valid positions	Pct. of reference
Salmonella-spp-02-03-008.ignored_snps	4848802	99.3745439938352
Salmonella-spp-02-03-002.ignored_snps	4847669	99.3513235450841
Salmonella-spp-05-102.ignored_snps	4861431	99.6333710435061
Salmonella-spp-07-022.ignored_snps	4821309	98.8110843314232

**Genomes covering each Position**



Download plot:

PDF

# https://bitbucket.org/account/user/genomicepidemiology/projects/CGE

Genomic Epidemiology

Repositories

Language

Create repository

Find repositories

Repository	Project	Last updated	Builds
assimpler2	CGE	an hour ago	
adv_bioinf	Teaching	3 hours ago	
resfinder_db	Databases	4 hours ago	
ccphylo	CGE	7 hours ago	
mlst_db	Databases	15 hours ago	
kma	CGE	3 days ago	
CGE Tools Docker	CGE	2019-12-27	
pyCoDa	CGE	2019-12-20	
cge_core_module	CGE	2019-12-20	
mydbfinder	CGE	2019-12-18	
CholeraeFinder	CGE	2019-12-18	
intfinder	CGE	2019-12-18	
PlasmidFinder	CGE	2019-12-18	
VirulenceFinder	CGE	2019-12-18	
ResFinder	CGE	2019-12-18	
cgMLSTFinder	CGE	2019-12-05	

ResFinder

Source

Commits

Branches

Pull requests

Pipelines

Deployments

Issues

Downloads

Genomic Epidemiology / CGE

## ResFinder

master



Name	Size	Last commit	Message
.gitignore	34 B	2015-07-29	Update
README.md	5.13 KB	2019-08-09	Warning Biopython
resfinder.pl	60.49 KB	2018-10-04	Script updated
resfinder.py	25.01 KB	2019-12-18	fix bug multiple and no hit
test.fsa	4.25 MB	2015-07-16	Updated

## Installation

### Setting up ResFinder script and database

```
# Go to wanted location for resfinder  
cd /path/to/some/dir  
  
# Clone and enter the resfinder directory  
git clone https://git@bitbucket.org/genomicepidemiology/resfinder.git  
cd resfinder  
  
# Installing up the ResFinder database  
# Go to wanted location for resfinder database  
cd /path/to/some/dir  
  
# Clone and enter the resfinder directory  
git clone https://git@bitbucket.org/genomicepidemiology/resfinder_db.git  
cd resfinder_db
```

# Usage

You can run resfinder command line using python3

*# Example of running resfinder*

```
python3 resfinder.py -i test.fsa -o . -p /path/to/resfinder_db \
-mp /path/to/blastn -d aminoglycoside -t 0.90 -l 0.60
```

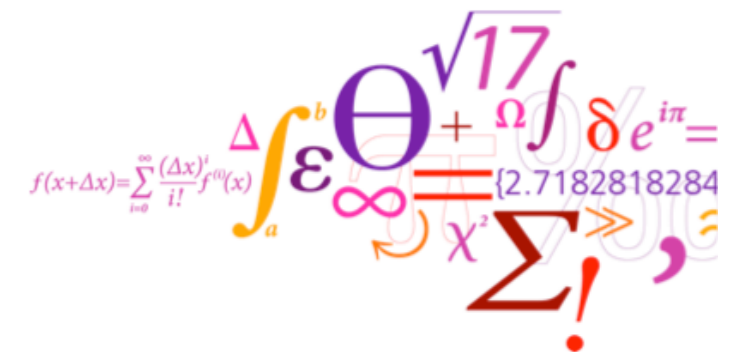
*# The program can be invoked with the -h option*

```
Usage: resfinder.py [-h] [-i INPUTFILE] [-o OUT_PATH]
                [-tmp TMP_DIR] [-mp METHOD_PATH] [-ao ACQ_OVERLAP]
                [-matrix MATRIX] [-p DB_PATH] [-d DATABASES] [-l MIN_COV]
                [-t THRESHOLD] [-x] [-q]
```

optional arguments:

```
-h, --help          show this help message and exit
-i INPUTFILE, --inputfile INPUTFILE
                    Input file (fasta or fastq(s) files)
-o OUT_PATH, --outputPath OUT_PATH
                    Path to blast output
-p DB_PATH, --databasePath DB_PATH
                    Path to the databases
-mp METHOD_PATH --methodPath METHOD_PATH
                    Path to the method to use (kma or blastn)
-d DATABASES, --databases DATABASES
                    Databases chosen to search in - if none are specified
                    all are used
-l MIN_COV, --min_cov MIN_COV
                    Minimum coverage default 0.6
-t THRESHOLD, --threshold THRESHOLD
                    Blast threshold for identity
                    default minimum 0.9
-ao ACQ_OVERLAP --acq_overlap ACQ_OVERLAP
                    Genes are allowed to overlap this number of nucleotides (30)
-matrix, --matrix
                    If used, gives the counts all all called bases at each position
                    in each mapped template. Columns are: reference base,
                    A count, C count, G count, T count, N count, - count.
-x --extended_output
                    If used, give extended output with alignment files,
                    "template and query hits in fasta and a tab
                    "seperated file with gene profile results
-q --quiet
```

# Genomic epidemiology for global surveillance AMR





# Global surveillance





# Whole genome sequencing vs Metagenomics

## Whole genome sequencing

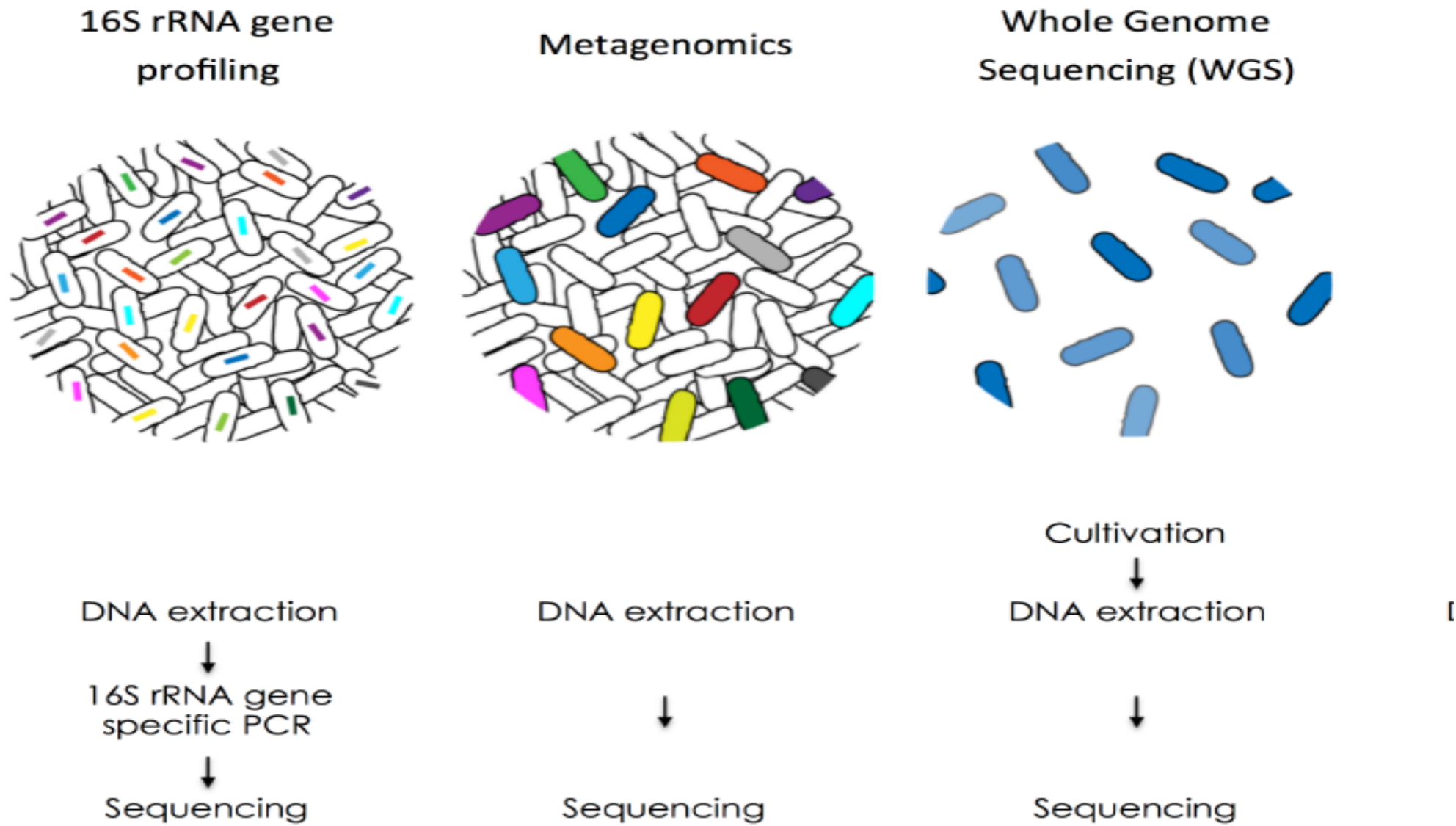
- Sequencing the entire genome of a pure culture (single isolation), including chromosome and plasmids
- Identification, Typing, identification of genetic markers (resistance genes) and phylogenetic relatedness

## Metagenomics

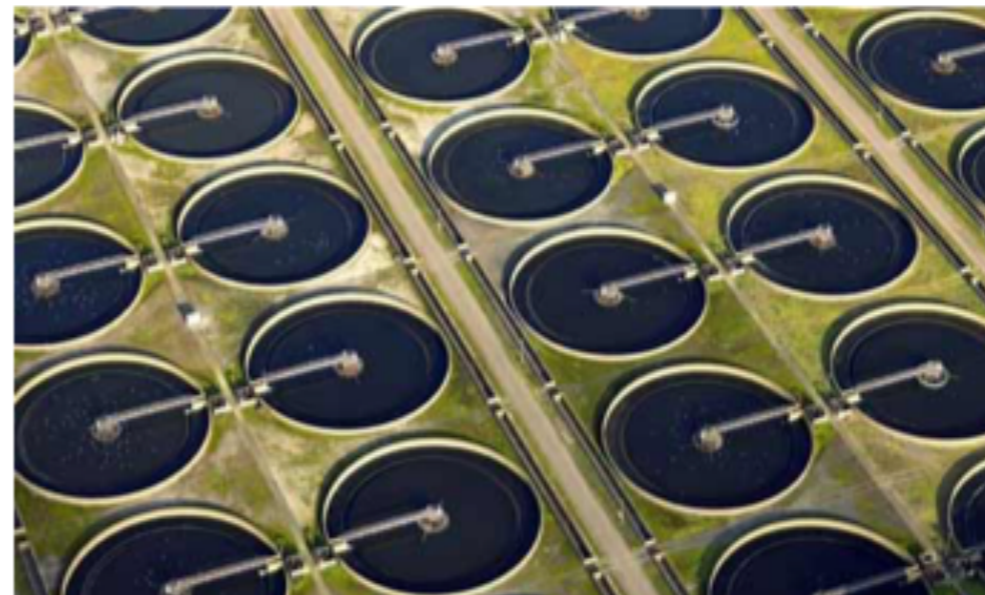
- Sequencing the DNA of the complex community without isolating the individual microorganisms (mixed of multiple organisms)



# Microbial Genomics



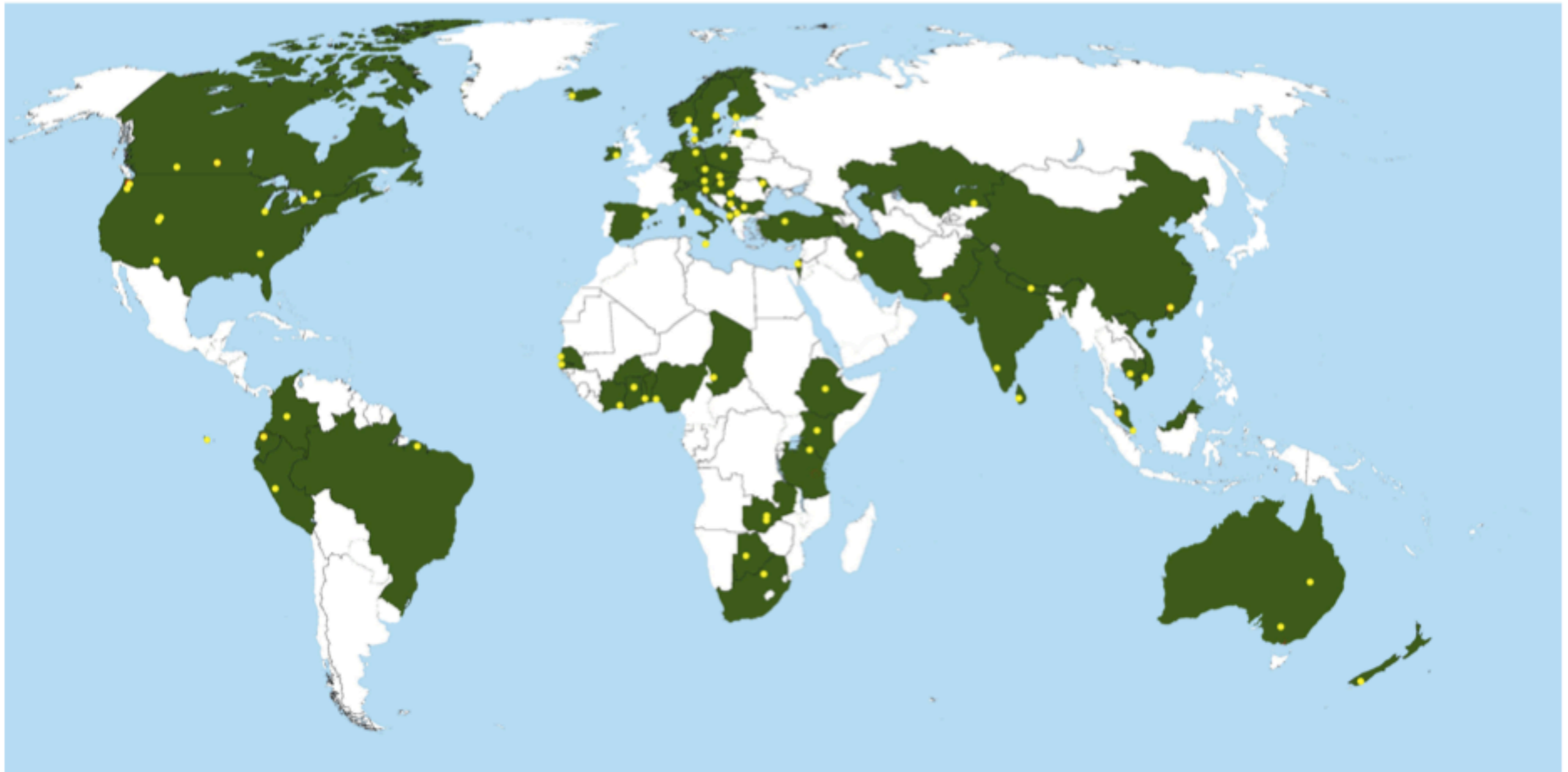
# Global surveillance of AMR



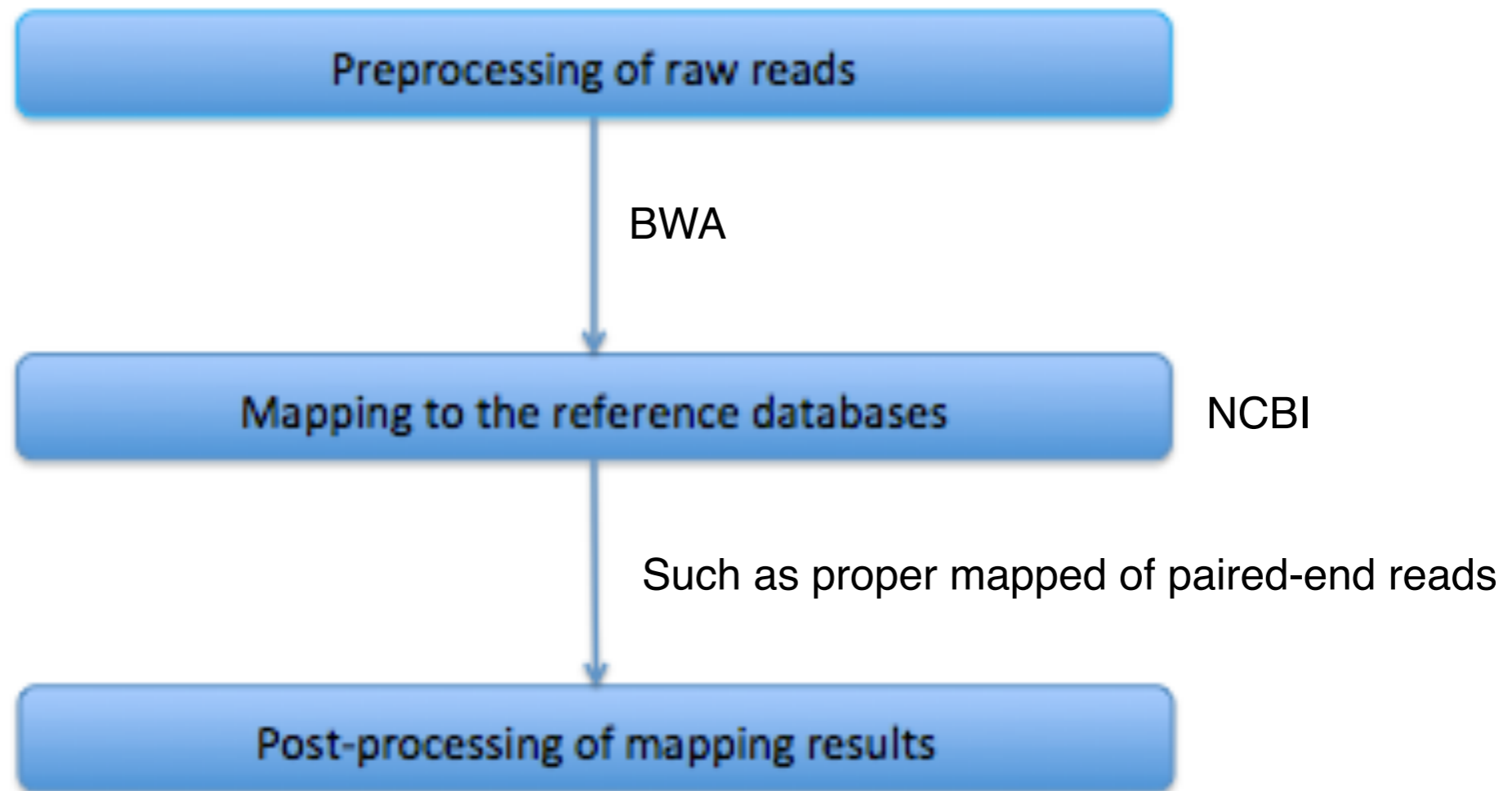
**Can human sewage be used to detect and combined with modelling explain global emergence and trends in AMR ?**

# Sample collection - 2016

79 samples from 60 countries have been collected and analysed



# MGmapper



# mapping AMR result

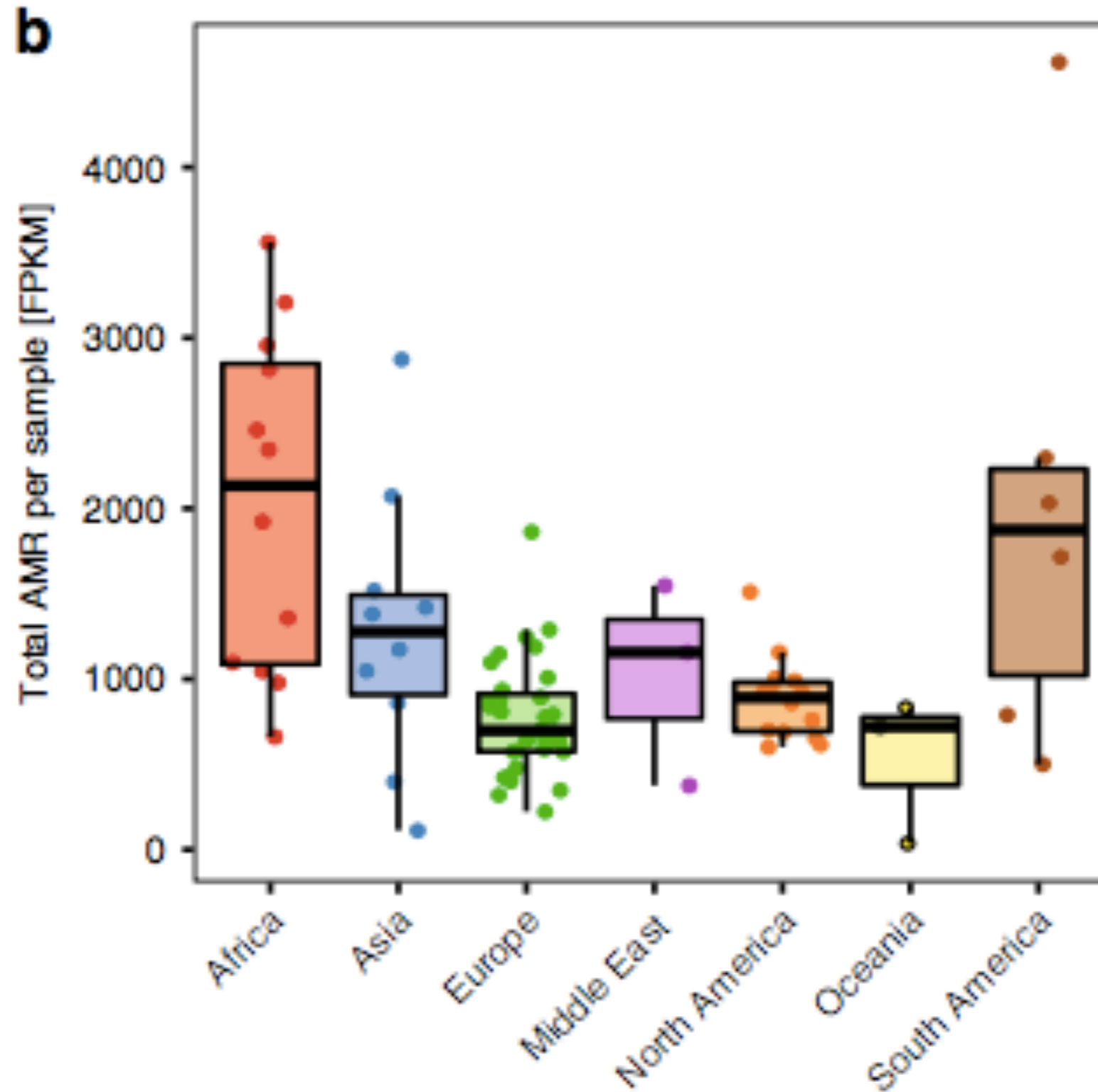
Gene	Description	GeneSize	ALB.17	AUS.18	AUS.18a	AUT.70		Sample	ReadCount_Bacteria
16S_rRNA_meth vltransferaseI	gentamicin	732	0	0	0	0	0	ALB.17	29715190
16S_rRNA_meth vltransferaseI	gentamicin	1263	0	0	0	0	0	AUS.18	11724196
16S_rRNA_meth vltransferaseI	gentamicin	765	0	0	0	0	0	AUS.18a	3778910
16S_rRNA_meth vltransferaseI	gentamicin	765	0	0	0	0	0	AUT.70	15734026
16S_rRNA_meth vltransferaseI	gentamicin	804	0	0	0	0	0	BGR.66	19108694
16S_rRNA_meth vltransferaseI	gentamicin	804	0	0	0	0	0	BRA.53	17108324
16S_rRNA_meth vltransferaseI	gentamicin	765	0	0	0	0	0	BRA.53a	7733328
16S_rRNA_meth vltransferaseI	gentamicin	867	50	2	2	6	6	BWA.19	6599146
16S_rRNA_meth vltransferaseI	gentamicin	867	50	2	2	6	6	CAN.22	1423112
AAC  GQ343136.1	gentamicin	867	50	2	2	6	6	CAN.22a	868336
AAC  GQ343186.1	sisomycin	429	0	0	0	0	0	CAN.22b	11949640
AAC  KJ695106.1	gentamicin	501	0	0	0	0	0	CAN.22c	21282768
beta_lactamase  GQ342998.1	amikacin	891	0	0	0	0	0		
beta_lactamase  GQ342999.1	amikacin	891	0	0	0	0	0		
beta_lactamase  GQ343000.1	amikacin	906	0	0	0	0	0		
beta_lactamase  GQ343002.1	amikacin	873	0	0	0	0	0		
beta_lactamase  GQ343003.1	amikacin	891	0	0	0	0	6		
beta_lactamase  GQ343008.1	amikacin	1263	2	0	0	0	0		
beta_lactamase  GQ343010.1	amikacin	1167	0	0	0	0	0		
beta_lactamase  GQ343015.1	amikacin	1263	0	0	0	0	6		
beta_lactamase  GQ343019.1	amikacin	891	0	0	0	0	0		

**FPKM**

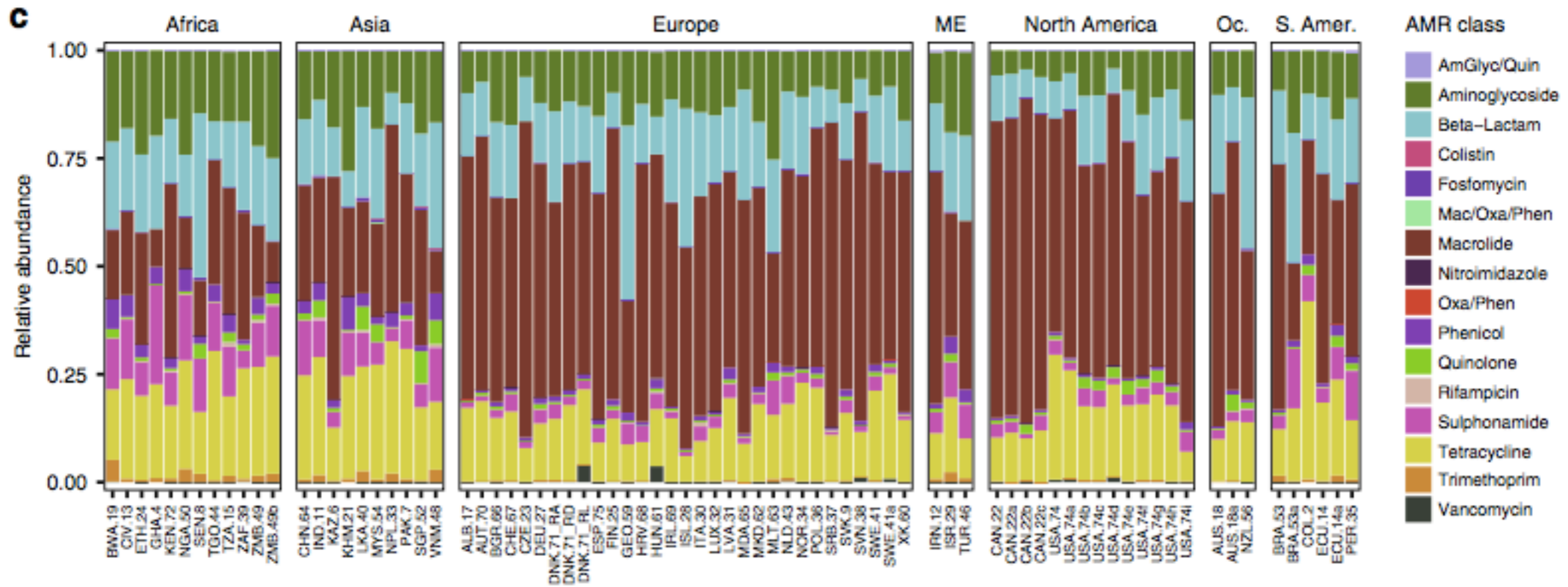
**Fragments Per Kilobase reference per Million bacterial fragments<sub>x</sub>**

# Total FPKM

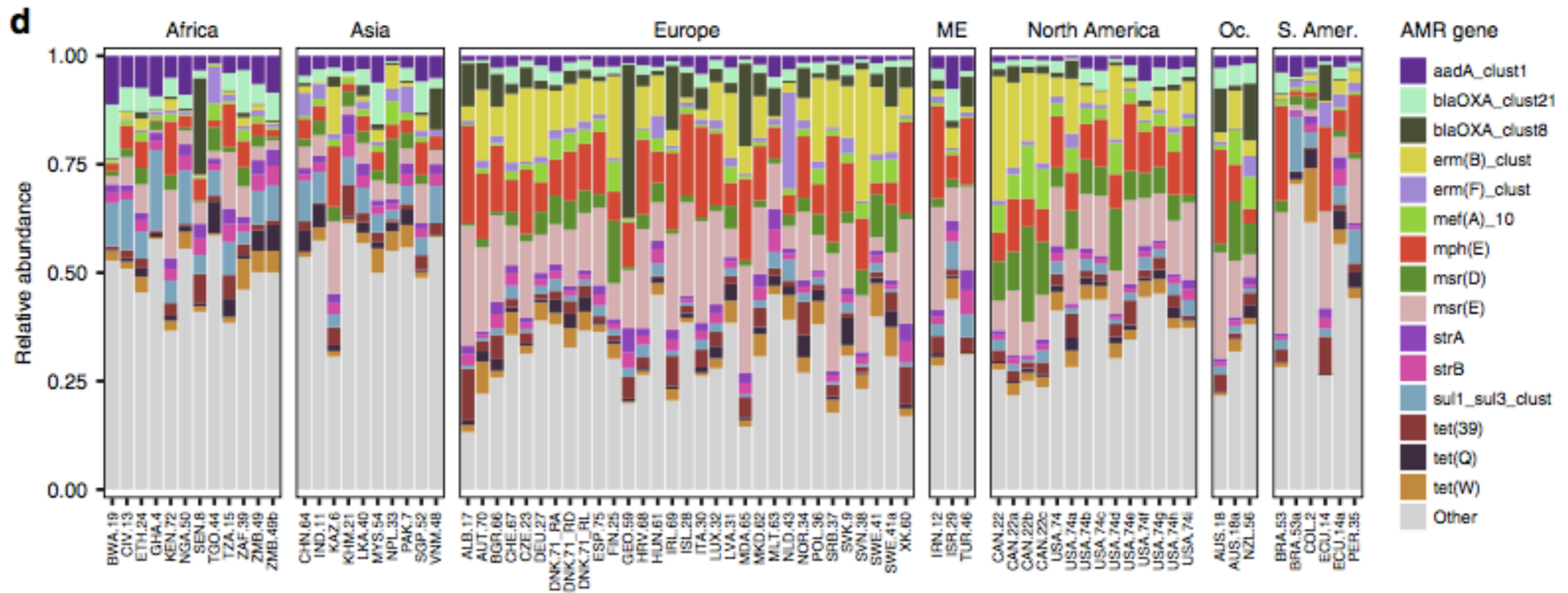
Fragments Per Kilobase reference per Million bacterial fragments



# AMR classes



# AMR genes

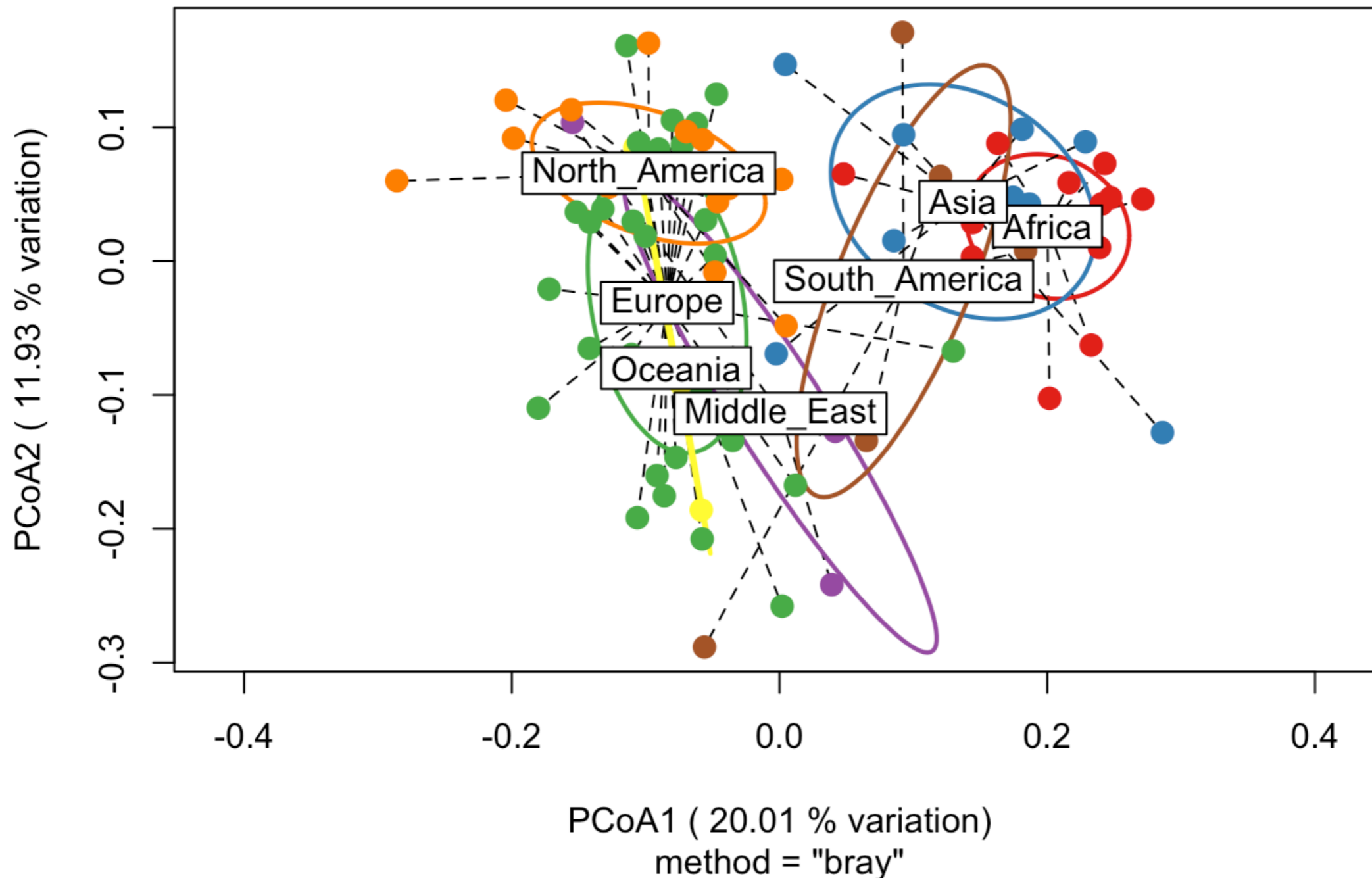




# Resistome clustering in sewage across regions

Hellinger-transformed (decostand function in vegan package) and Bray-Curtis dissimilarity

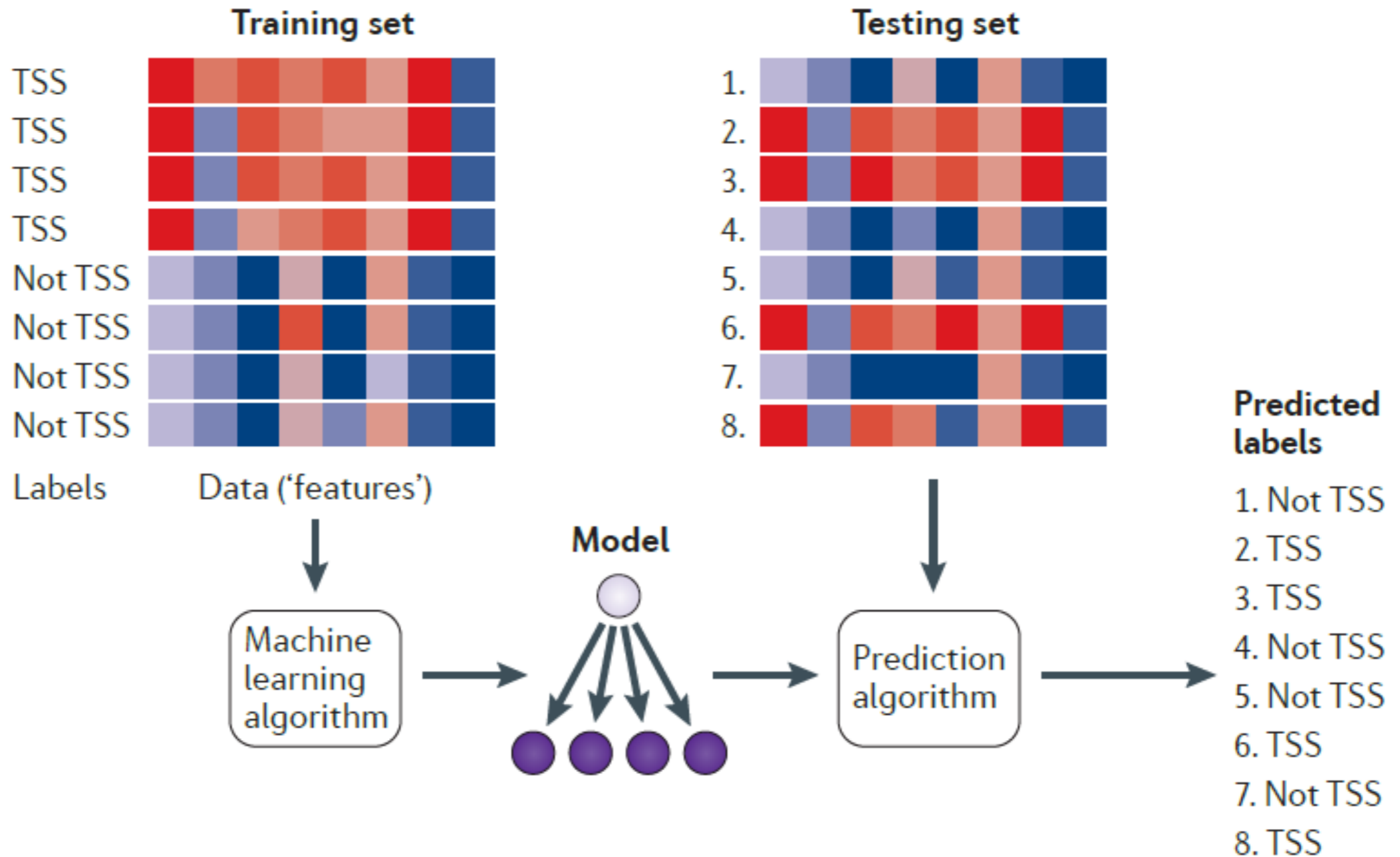
## PCoA ResFinder



# Drivers for AMR

Factor	Significance
Temperature	-
Flight connections	-
Antimicrobial use	*
Human development index	***

# Predict AMR level using socio-economic data



## Deeper look into socio-economic data from World bank

### Predictors of higher AMR

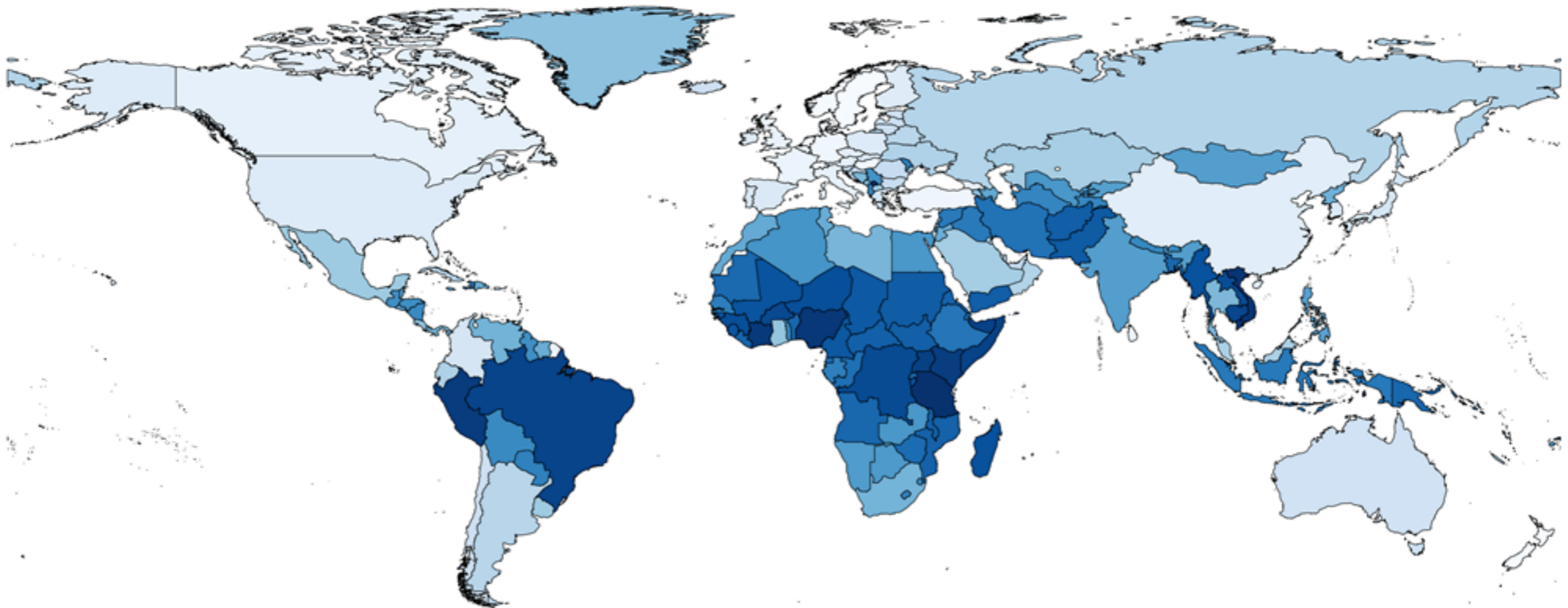
- Mortality rate
- Death, by communicable diseases and maternal, prenatal and nutrition conditions
- Risk of maternal death
- Open defecation
- Diarrhoea prevalence in children
- Risk of impoverishing expenditure for surgical care
- Informal employment
- Time to import

### Predictors of lower AMR

- Investment in water and sanitation
- Completeness of death reporting
- Educational attainment
- Number of surgical procedures
- Life expectancy at birth
- Number of Physicians
- Births attended by skilled health staff
- Grace period on external debt



# Global resistance prediction



# Genomic analysis of sewage from 101 countries reveals global landscape of antimicrobial resistance

Received: 25 July 2022

Accepted: 20 October 2022

Published online: 01 December 2022

 Check for updates

Patrick Munk <sup>1</sup>✉, Christian Brinch <sup>1</sup>, Frederik Duus Møller<sup>1</sup>, Thomas N. Petersen<sup>1</sup>, Rene S. Hendriksen<sup>1</sup>, Anne Mette Seyfarth<sup>1</sup>, Jette S. Kjeldgaard<sup>1</sup>, Christina Aaby Svendsen<sup>1</sup>, Bram van Bunnik <sup>2</sup>, Fanny Berglund <sup>3</sup>, Global Sewage Surveillance Consortium\*, D. G. Joakim Larsson <sup>3</sup>, Marion Koopmans <sup>4</sup>, Mark Woolhouse <sup>2</sup> & Frank M. Aarestrup <sup>1</sup>

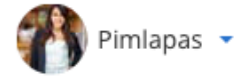
Antimicrobial resistance (AMR) is a major threat to global health. Understanding the emergence, evolution, and transmission of individual antibiotic resistance genes (ARGs) is essential to develop sustainable strategies combatting this threat. Here, we use metagenomic sequencing to analyse ARGs in 757 sewage samples from 243 cities in 101 countries, collected from 2016 to 2019. We find regional patterns in resistomes, and these differ between subsets corresponding to drug classes and are partly driven by taxonomic variation. The genetic environments of 49 common ARGs are highly diverse, with most common ARGs carried by multiple distinct genomic contexts globally and sometimes on plasmids. Analysis of flanking sequence revealed ARG-specific patterns of dispersal limitation and global transmission. Our data furthermore suggest certain geographies are more prone to transmission events and should receive additional attention.

<https://www.coursera.org/learn/wgs-bacteria/>



Catalog Search catalog

For Enterprise



Home > Life Sciences > Clinical Science

# Whole Genome Sequencing of bacterial genomes – tools and applications

**About this course:** This course will cover the topic of Whole genome sequencing (WGS) of bacterial genomes which is becoming more and more relevant for the medical sector. WGS technology and applications are high on international political agenda, as the classical methods are being replaced by WGS technology and therefore bioinformatic tools are extremely important for allowing the people working in

[More](#)

**Who is this class for:** This course is for you if you are interested in getting to know more about Whole genome sequencing applied to bacterial characterization and surveillance. We aim at having a broad scope and international reach in different sectors. So this course is for you whether you are an undergraduate or graduate student, a researcher, medical or veterinary related professional, technical staff or simply interested in the subject!

**Created by:** Technical University of Denmark (DTU)

Financial Aid is available for learners who cannot afford the fee. [Learn more and apply.](#)

Overview  
Syllabus  
FAQs  
Creators  
Ratings and Reviews

Whole genome sequencing of bacterial genomes – tools and applications

[Go to Course](#)

Already enrolled

# 23262 Infectious disease bioinformatics

## Course information

<b>Danish title</b>	Bioinformatik for smitsomme sygdomme
<b>Language of instruction</b>	English
<b>Point( ECTS )</b>	5
<b>Course type</b>	MSc Offered as a single course Technological Specialization Course, MSc. Eng., Bioinformatics and System Biologi
<b>Schedule</b>	Spring F3A (Tues 8-12)
<b>Location</b>	Campus Lyngby
<b>Scope and form</b>	Lectures and computer-based exercises.
<b>Duration of Course</b>	13 weeks
<b>Type of assessment</b>	Oral examination Project presentation in groups followed by individual questions based on the presentation. One final grade for group presentation and individual questions.
<b>Aid</b>	All Aid
<b>Evaluation</b>	7 step scale , internal examiner
<b>Previous Course</b>	27683 and 36683
<b>Not applicable together with</b>	<b>27683/ 36683</b>
<b>Recommended prerequisites</b>	<b>23205</b> , Unix/Linux background. Basic understanding of bioinformatics in sequence analysis and infectious disease.



## 23205 Fighting infectious diseases

### Course information

<b>Danish title</b>	Bekæmpelse af infektionssygdomme
<b>Language of instruction</b>	English
<b>Point( ECTS )</b>	5
<b>Course type</b>	MSc Offered as a single course Technological specialization course, MSc. Eng., Quantitative Biology and Disease Modelling Technological specialization course, MSc. Eng., Biotechnology Technological specialization course, MSc. Eng., Pharmaceutical Design and Engineering Technological Specialization Course, MSc. Eng., Bioinformatics and System Biologi
<b>Schedule</b>	Autumn E4B (Fri 8-12)
<b>Location</b>	Campus Lyngby
<b>Scope and form</b>	Lectures, exercises, team work
<b>Duration of Course</b>	13 weeks
<b>Date of examination</b>	E4B
<b>Type of assessment</b>	Written examination and exercises Written examination (75%), Team presentation of a scientific article (25%)
<b>Exam duration</b>	3 hours
<b>Aid</b>	All Aid
<b>Evaluation</b>	7 step scale , internal examiner
<b>Recommended prerequisites</b>	<a href="#">23258.23256</a> , Knowledge in microbial diversity and the basic biological processes related to molecular microbiology, microbial physiology, and host immune defenses. Basic microbiological experimental skills/knowledge, such as microbial cultivation and PCR. Basic statistical skills/knowledge, such as understanding probability and interpreting results of statistical tests as applied in e.g. experimental and epidemiological studies.



Pimlapas Leekitcharoenphon (Shinny)  
Research Group for Genomic Epidemiology, DTU-Food

[pile@food.dtu.dk](mailto:pile@food.dtu.dk)



@ShinnyPimlapas