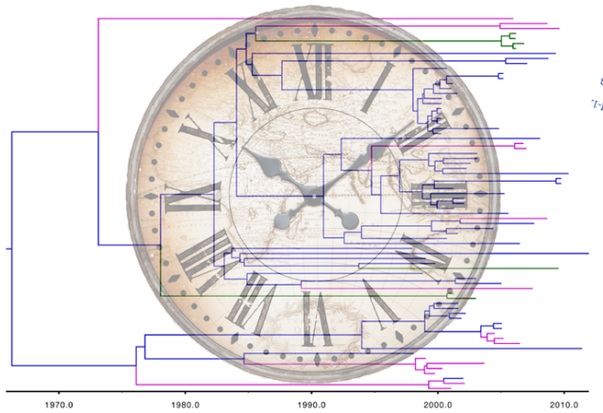
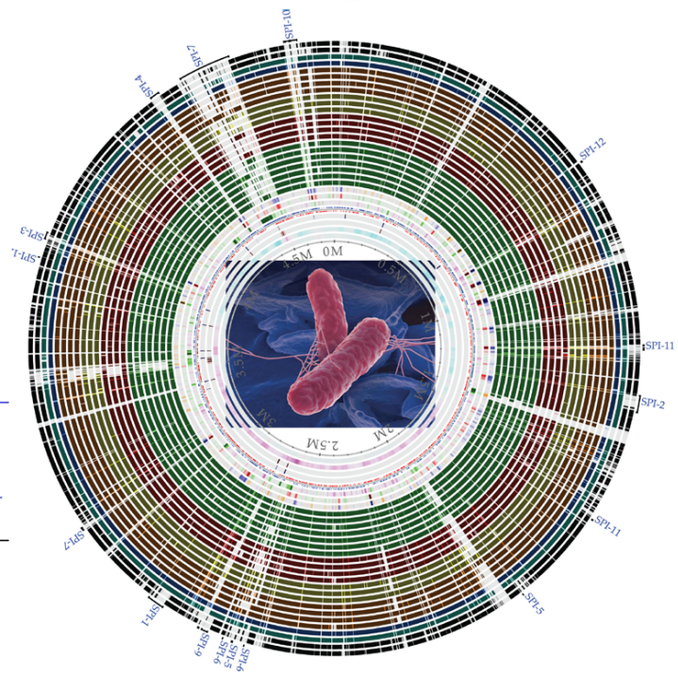
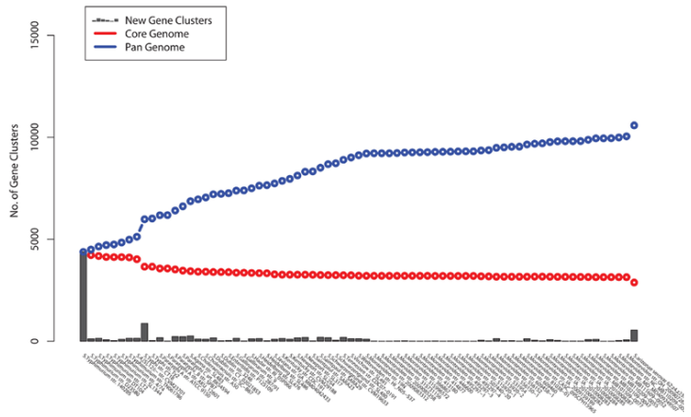
Genomic Epidemiology



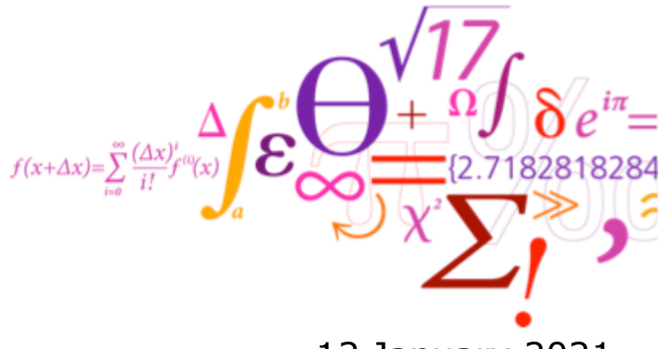




Pimlapas Leekitcharoenphon (Shinny)
Research Group for Genomic Epidemiologies
National Food Institute (DTU Food)

WHO Collaborating Centre for Antimicrobial Resistance in Foodborne Pathogens and Genomics

European Union Reference Laboratory for Antimicrobial Resistance (EURL-AMR)
pile@food.dtu.dk
@ShinnyPimlapas



12 January 2021

About myself

- Biotechnology and Bioinformatics backgrounds

- WGS analysis of foodborne pathogens and metagenomics for surveillance of AMR

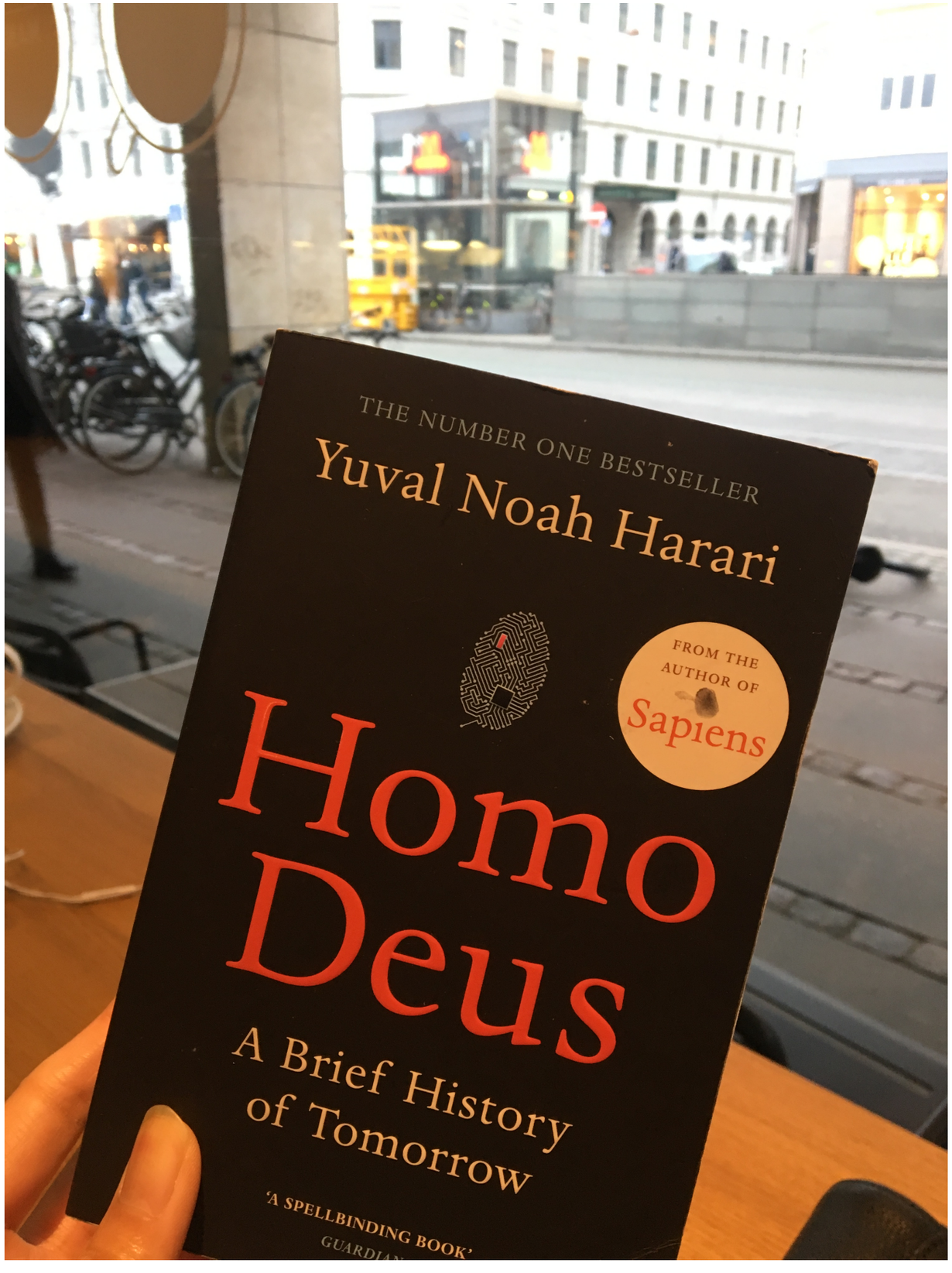
- Researcher at DTU Food

- Onsite course on “23259- Whole genome analysis in diagnostic microbiology”

- Online courses 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 ?

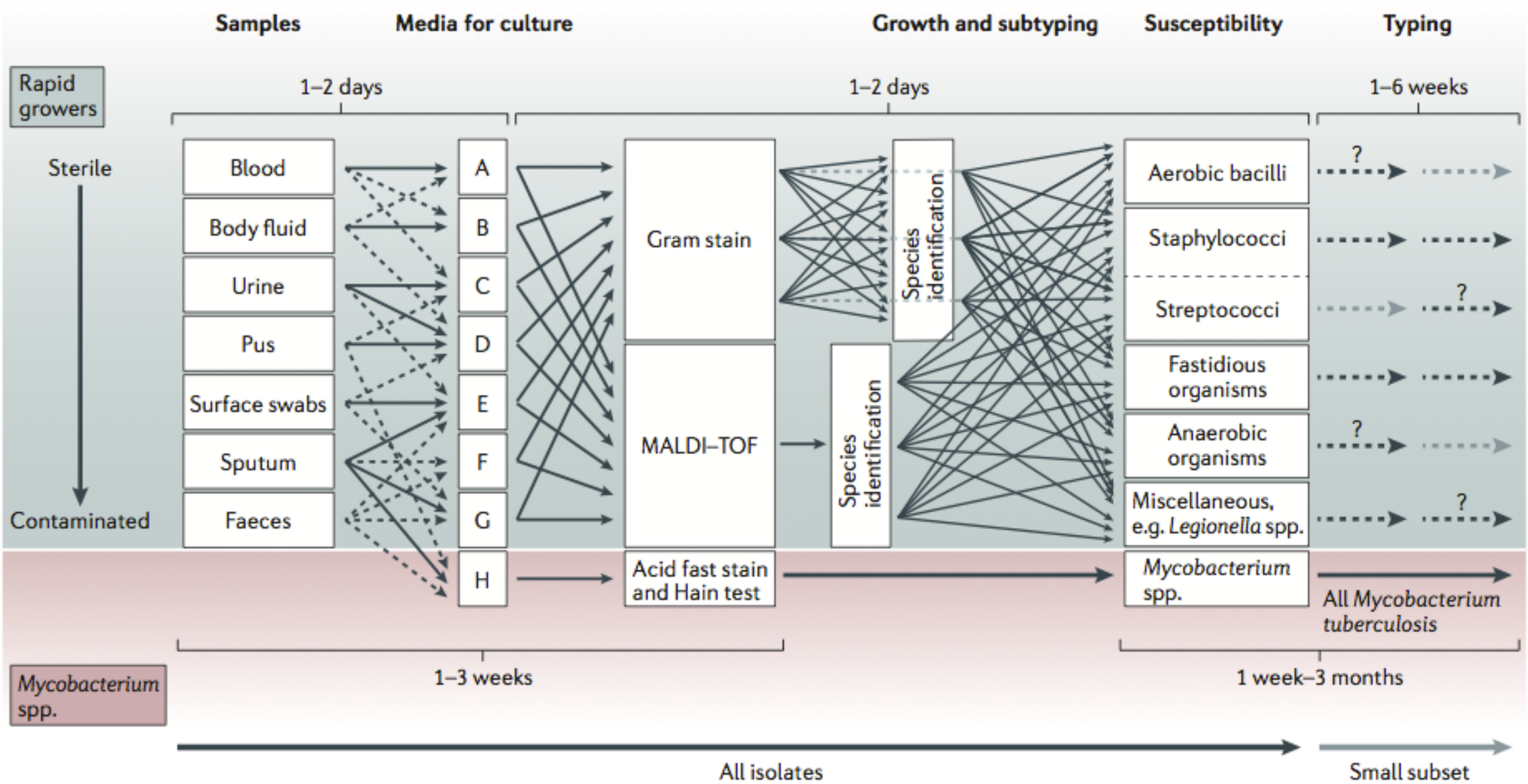
  - Is it an outbreak ?

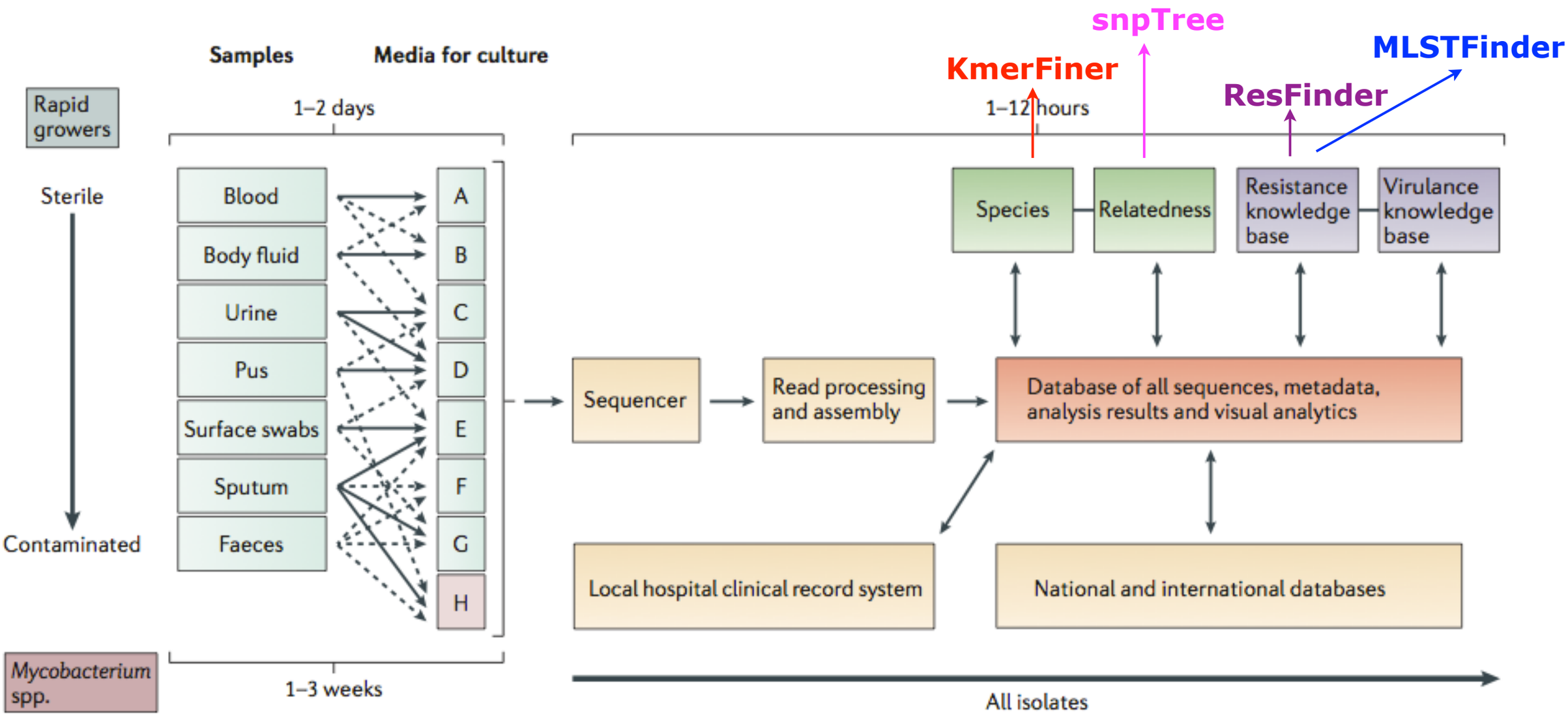
Identification and Typing

- 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**


Samples
Media for culture
Growth and subtyping
Susceptibility
Typing
Rapid growers
1–2 days
1–2 days
1–6 weeks
Sterile
Blood
Body fluid
Urine
Pus
Surface swabs
Sputum
Faeces
Contaminated
A
B
C
D
E
F
G
H
Gram stain
MALDI–TOF
Acid fast stain and Hain test
Species identification
Species identification
Aerobic bacilli
Staphylococci
Streptococci
Fastidious organisms
Anaerobic organisms
Miscellaneous, e.g. Legionella spp.
Mycobacterium spp.
?
?
?
?
All Mycobacterium tuberculosis
Mycobacterium spp.
1–3 weeks
1 week–3 months
All isolates
Small subset


Samples
Media for culture
Rapid growers
1–2 days
Sterile
Blood
Body fluid
Urine
Pus
Surface swabs
Sputum
Faeces
A
B
C
D
E
F
G
H
Contaminated
Mycobacterium spp.
1–3 weeks
Sequencer
Read processing and assembly
Database of all sequences, metadata, analysis results and visual analytics
1–12 hours
KmerFiner
snpTree
ResFinder
MLSTFinder
Species
Relatedness
Resistance knowledge base
Virulance knowledge base
Local hospital clinical record system
National and international databases
All isolates

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

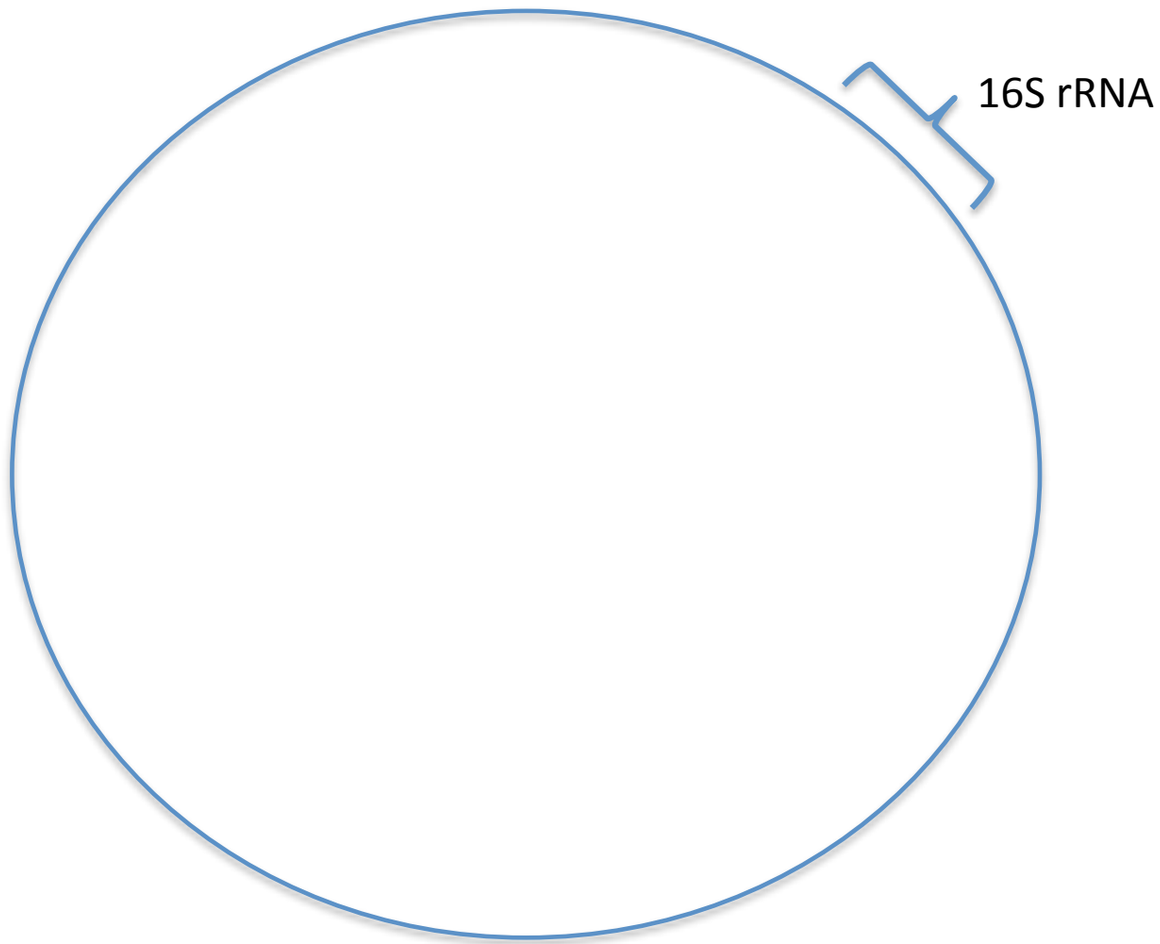
CGE implementation of
16S rRNA species identification - SpeciesFinder

Reference database

- 16S rRNA genes are isolated from genomes in NCBI

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

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



K-mer ?

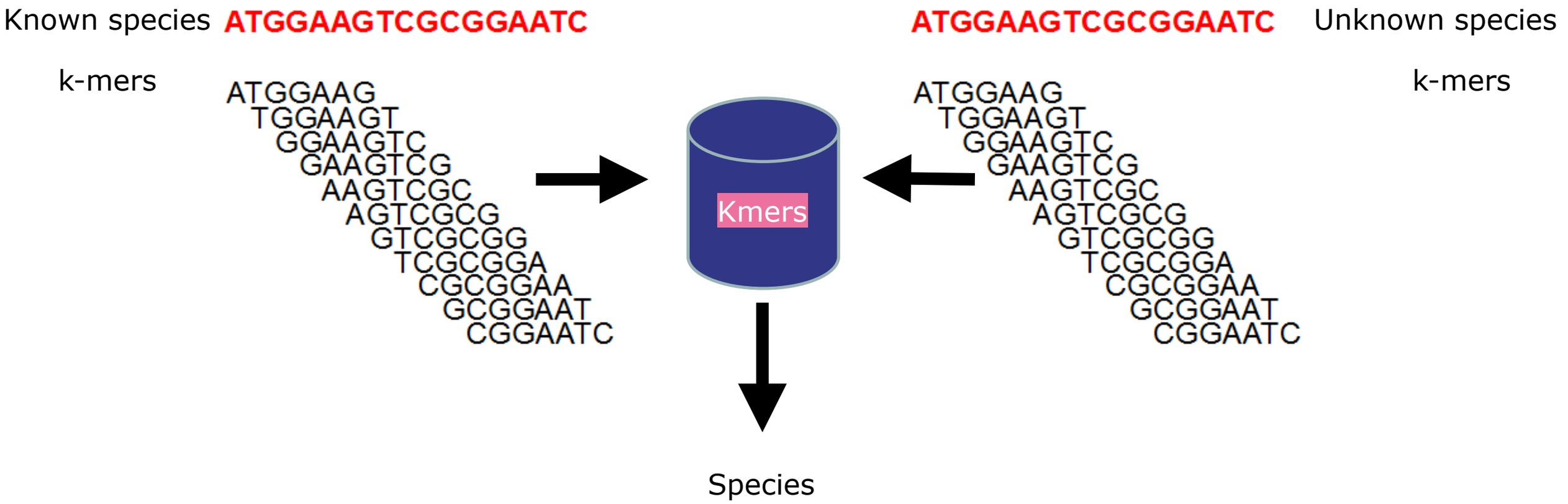
- A k-mer is a contiguous sequence of k bases
- k is any positive integer
- Sequences with high similarity must share k-mers

sequence **ATGGAAGTCGCGGAATC**

7 mers

ATGGAAG
TGGAAGT
GGAAGTC
GAAGTCG
AAGTCGC
AGTCGCG
GTCGCGG
TCGCGGA
CGCGGAA
GCGGAAT
CGGAATC

Species identification by K-mer



Study on how to use WGS data for bacterial species identification

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)

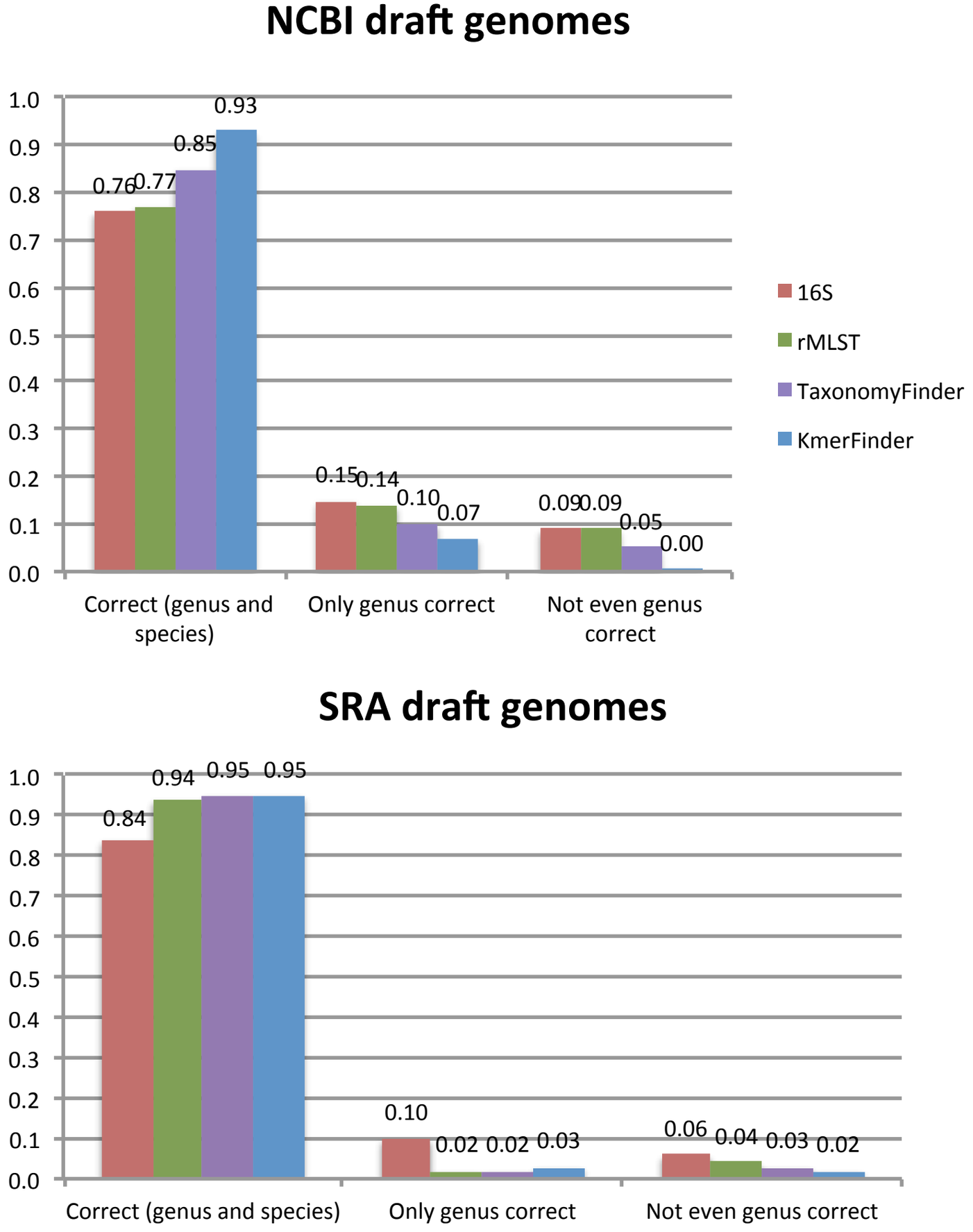
✧ NCBI draft genomes

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

✧ SRA draft genomes

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

Performance



Epidemiology

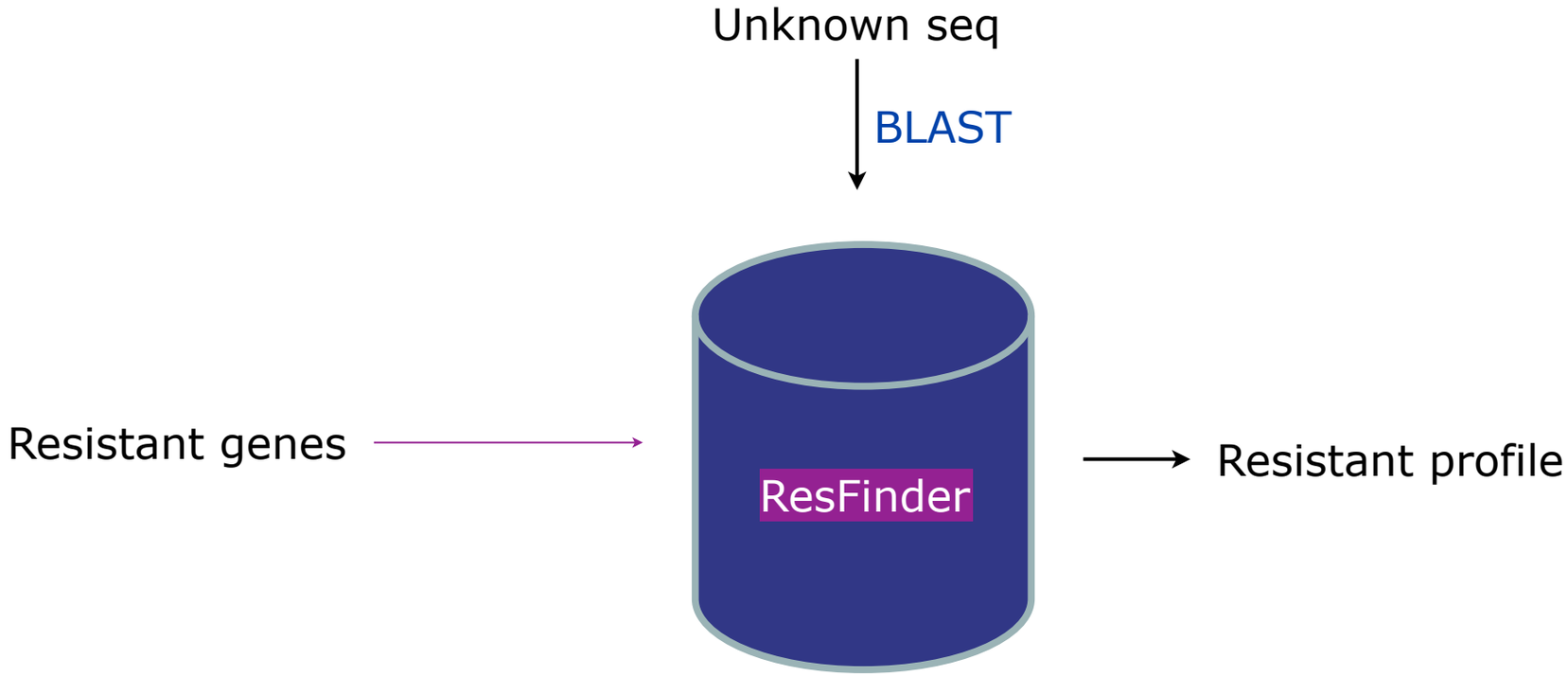
- The science that studies the patterns, causes, and effects of health and disease conditions in defined populations

- 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

Zankari E, Hasman H, Kaas RS, Seyfarth AM, Agersø Y, Lund O, Larsen MV,. Aarestrup FM. Genotyping using whole-genome sequencing is a realistic alternative to surveillance based on phenotypic antimicrobial susceptibility testing. J. Antimicrob Chemother. 2013 Apr;68(4):771-7.

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

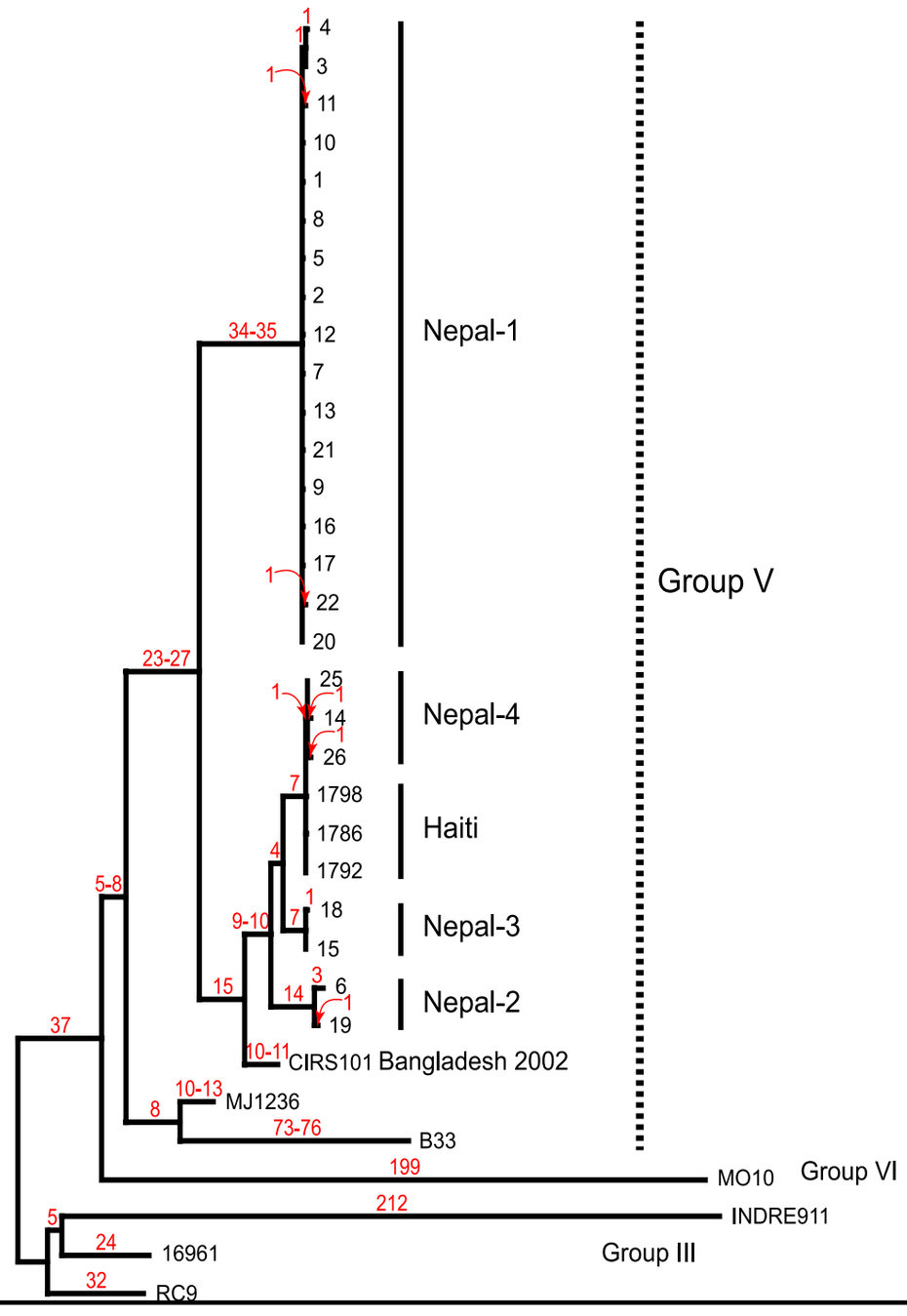
Case story - Results

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

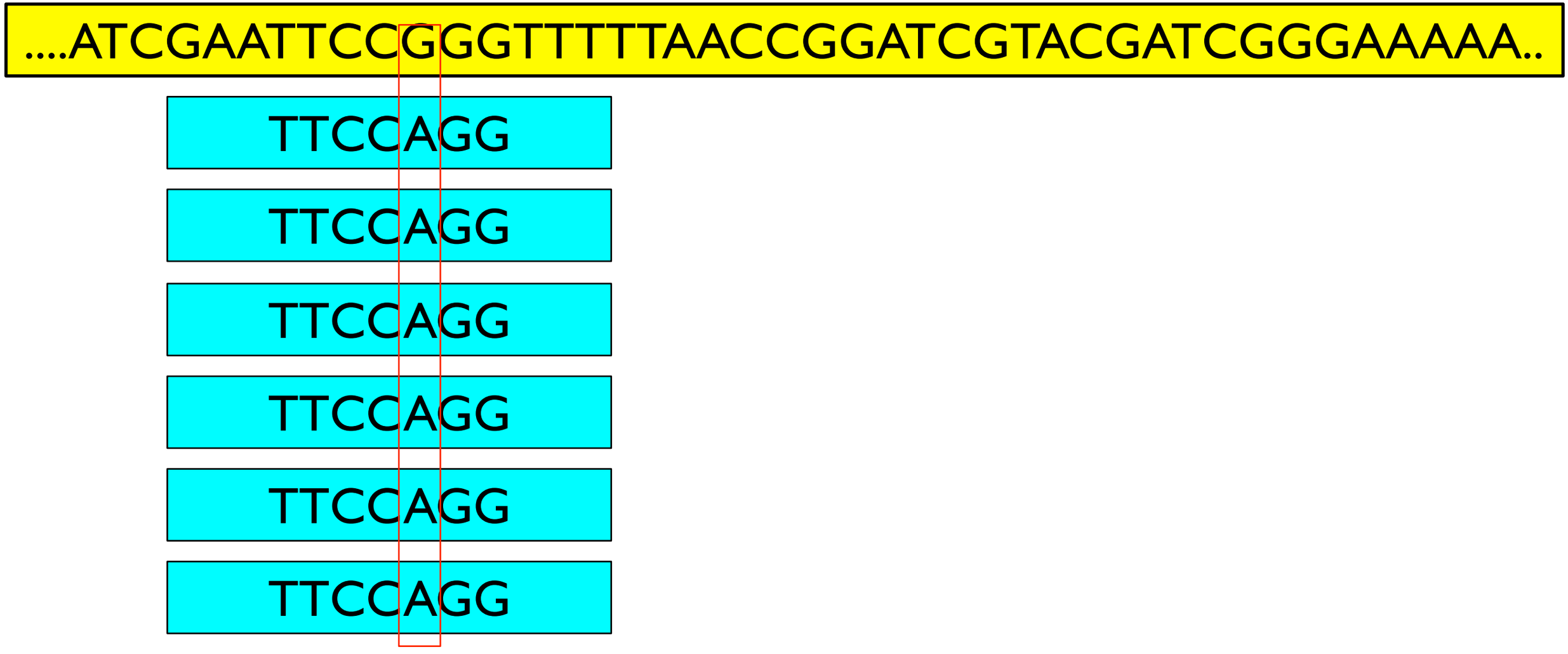
Case story - Results

- 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



SNPs detection

Variant calling format (VCF)

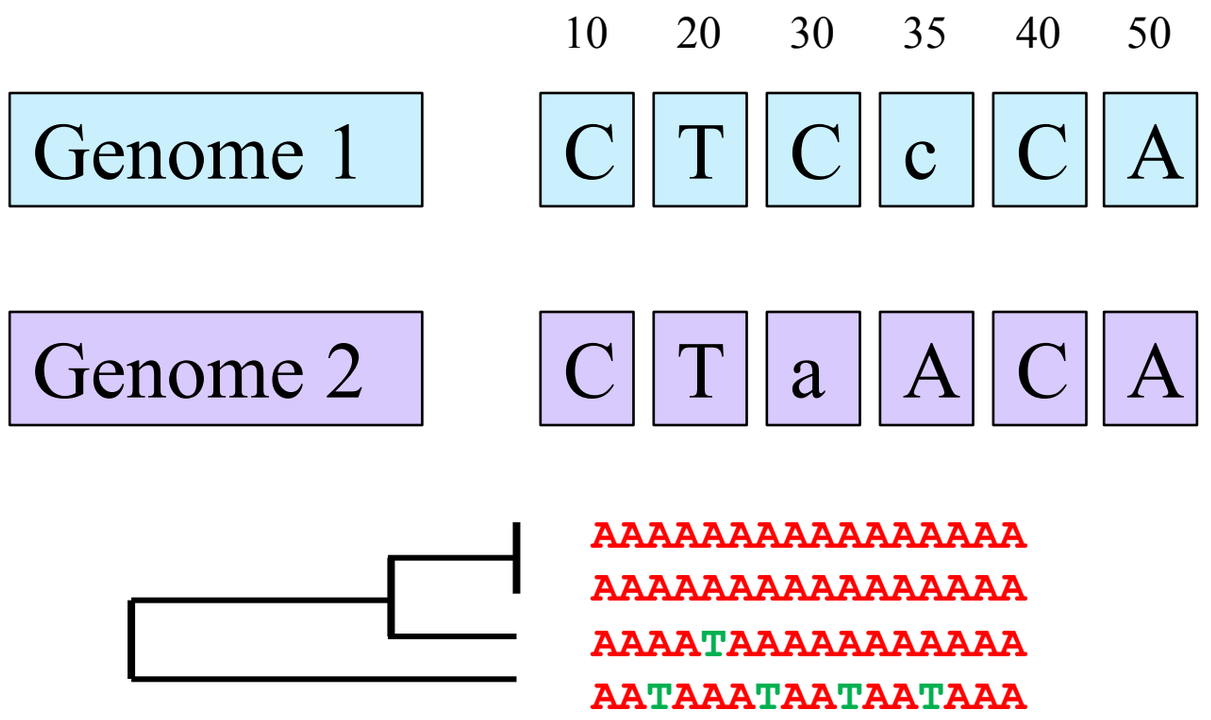
Genome 1	position	ref	change
Ref_genome	10	T	C
Ref_genome	20	C	T
Ref_genome	30	A	C
Ref_genome	40	A	C
Ref_genome	50	G	A

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

HELP DOCS BLOG LOGIN

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.

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.

DOCS HELP LOGIN

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.

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

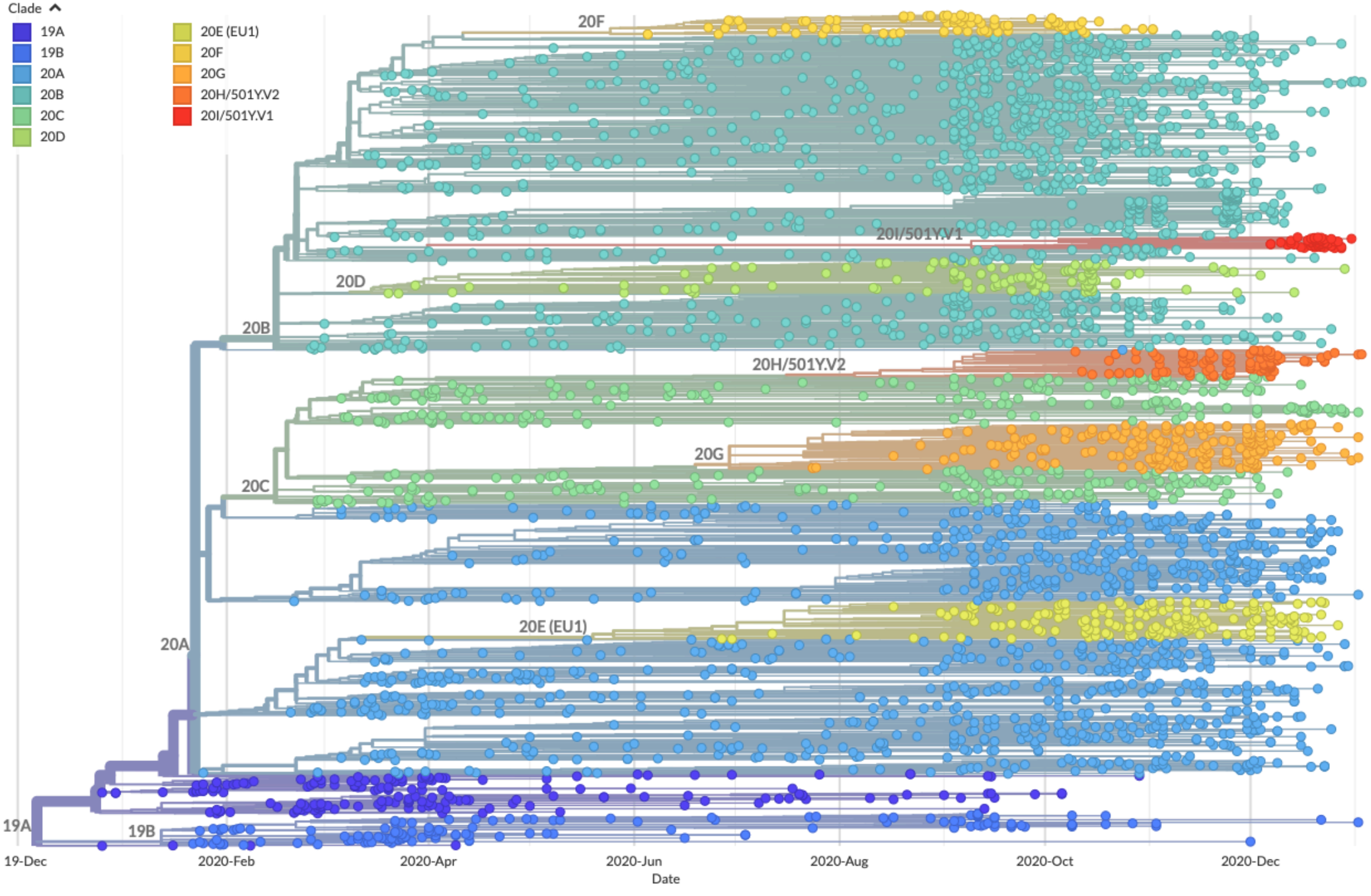
Sample Name

Second Tree

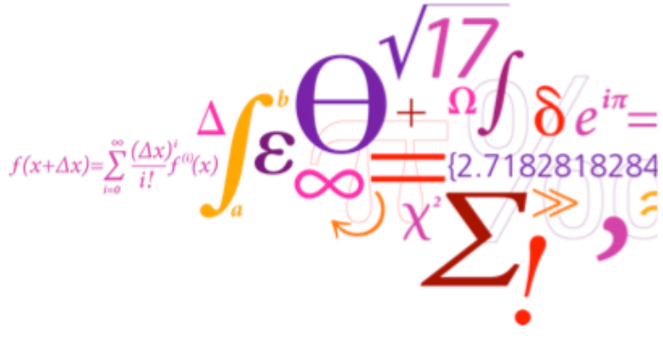
Select...

Phylogeny

ZOOM TO SELECTED RESET LAYOUT



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



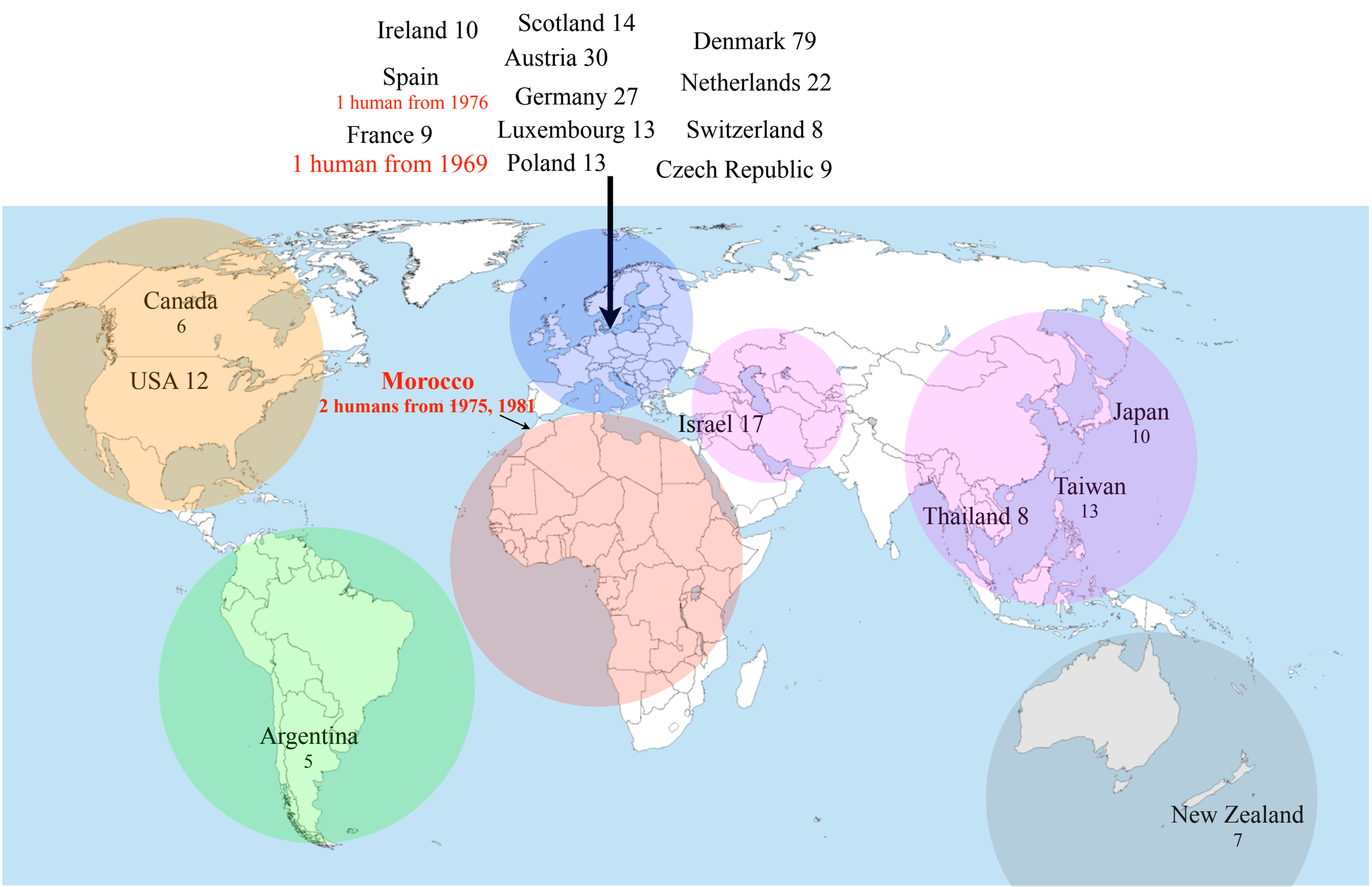
S. Typhimurium DT104

- 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]

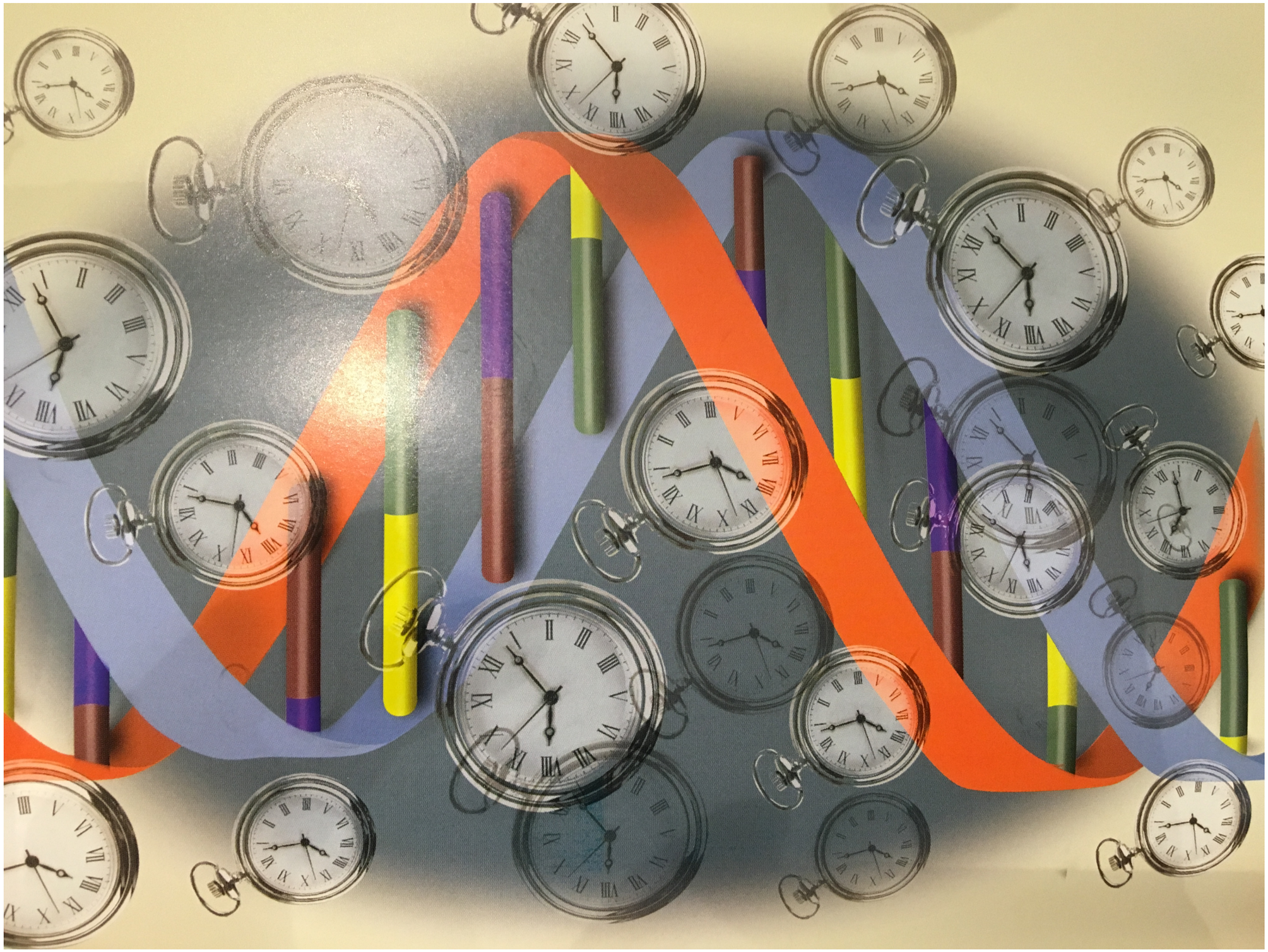


315 genomes

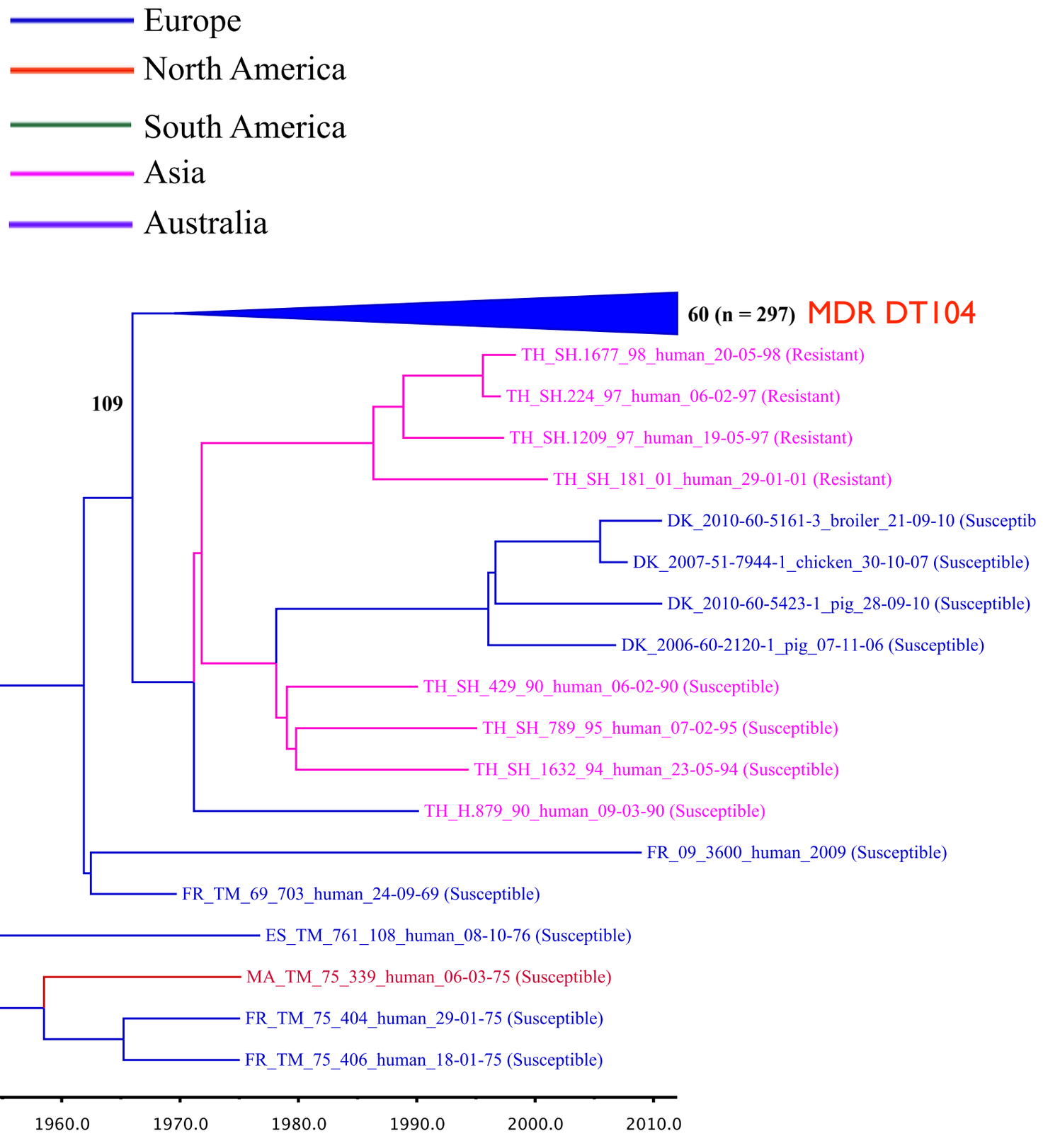
197 animal isolates

118 human isolates

BEAST



Global phylogeny of DT104

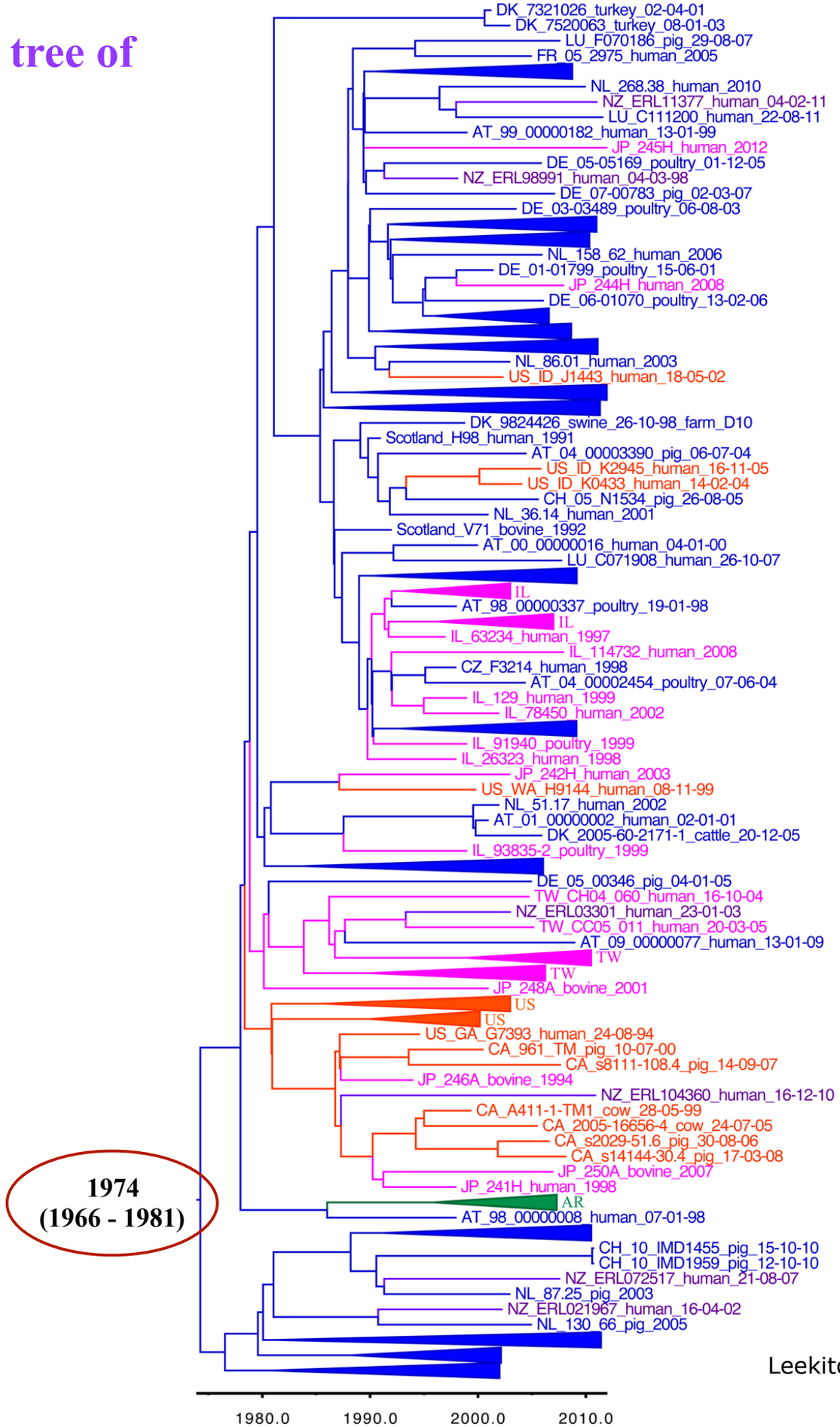


4619 variable sites

2.97 x 10^{-7} SNP/site/year ~ 1.47 SNP/year

Leekitcharoenphon P. *et al*, Appl Environ Microbiol. 2016

Temporal phylogenetic tree of MDR DT104



3621 variable sites

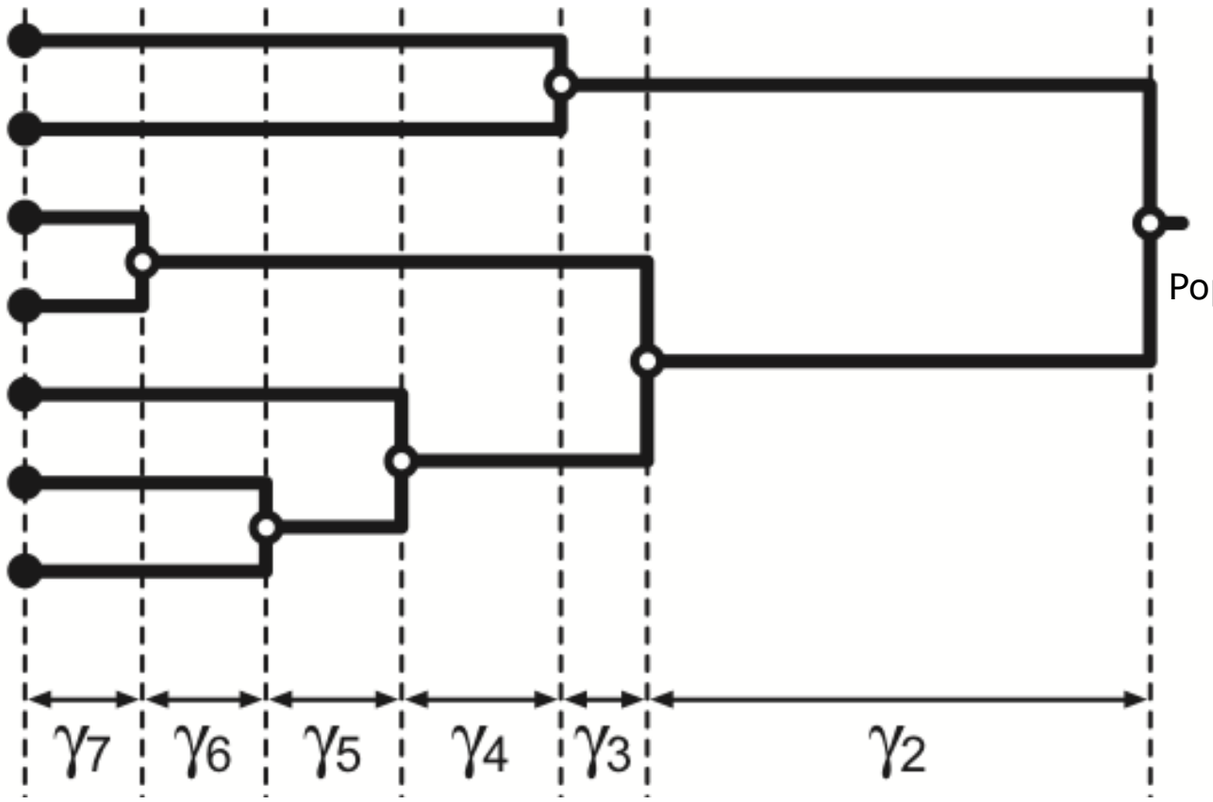
2.97E-7 SNP/site/year 1.47 SNP/year

Leekitcharoenphon P. *et al*, Appl Environ Microbiol. 2016

Evolution and global dissemination of DT104

Leekitcharoenphon P. *et al*, Appl Environ Microbiol. 2016

Demographic history using Bayesian skyline plot



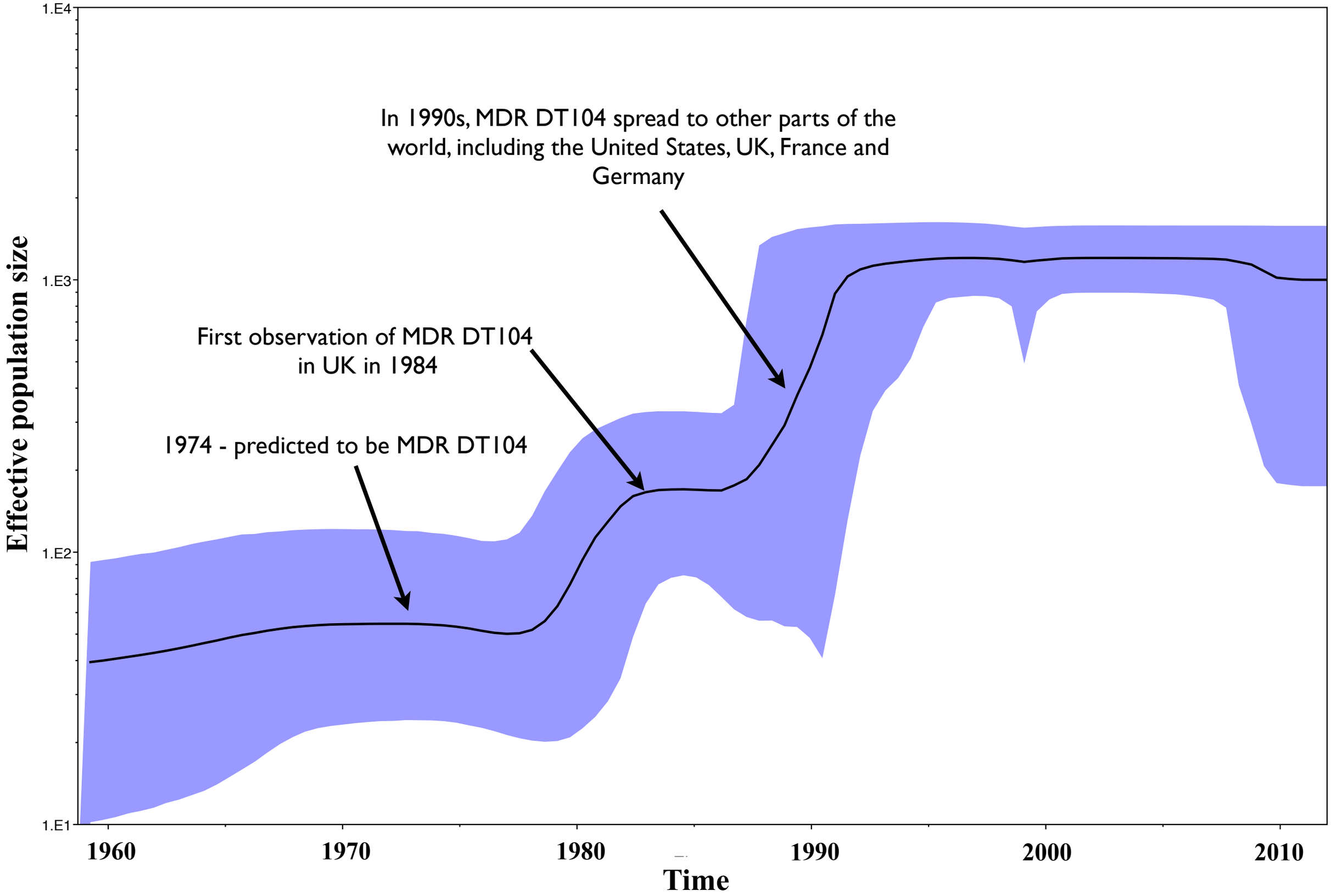
Population size

$$N_i = \gamma_i i(i-1)/2$$

interval size

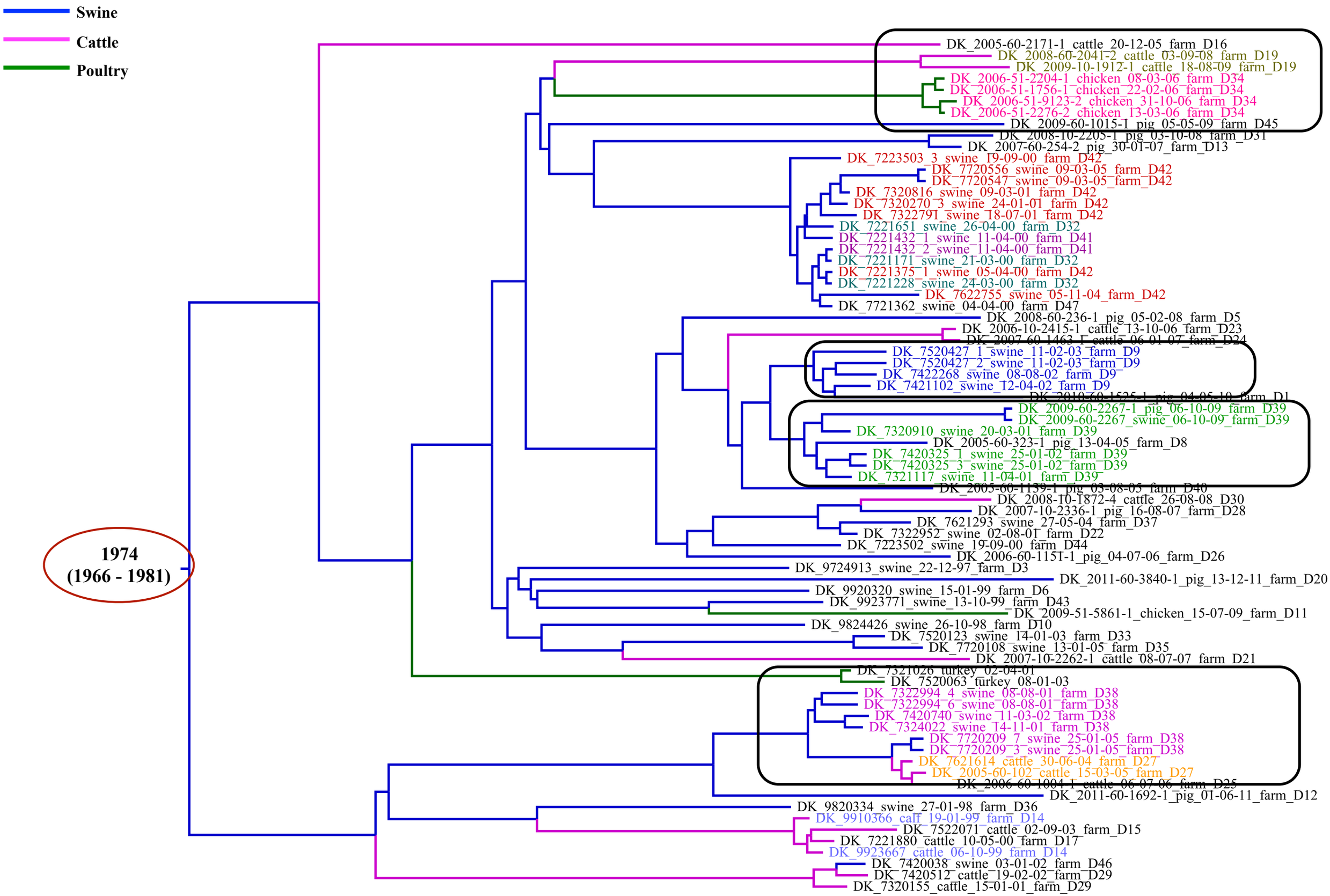
number of lineages (branches)

Demographic history of global DT104



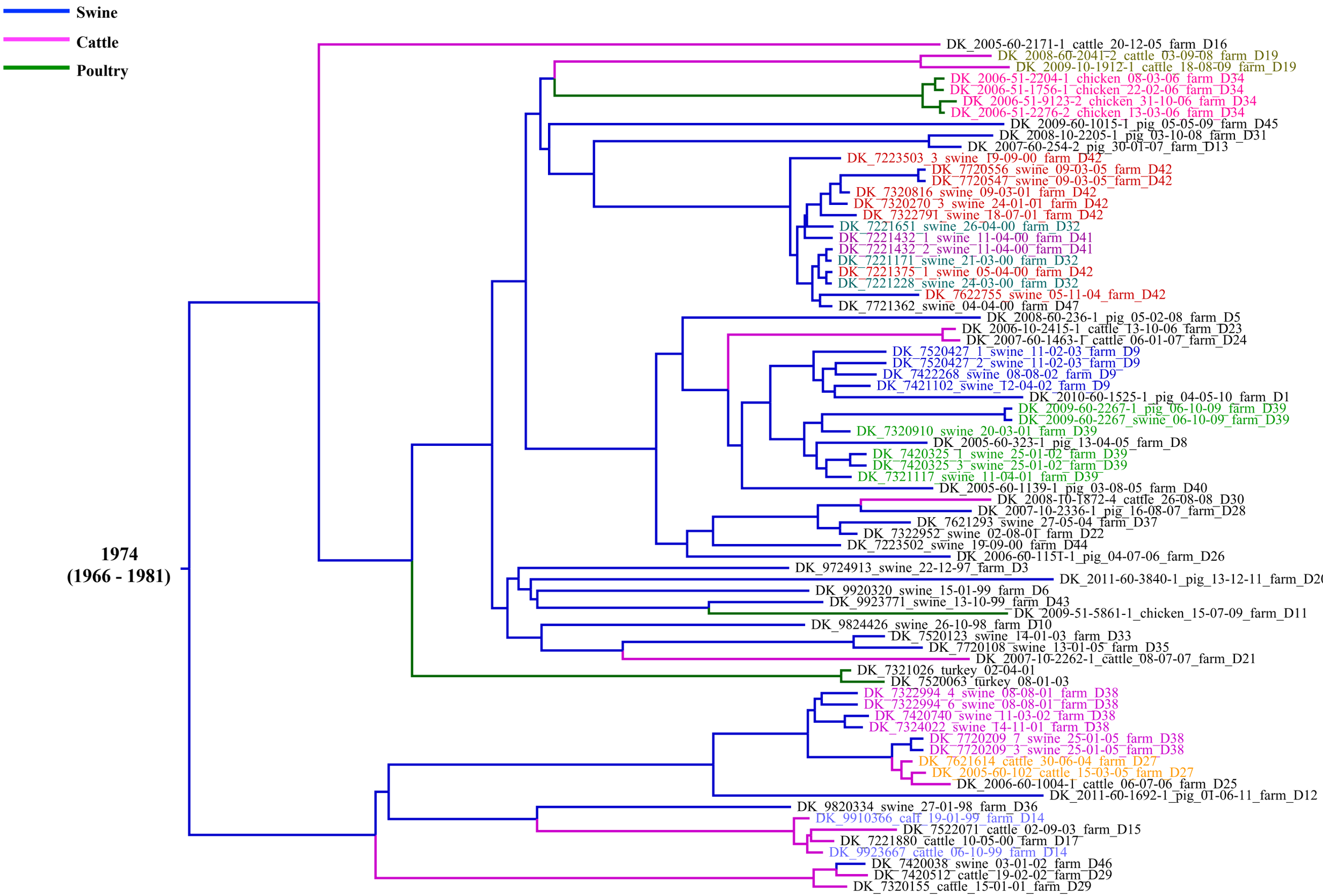
Leekitcharoenphon P. *et al*, Appl Environ Microbiol. 2016

Local phylogeny of DT104 (Denmark)

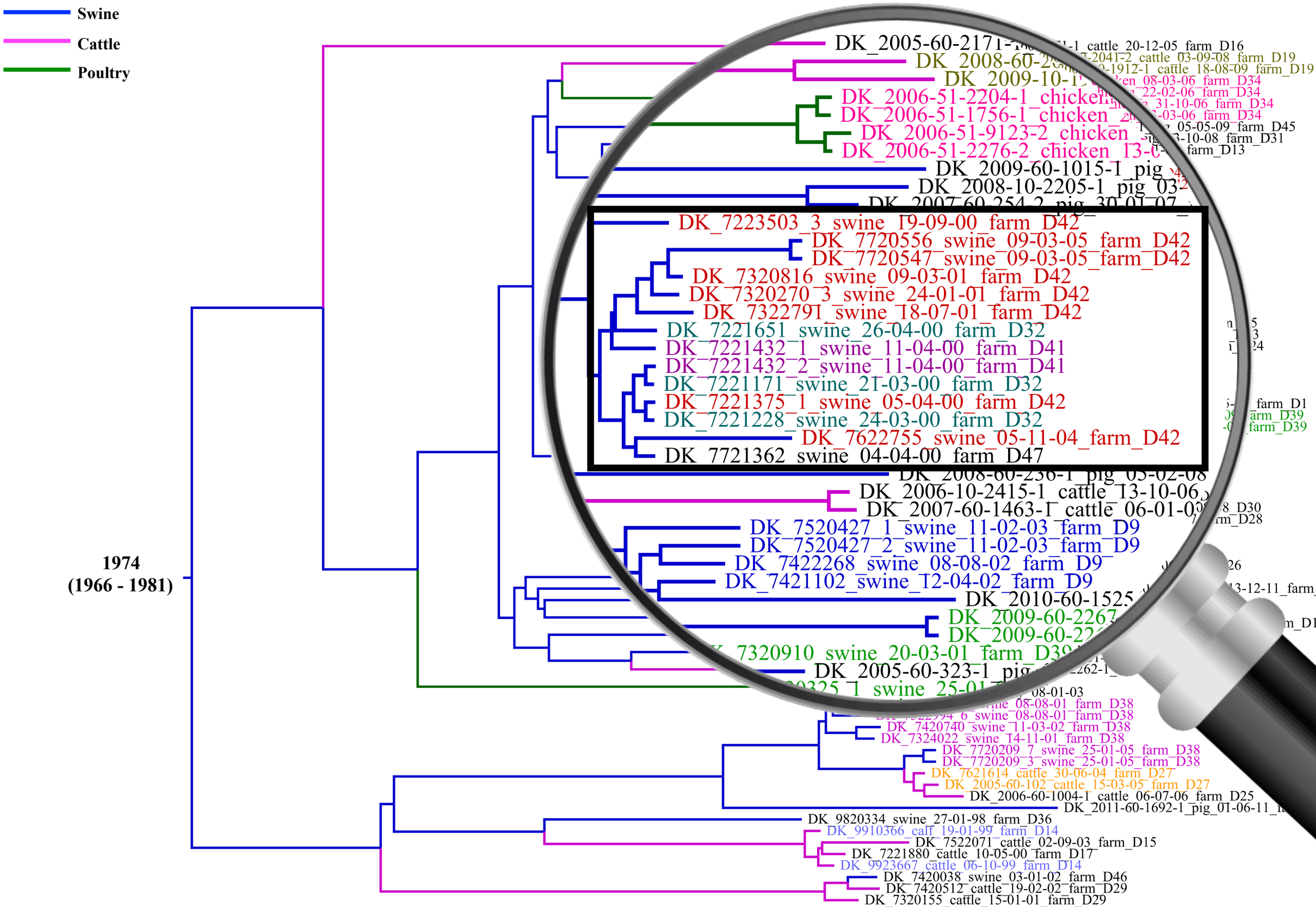


Leekitcharoenphon P. *et al*, Appl Environ Microbiol. 2016

Local phylogeny of DT104 (Denmark)



Local phylogeny of DT104 (Denmark)



File Edit View Share Window Help

Stop Recording 11 Wed Mar 12 16:29

Google Earth

Search

Get Directions History

Earth Gallery

Sign in

3/1960

1960 2011

Little Fisher Bank

Jutland Bank

Hjorring

Frederikshavn

Kattegatt

Gothenburg

Boras

Limfjord

Aalborg

Kattegat

Varberg

Halland

Viborg

Randers

Central Denmark

Denmark

Vesterhavet

Grenaa

Halmstad

Aarhus

Vejle County

Hesselø Bugt

Helsingborg

Skane

Kristians

Syddanmark

Odense

Great Belt

Region Zealand

Copenhagen

Øresund

Lund

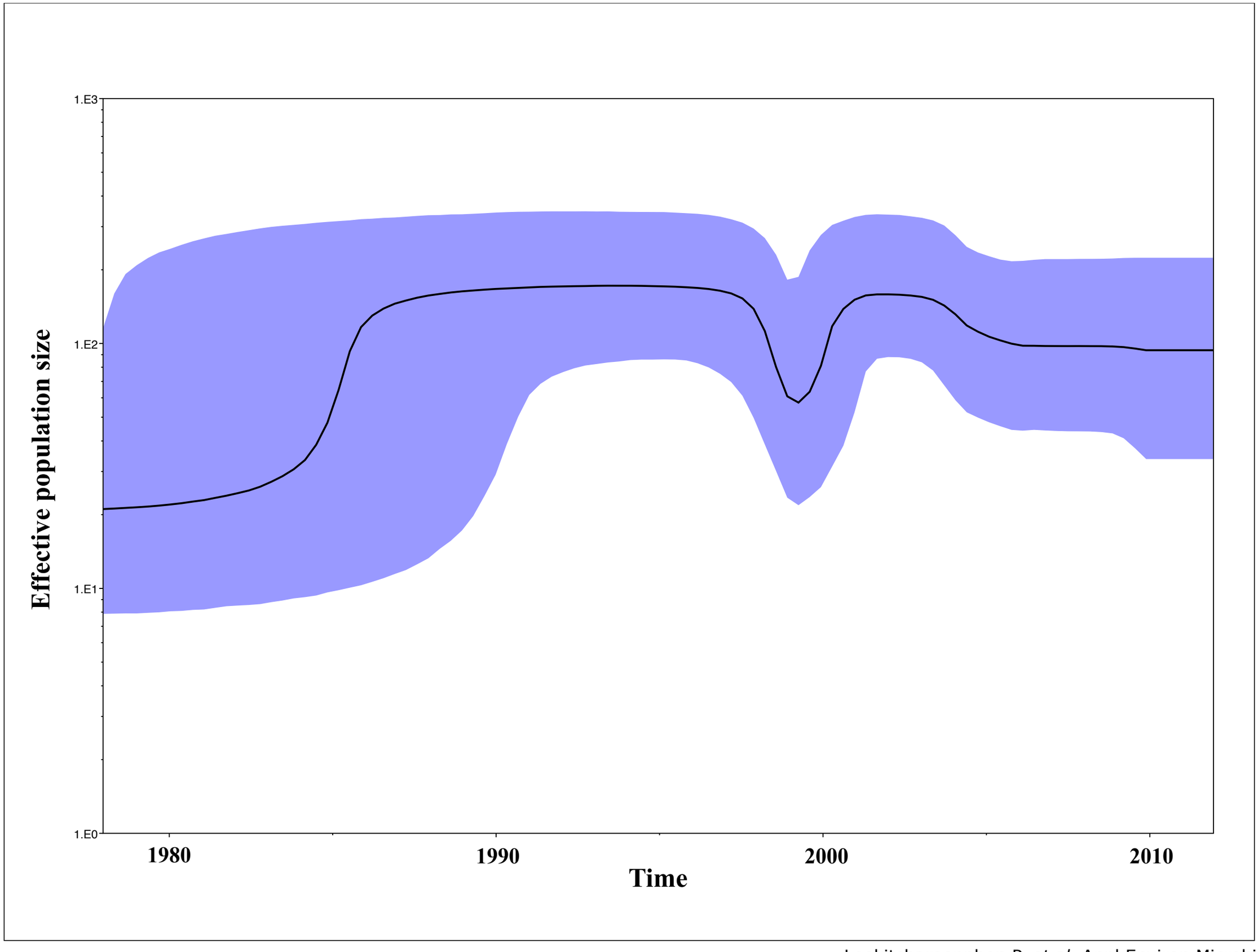
Sylt

D20 D21 D34 D38 D9 D39 D19 D17 D2 D3 D18 D4 D7 D28 D46 D29 D37 D22 D30 D45 D44 D1 D11 D16 D15 D32 D33 D14 D42 D35 D27 D24 D41 D23 D40 D26 D25 D8 D36 D47 D10 D12 D31 D5 D13 D6

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

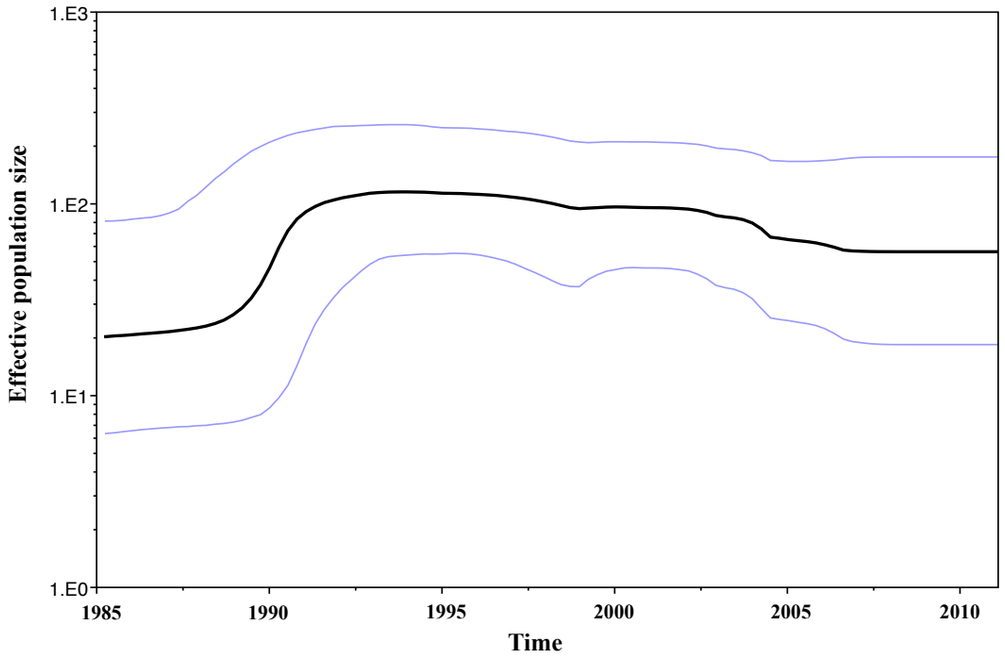
Google earth

Demographic history of Danish MDR DT104

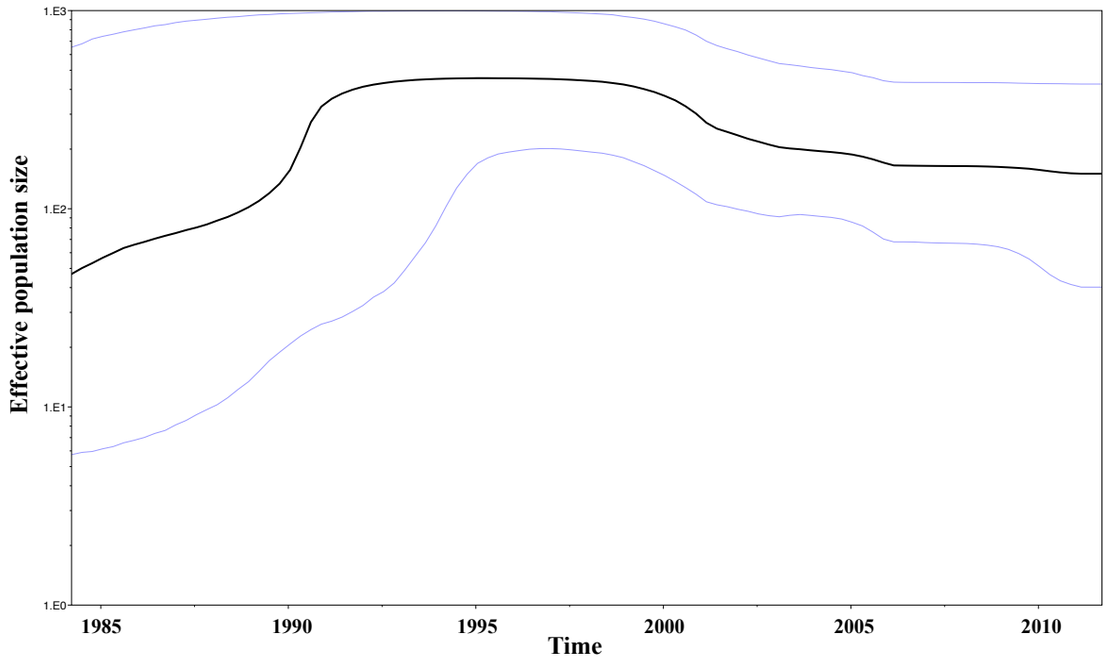


Leekitcharoenphon P. *et al*, Appl Environ Microbiol. 2016

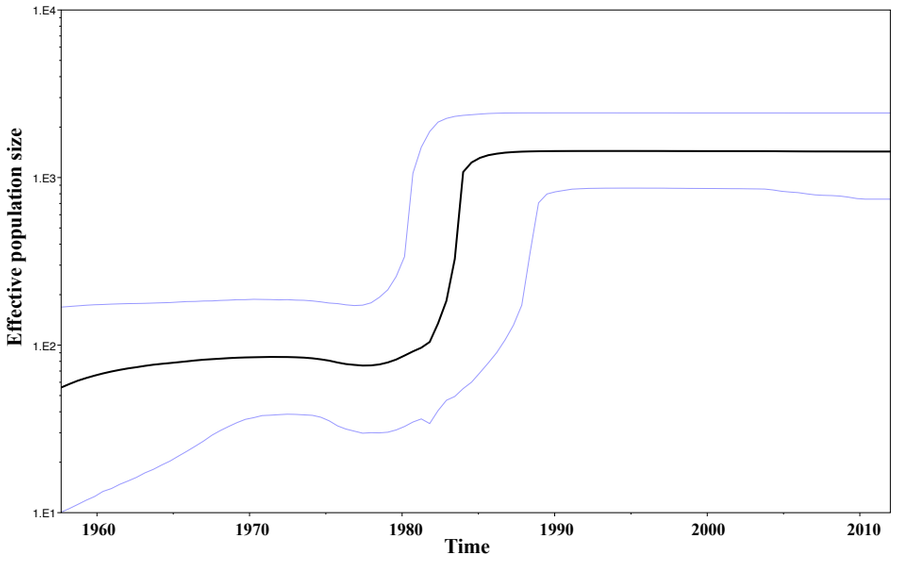
Cattle



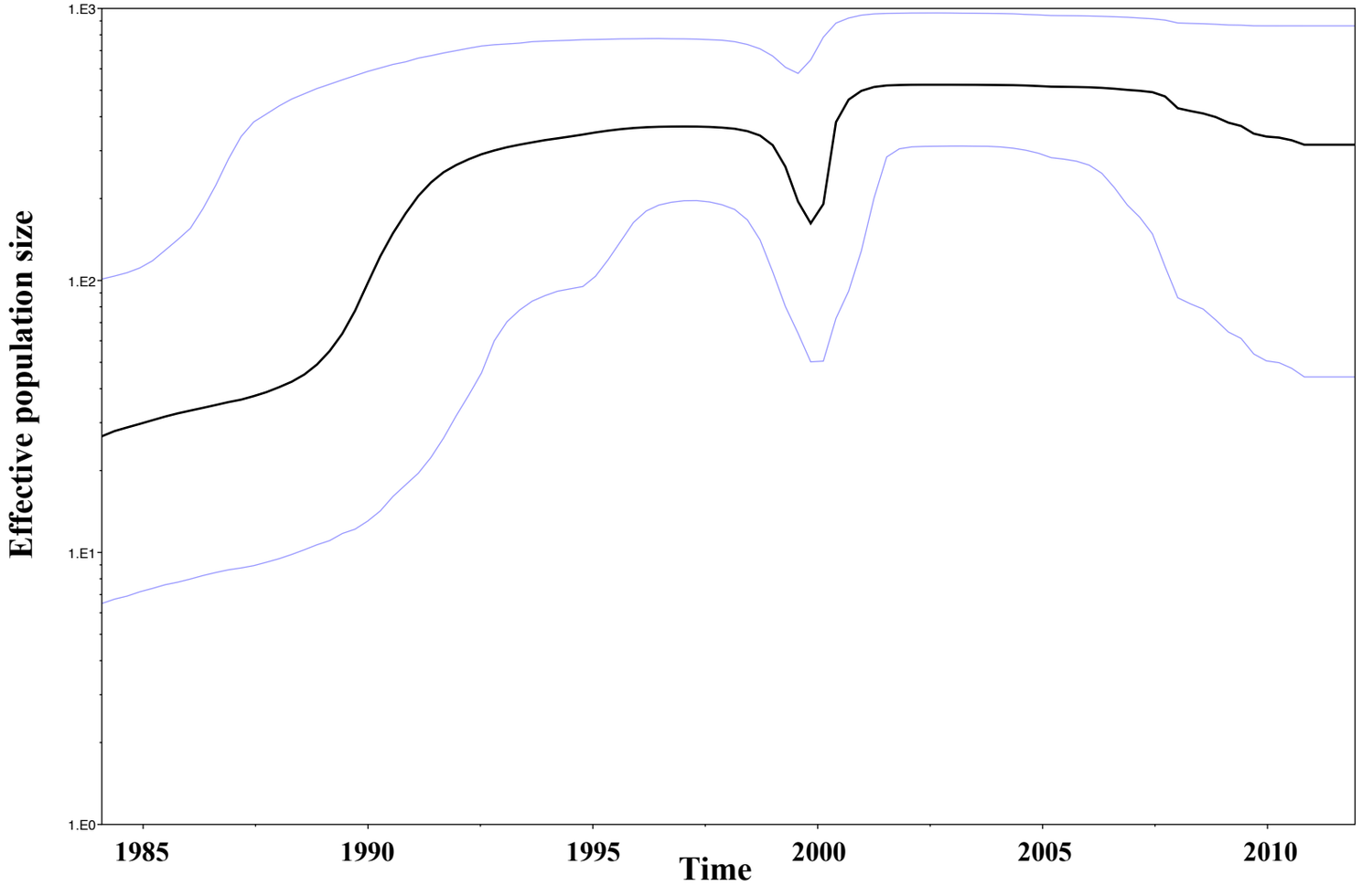
Poultry



Human



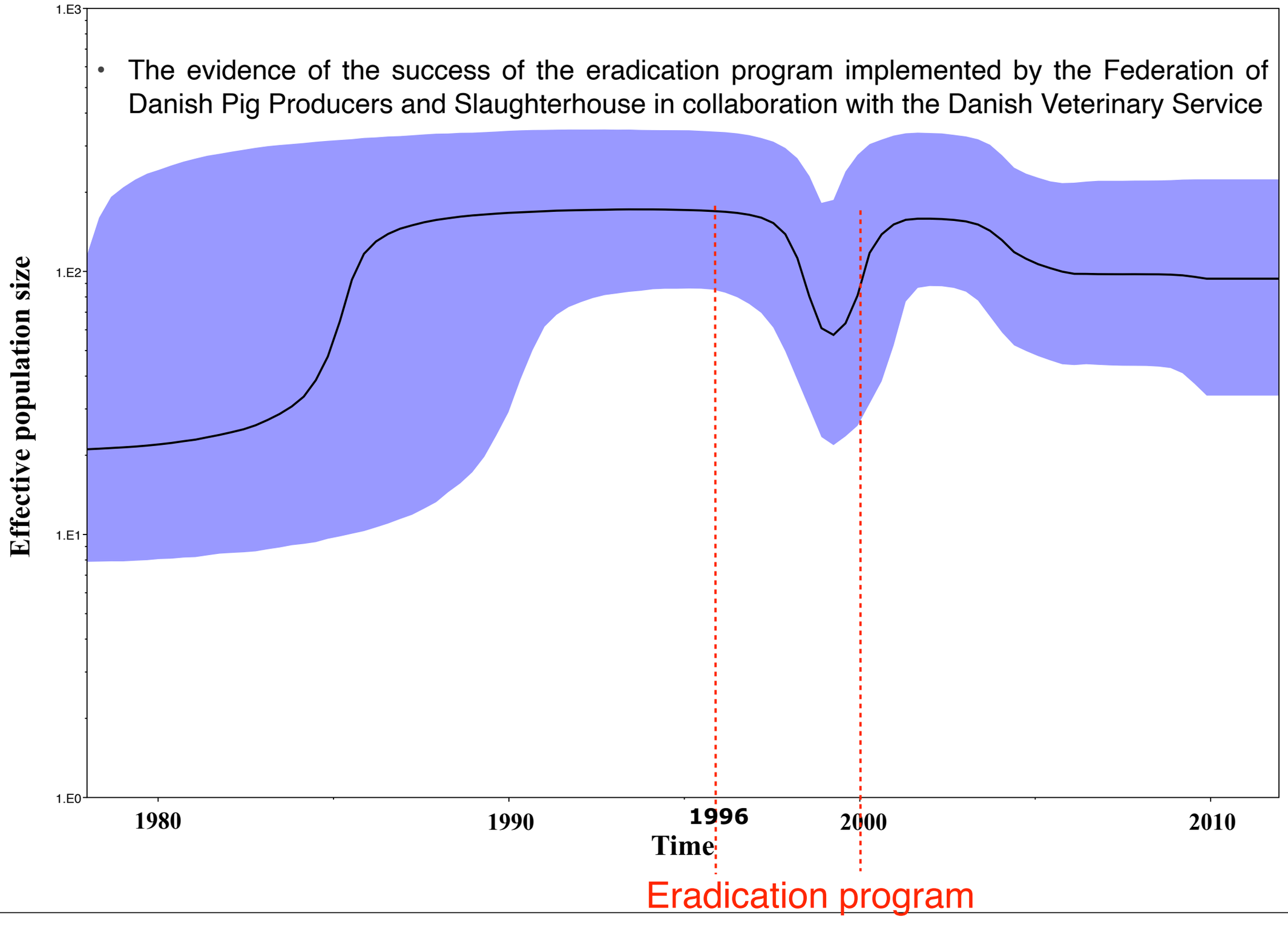
Swine



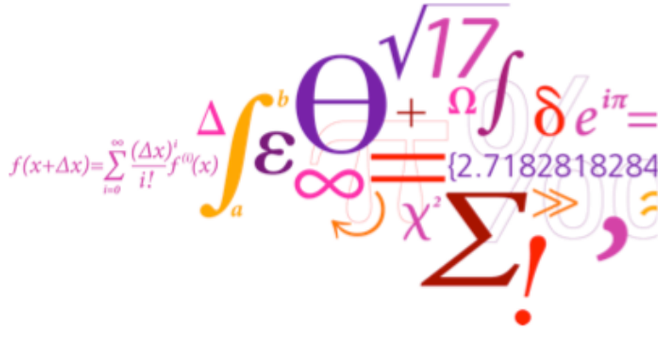
Demographic history of Danish MDR DT104

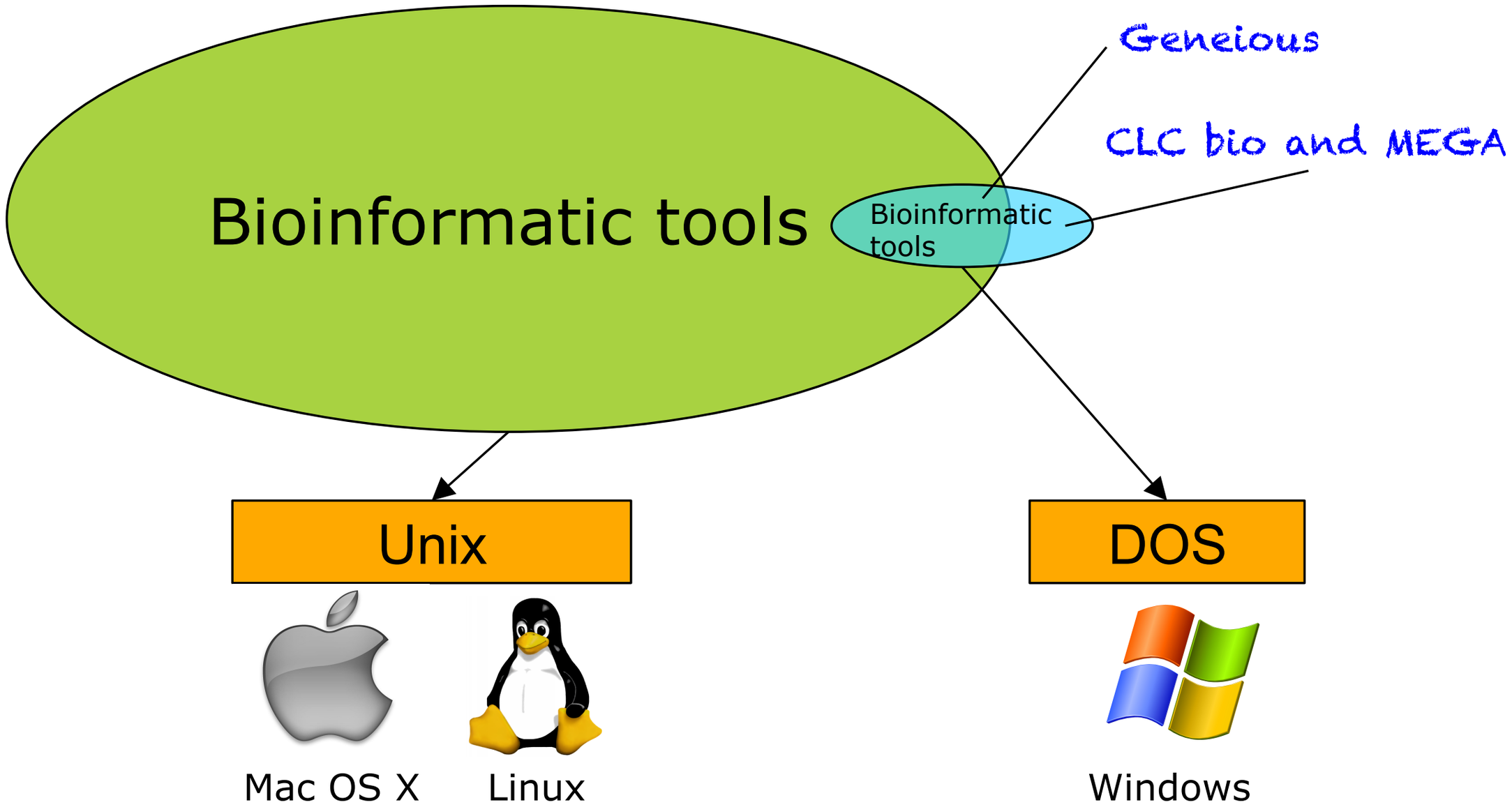
- The program aimed to eradicate MDR DT104 from infected pig herds by depopulation of pig herds, cleaning and disinfection of building before repopulation with pigs free from DT104

- 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




Bioinformatic tools
Bioinformatic tools
Geneious
CLC bio and MEGA
Unix
DOS
Mac OS X
Linux
Windows

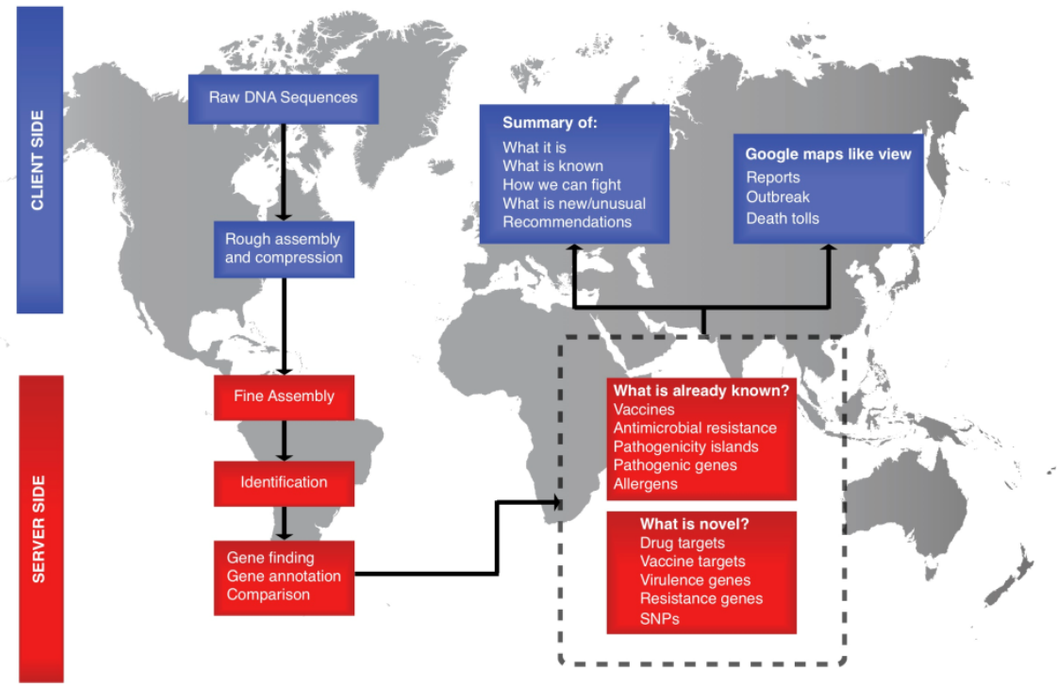
Preparing for Global Surveillance - Center for Genomic Epidemiology

- Provide a proof of concept of combining bioinformatics with global epidemiology in real-time

- Provide foundation for web based solutions (plug and play tool)

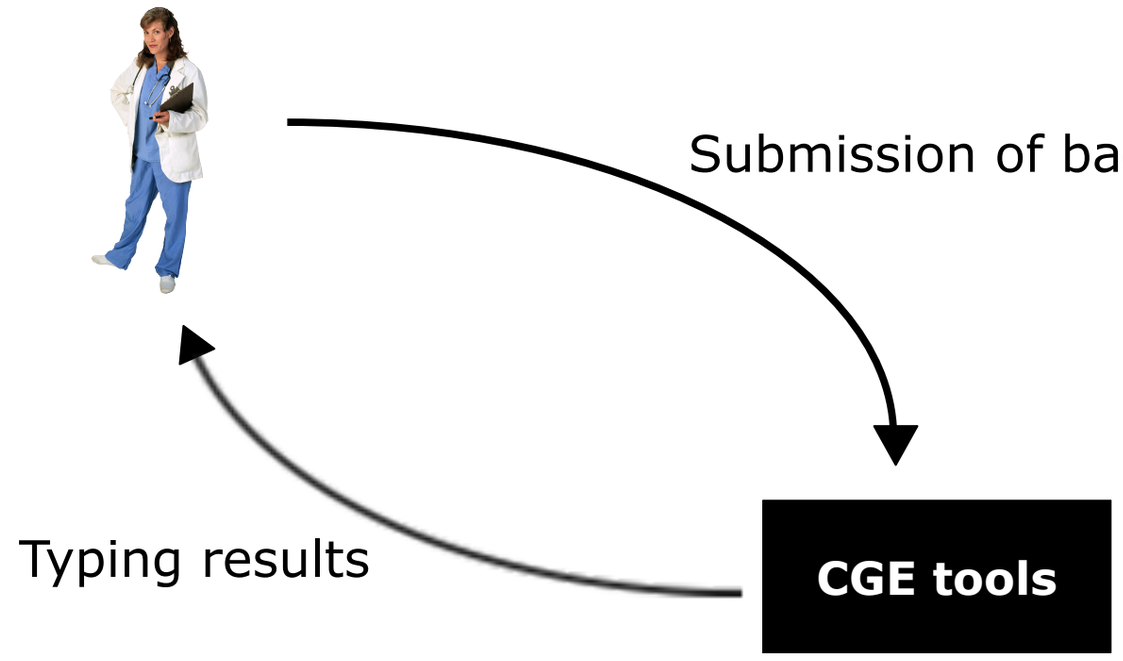
    - What is it
    - How dangerous is it
    - Have we seen it before
    - With what can it be treated

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



www.genomicepidemiology.org

CGE tools


Submission of bacterial sequence
Typing results
CGE tools

http://www.genomicepidemiology.org

Center for Genomic Epidemiology

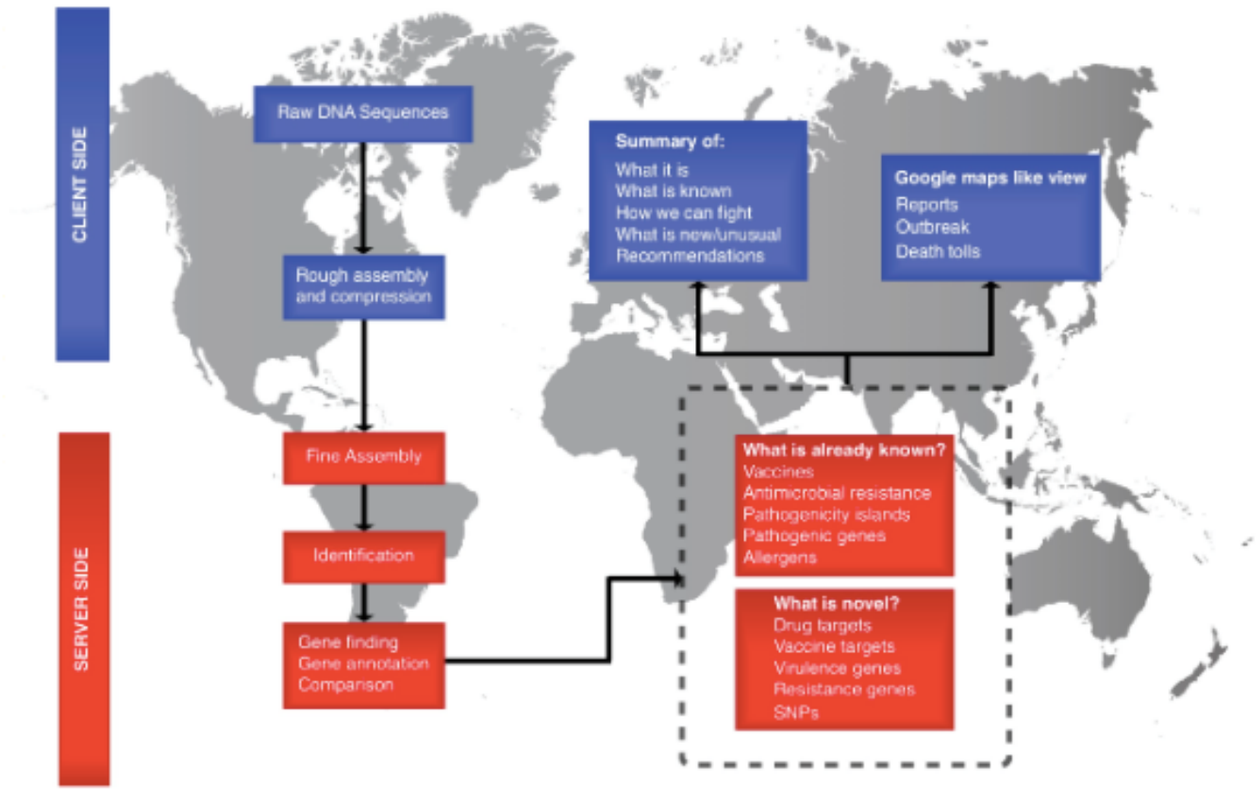
Services

Phenotyping:

- Identifcation of acquired antibiotic resistance genes. ResFinder
- Prediction of a bacteria's pathogenicity towards human hosts. PathogenFinder
- Identifcation 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
New Reset Login

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

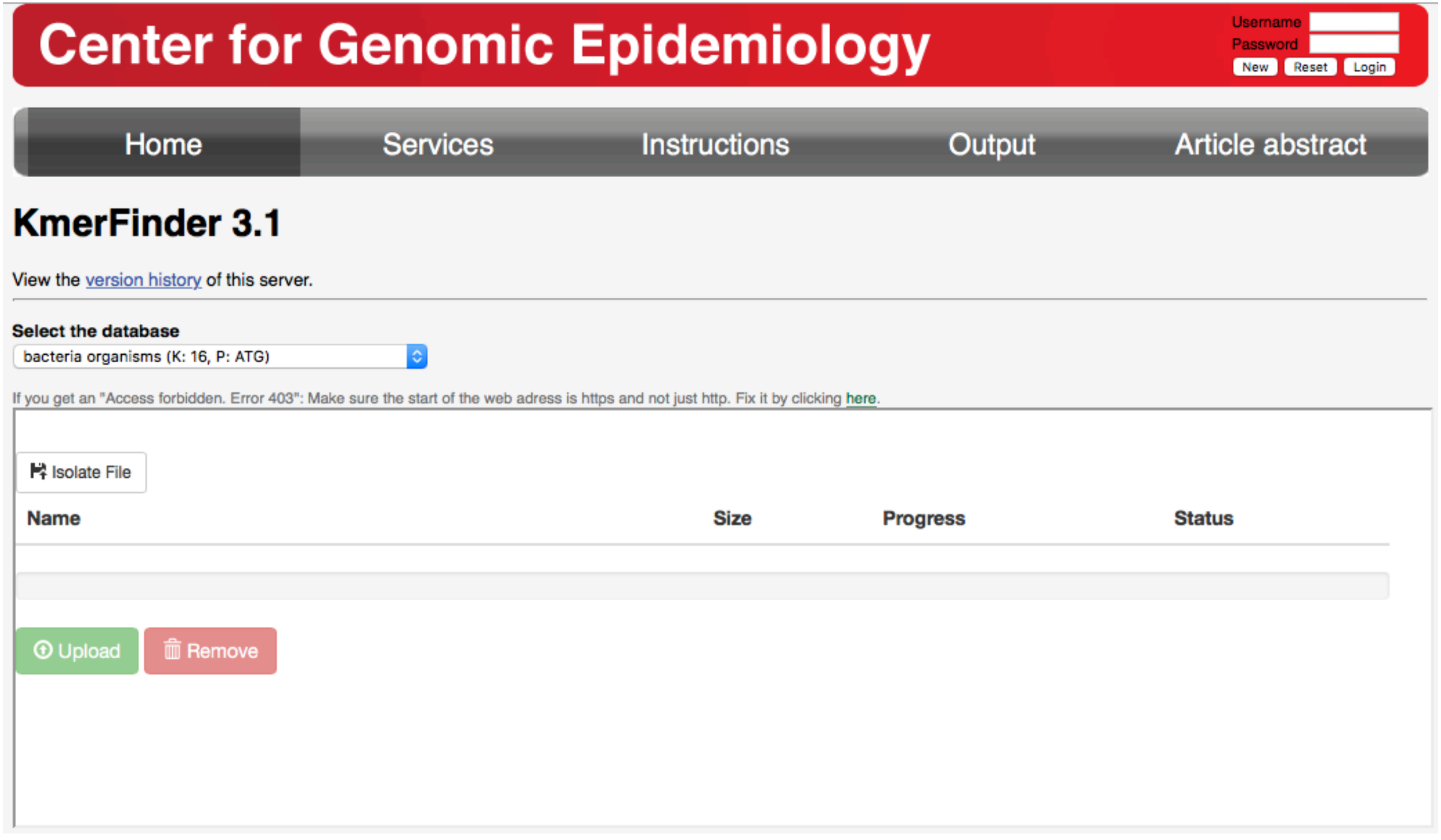
snpTree (Works)
NDtree (Works)
CSIPhylogeny (Works)
TreeViewer (Works)

Other

Assembler (Works)
ENAUploader (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.cbs.dtu.dk/services/KmerFinder/



Center for Genomic Epidemiology

Username
Password
New Reset Login

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 adress is https and not just http. Fix it by clicking here.

Name	Size	Progress	Status

Upload Remove

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: Notify me via 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.cbs.dtu.dk/services/ResFinder/

Center for Genomic Epidemiology

Username
Password
New Reset Login

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 adress is https and not just http. Fix it by clicking here.

Isolate File

Name	Size	Progress	Status

Upload Remove

Center for Genomic Epidemiology

Username
Password
New Reset Login

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
New Reset Login

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 pile
Password ••••••••
New Reset Login

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 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
90 %

Select minimum length
60 %

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 adress is https and not just http. Fix it by clicking here.

Isolate File

Name	Size	Progress	Status
strain01.fasta	4.63 MB		

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	15848289

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	21653764

Center for Genomic Epidemiology

ResFinder-2.1 Server - Results

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

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

Colistin
No resistance genes found.

www.eurosurveillance.org

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 [1], AM Hammerum [1], F Hansen [1], RS Hendriksen [2], B Olesen [3], Y Agersø [2], E Zankari [2], P Leekitcharoenphon [2], M Stegger [14], RS Kaas [2], LM Cavaco [2], DS Hansen [3], FM Aarestrup [2], RL Skov [1]

1. 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.cbs.dtu.dk/services/CSIPhylogeny/

Center for Genomic Epidemiology

Username
Password
New Reset Login

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 suported.

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.

Choose File 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
New Reset Login

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 suported.

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.

Choose File 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
New Reset Login

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 suported.

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.

Choose File 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

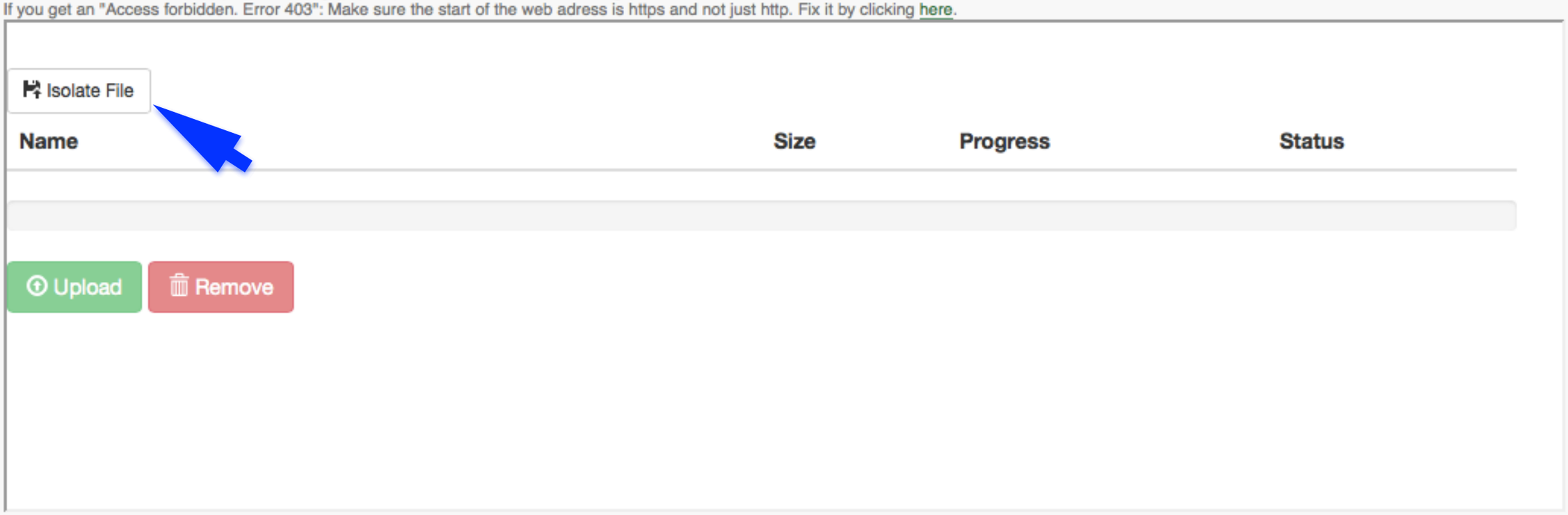
Select min. SNP quality
30

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 adress is https and not just http. Fix it by clicking here.



***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 adress is https and not just http. Fix it by clicking here.

Isolate File

Name	Size	Progress	Status
Salmonella-spp-02-03-002.fna	4.80 MB		
Salmonella-spp-02-03-008.fna	4.81 MB		
Salmonella-spp-05-102.fna	4.81 MB		
Salmonella-spp-07-022.fna	4.80 MB		

Upload Remove

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

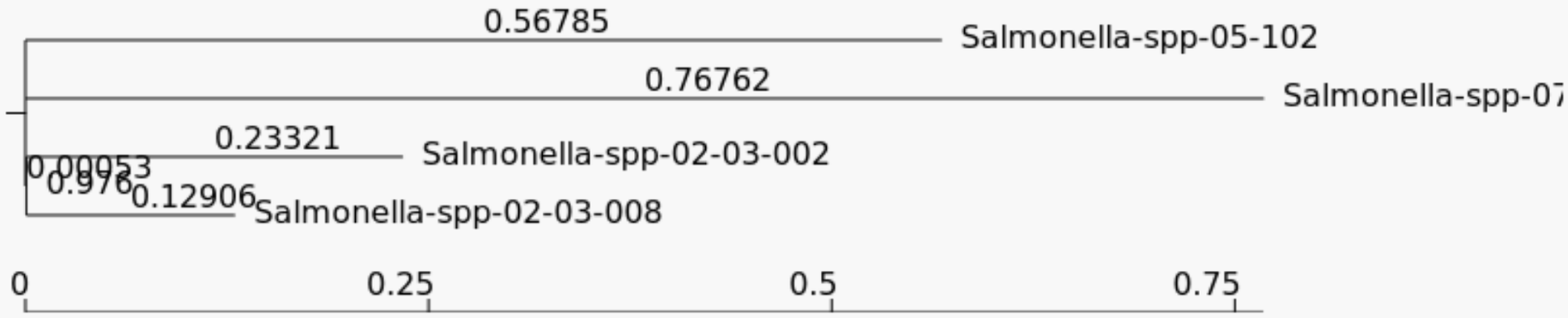
Confidentiality:
The sequences are kept confidential and will be deleted after 48 hours.

Center for Genomic Epidemiology

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).



Dowload phylogeny as: Newick PDF SVG

Download the filtered SNP calls in Variant Calling Format (VCF):
Note: VCF files are compressed with gzip.
VCF files

Download matrix of SNP pair counts:
Dowload matrix as: TXT EPS

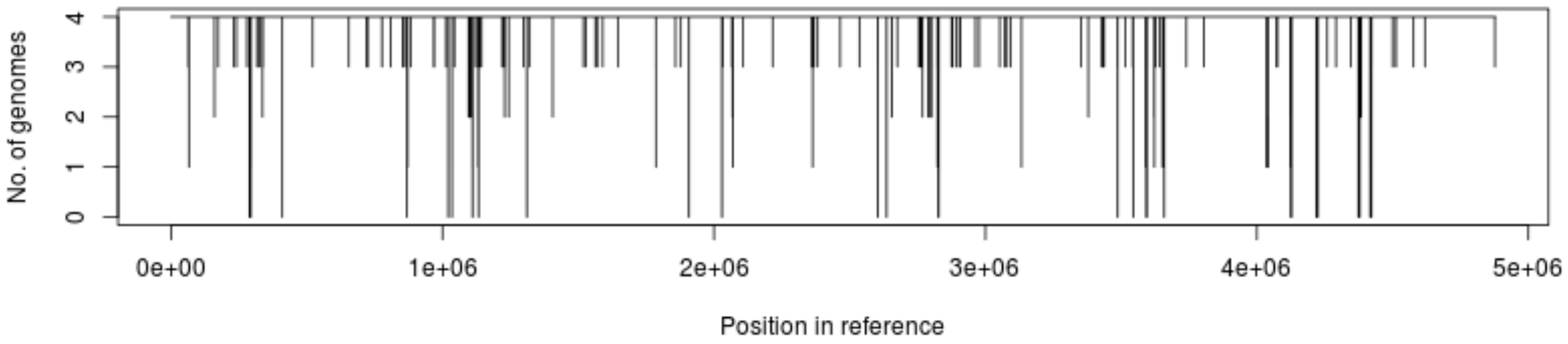
Dowload SNP alignment: FASTA

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

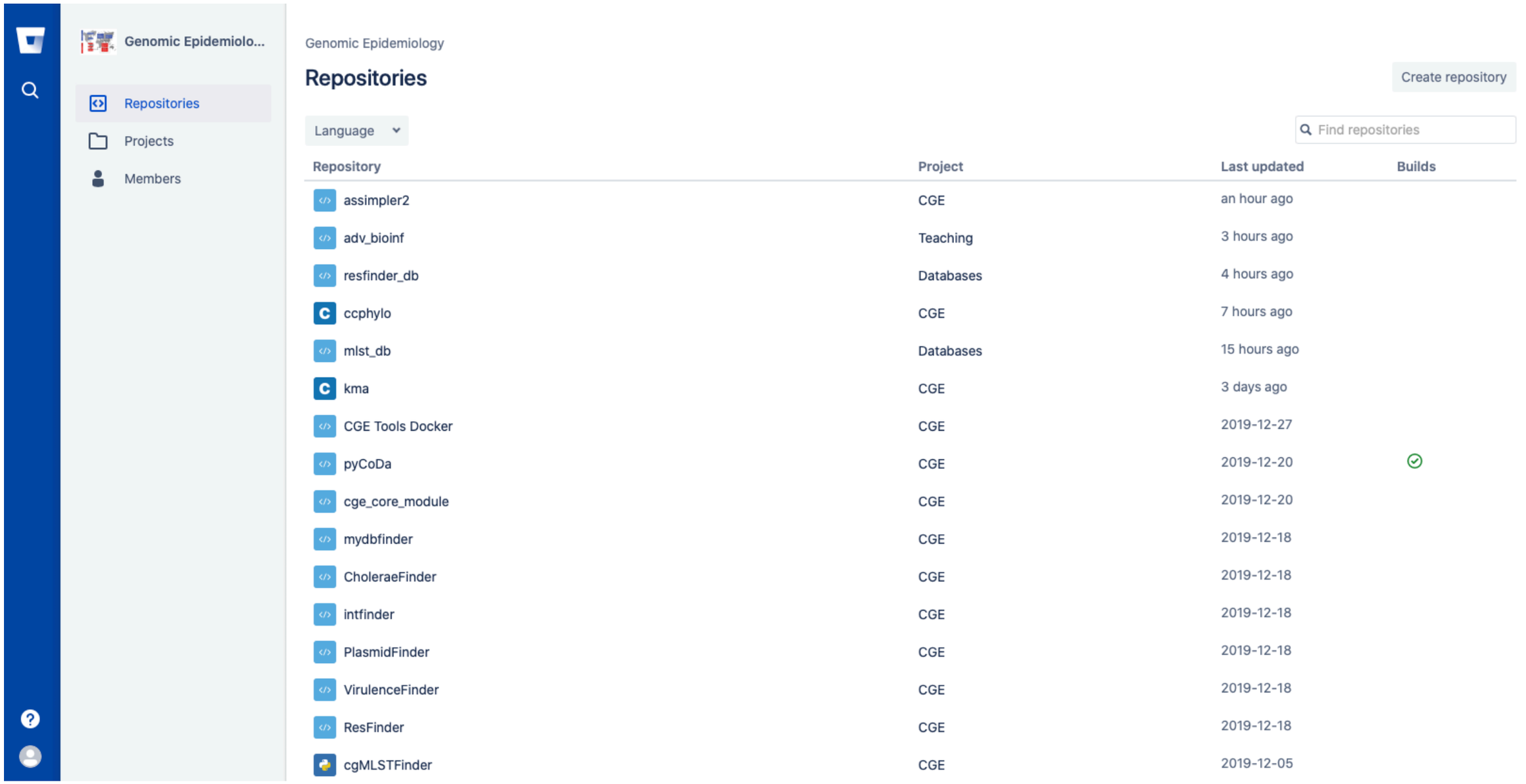
Genomes covering each Position

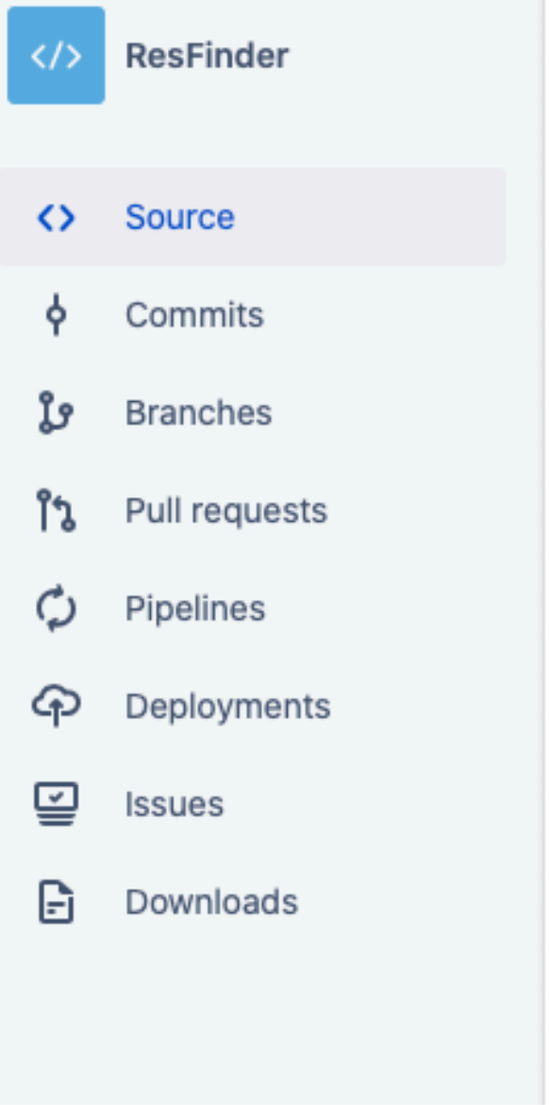


Download plot:

PDF

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





Genomic Epidemiology / CGE

ResFinder

master

Filter files

/

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 extented output with allignment files,
                       "template and query hits in fasta and a tab
                       "seperated file with gene profile results
-q --quiet
```

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

coursera | Catalog | Search catalog | For Enterprise | Pimlapas

- Overview
- Syllabus
- FAQs
- Creators
- Ratings and Reviews

Whole genome sequencing of bacterial genomes – tools and applications

Go to Course

Already enrolled

Financial Aid is available for learners who cannot afford the fee.
Learn more and apply.

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 us 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)

23259 Whole genome analysis in diagnostic microbiology

Course information

Danish title	Helgenom analyser i mikrobiologisk diagnostik
Language of instruction	English
Point(ECTS)	5
Course type	MSc Offered as a single course
Schedule	Spring F3A (Tues 8-12)
Location	Campus Lyngby
Scope and form	Lectures and computer-based exercises. It is mandatory to hand in a report about the exercises.
Duration of Course	13 weeks
Date of examination	F3A, Poster presentation on the last day of the course
Type of assessment	Written and oral examination Written examination (counts 50%) as well as poster presentation of group based project (counts 50%). One overall grade will be given.
Exam duration	2 hours
Aid	All Aid
Evaluation	7 step scale , internal examiner
Previous Course	27683 and 36683
Not applicable together with	27683/ 36683

Genomic epidemiology for global surveillance AMR

Global surveillance

Whole genome sequencing vs Metagenomics

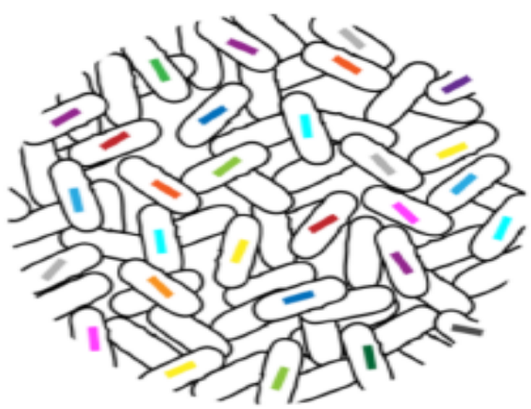
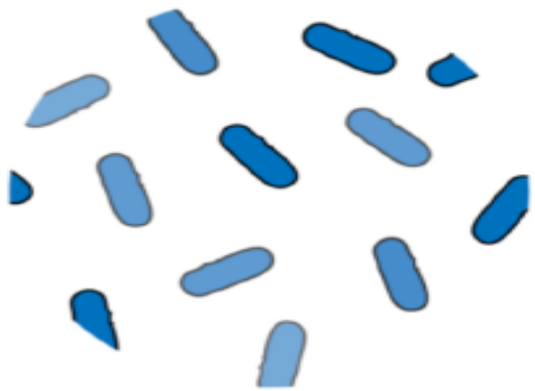
Whole genome sequencing

- 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

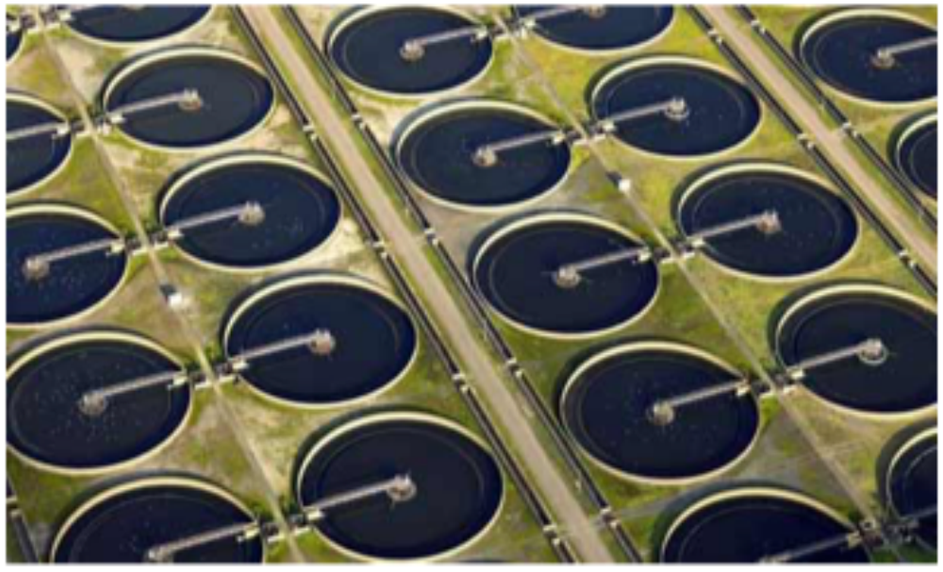
- Sequencing the DNA of the complex community without isolating the individual microorganisms (mixed of multiple organisms)

Microbial Genomics

16S rRNA gene profiling	Metagenomics	Whole Genome Sequencing (WGS)
		
DNA extraction	DNA extraction	Cultivation
↓	↓	↓
16S rRNA gene specific PCR		DNA extraction
↓	↓	↓
Sequencing	Sequencing	Sequencing

Environmental Chemicals, the Human Microbiome, and Health Risk: A Research Strategy, 2018

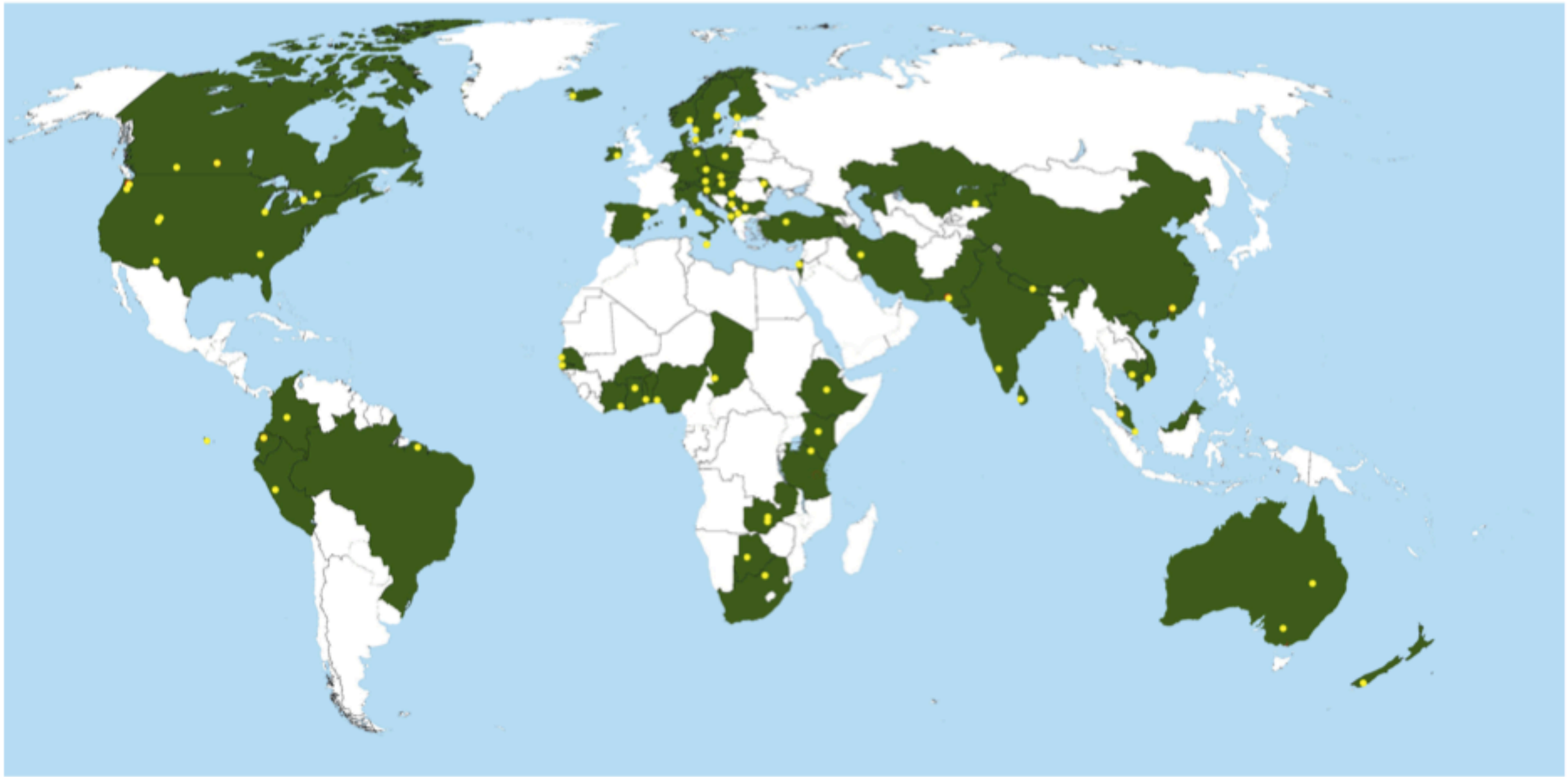
Global surveillance of AMR

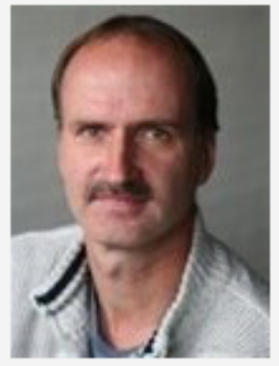


Can human sewage be used to detect and combined with modelling explain global emergence and trends in AMR ?

Sample collection - 2016

79 samples from 60 countries have been collected and analysed





MGmapper: Reference based mapping and taxonomy annotation of metagenomics sequence reads

Thomas Nordahl Petersen[1]*, Oksana Lukjancenko[2], Martin Christen Frølund Thomsen[1], Maria Maddalena Sperotto[1], Ole Lund[1], Frank Møller Aarestrup[2], Thomas Sicheritz-Pontén[1]*

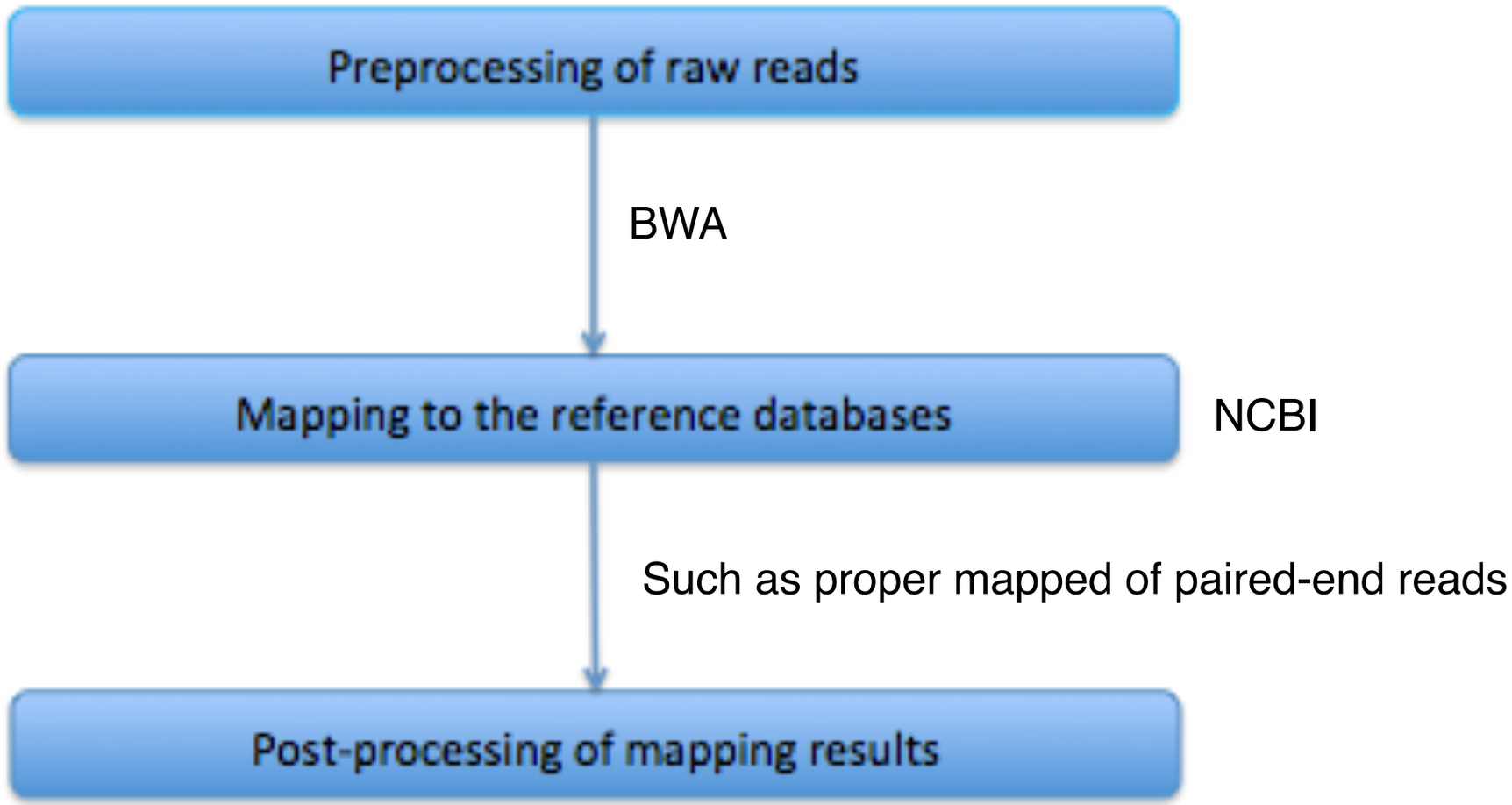
1 Department of Bio and Health Informatics, Technical University of Denmark, Kongens Lyngby, Denmark,
2 National Food Institute, Technical University of Denmark, Kongens Lyngby, Denmark

* tnp@cbs.dtu.dk (TNP); thomas@cbs.dtu.dk (TSP)

Abstract

An increasing amount of species and gene identification studies rely on the use of next generation sequence analysis of either single isolate or metagenomics samples. Several methods are available to perform taxonomic annotations and a previous metagenomics benchmark study has shown that a vast number of false positive species annotations are a problem unless thresholds or post-processing are applied to differentiate between correct and false annotations. MGmapper is a package to process raw next generation sequence data and perform reference based sequence assignment, followed by a post-processing analysis to produce reliable taxonomy annotation at species and strain level resolution. An in-vitro bacterial mock community sample comprised of 8 genuses, 11 species and 12 strains was previously used to benchmark metagenomics classification methods. After applying a post-processing filter, we obtained 100% correct taxonomy assignments at species and genus level. A sensitivity and precision at 75% was obtained for strain level annotations. A comparison between MGmapper and Kraken at species level, shows MGmapper assigns taxonomy at species level using 84.8% of the sequence reads, compared to 70.5% for Kraken and both methods identified all species with no false positives. Extensive read count statistics are provided in plain text and excel sheets for both rejected and accepted taxonomy annotations. The use of custom databases is possible for the command-line version of MGmapper, and the complete pipeline is freely available as a bitbucked package (https://bitbucket.org/genomicepidemiology/mgmapper). A web-version (https://cge.cbs.dtu.dk/services/MGmapper) provides the basic functionality for analysis of small fastq datasets.

MGmapper


Preprocessing of raw reads
BWA
Mapping to the reference databases
NCBI
Such as proper mapped of paired-end reads
Post-processing of mapping results

6

mapping AMR result

Gene	Description	GeneSize	ALB.17	AUS.18	AUS.18a	AUT.70
16S_rRNA_methyltransferase|	gentamicin	732	0	0	0	0
16S_rRNA_methyltransferase|	gentamicin	1263	0	0	0	0
16S_rRNA_methyltransferase|	gentamicin	765	0	0	0	0
16S_rRNA_methyltransferase|	gentamicin	765	0	0	0	0
16S_rRNA_methyltransferase|	gentamicin	804	0	0	0	0
16S_rRNA_methyltransferase|	gentamicin	804	0	0	0	0
16S_rRNA_methyltransferase|	gentamicin	765	0	0	0	0
AAC|GQ343136.1|	gentamicin	867	50	2	2	6
AAC|GQ343186.1|	sisomycin	429	0	0	0	0
AAC|KJ695106.1|	gentamicin	501	0	0	0	0
beta_lactamase|GQ342998.1|	amikacin	891	0	0	0	0
beta_lactamase|GQ342999.1|	amikacin	891	0	0	0	0
beta_lactamase|GQ343000.1|	amikacin	906	0	0	0	0
beta_lactamase|GQ343002.1|	amikacin	873	0	0	0	0
beta_lactamase|GQ343003.1|	amikacin	891	0	0	0	6
beta_lactamase|GQ343008.1|	amikacin	1263	2	0	0	0
beta_lactamase|GQ343010.1|	amikacin	1167	0	0	0	0
beta_lactamase|GQ343015.1|	amikacin	1263	0	0	0	6
beta_lactamase|GQ343019.1|	amikacin	891	0	0	0	0

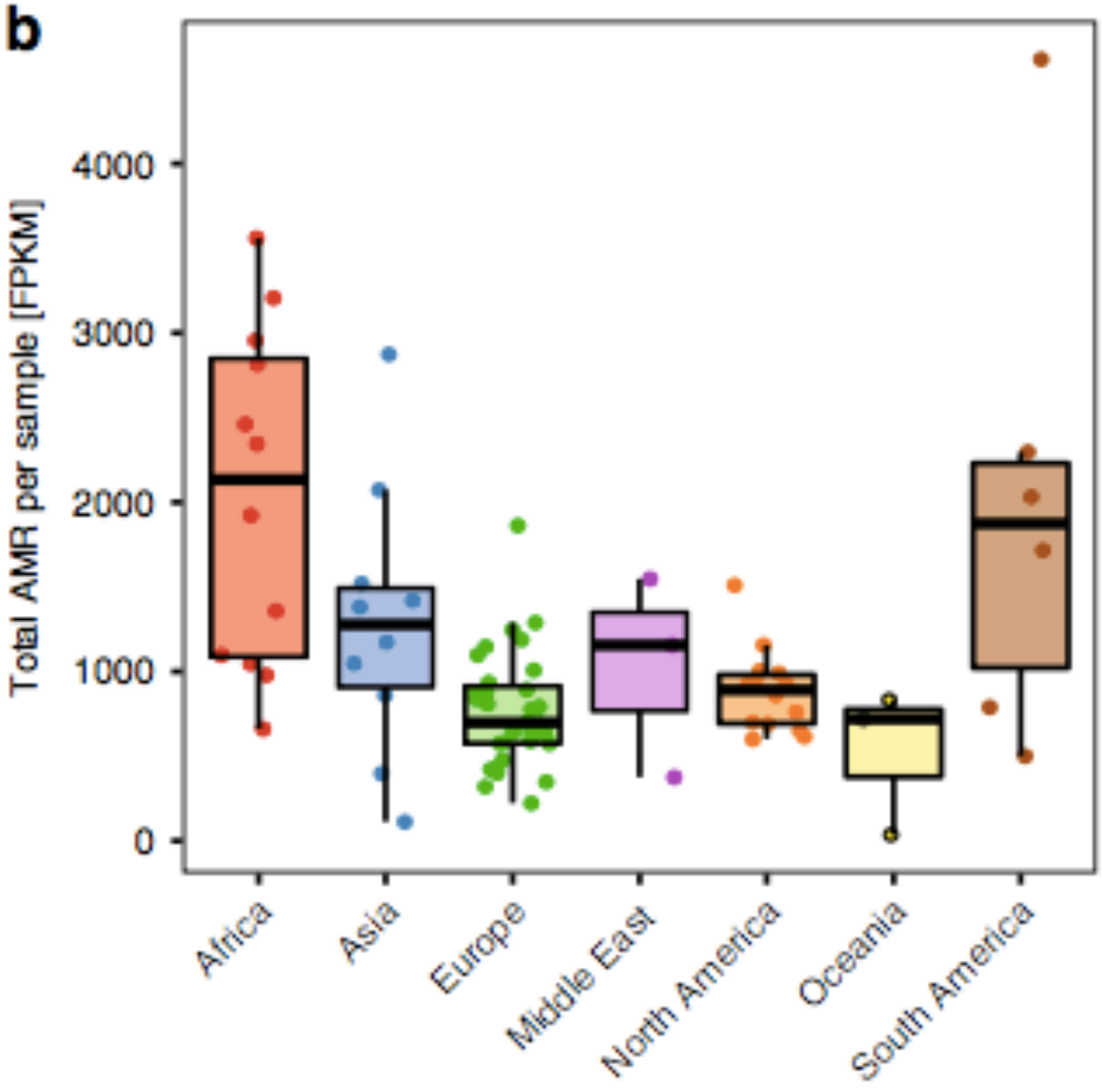
Sample	ReadCount_Bacteria
ALB.17	29715190
AUS.18	11724196
AUS.18a	3778910
AUT.70	15734026
BGR.66	19108694
BRA.53	17108324
BRA.53a	7733328
BWA.19	6599146
CAN.22	1423112
CAN.22a	868336
CAN.22b	11949640
CAN.22c	21282768

FPKM

Fragments Per Kilobase reference per Million bacterial fragments

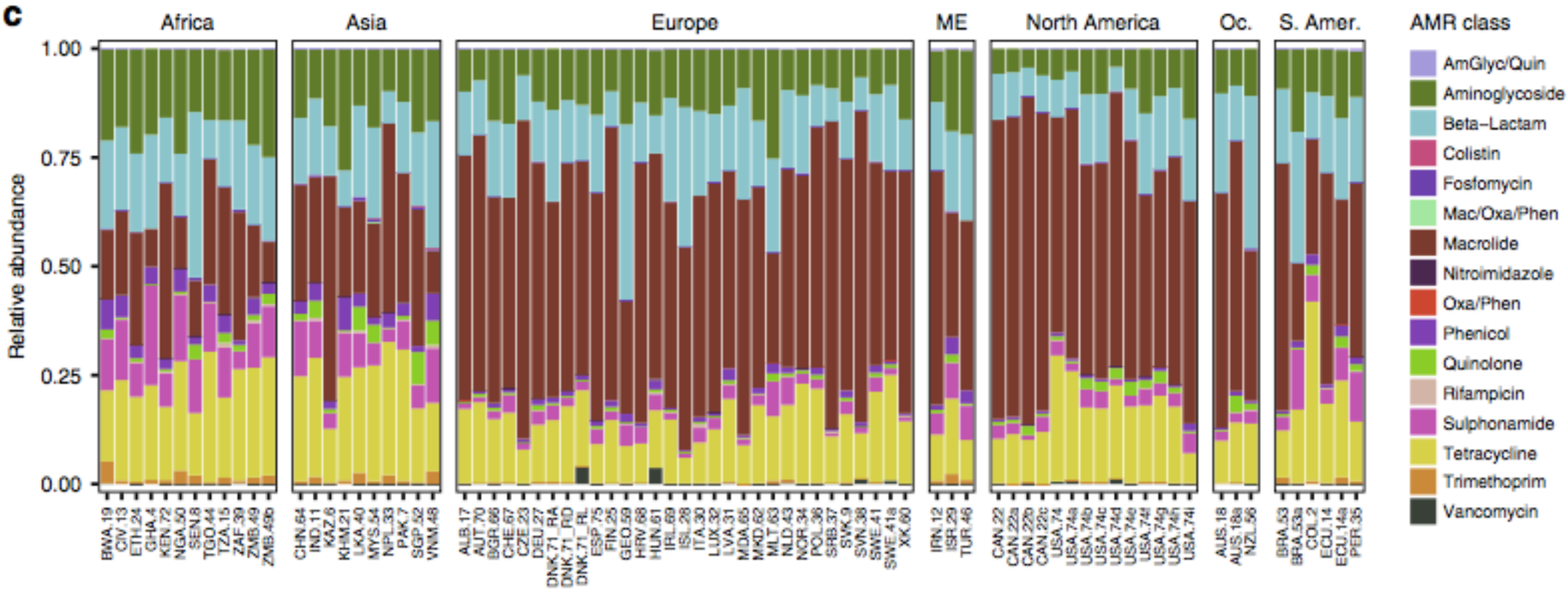
Total FPKM

Fragments Per Kilobase reference per Million bacterial fragments



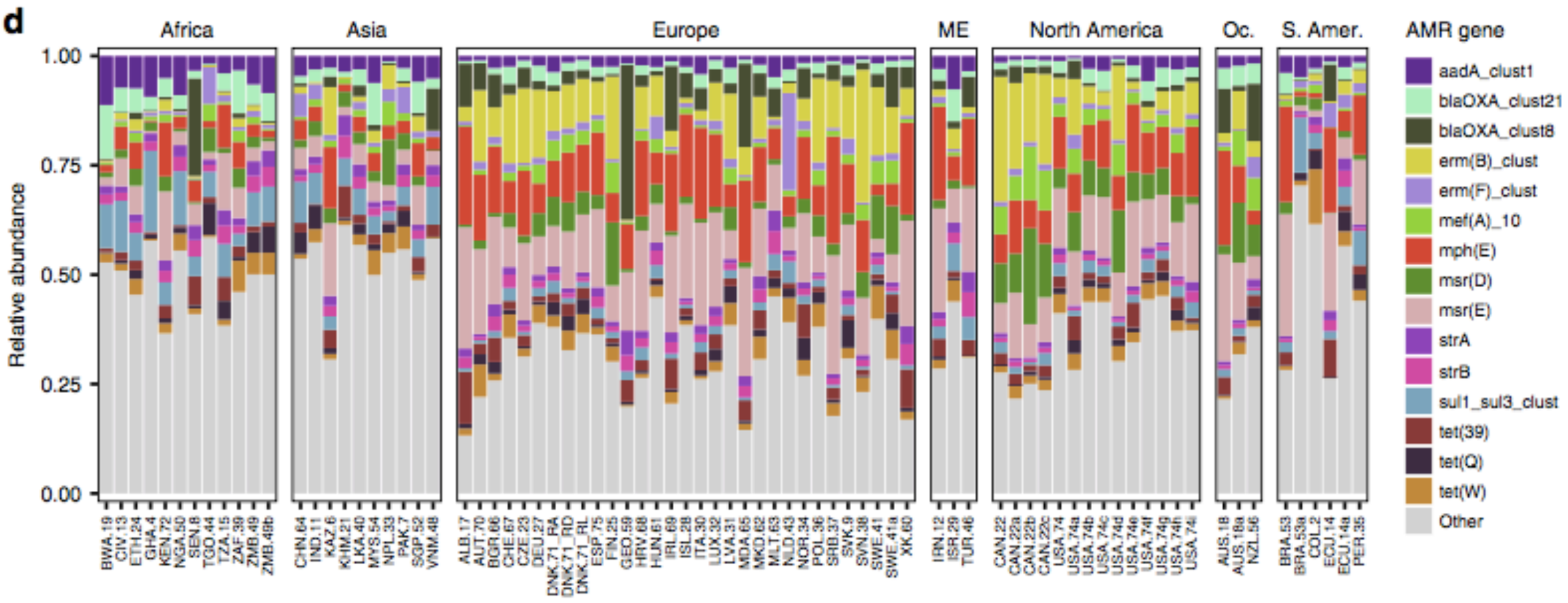
Hendriksen RS, et al. Nature Communications, 2019

AMR classes



Hendriksen RS, et al. Nature Communications, 2019

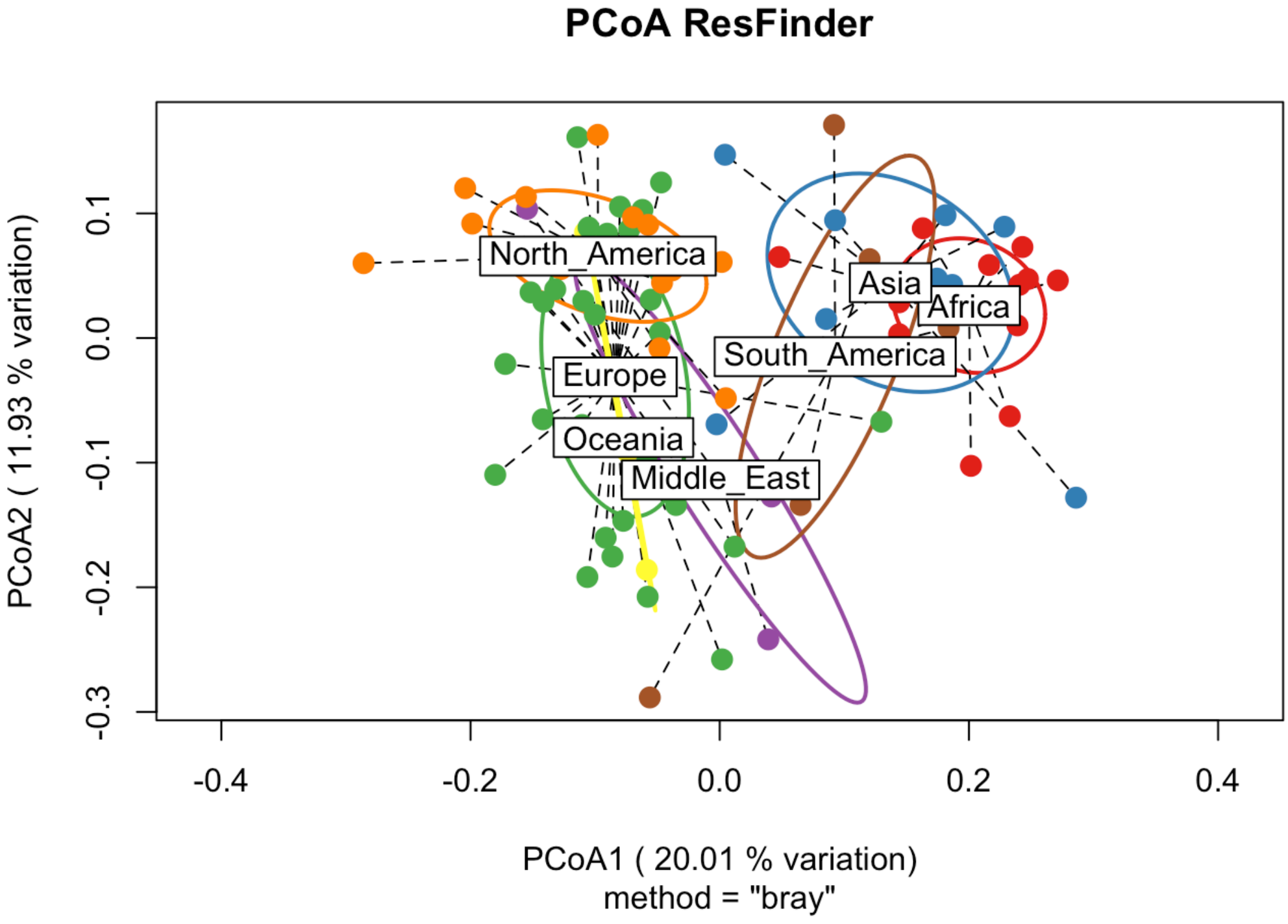
AMR genes



Hendriksen RS, et al. Nature Communications, 2019

Resistome clustering in sewage across regions

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

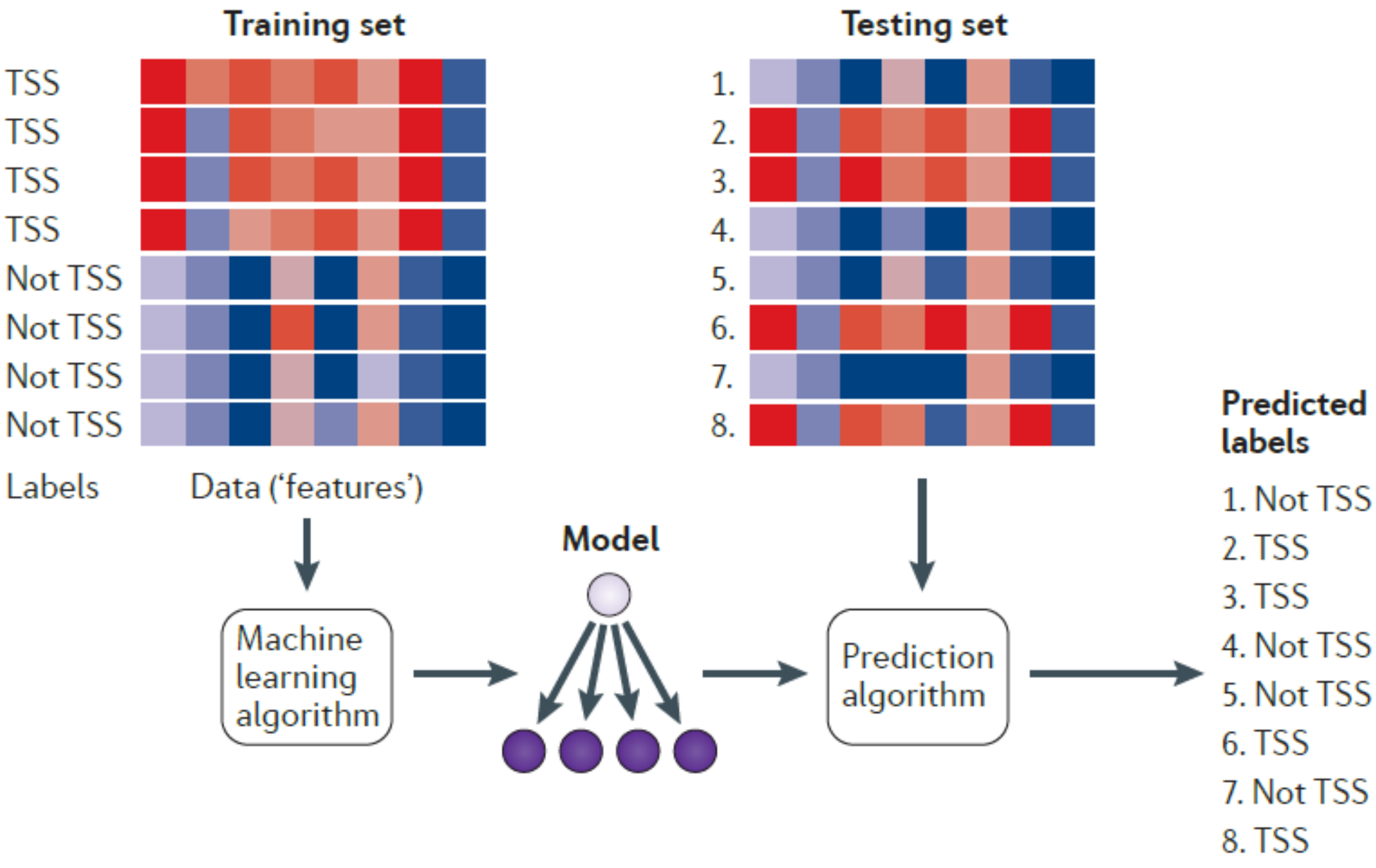


Drivers for AMR

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

Hendriksen RS, et al. Nature Communications, 2019

Predict AMR level using socio-economic data



Deeper look into socio-economic data from World bank

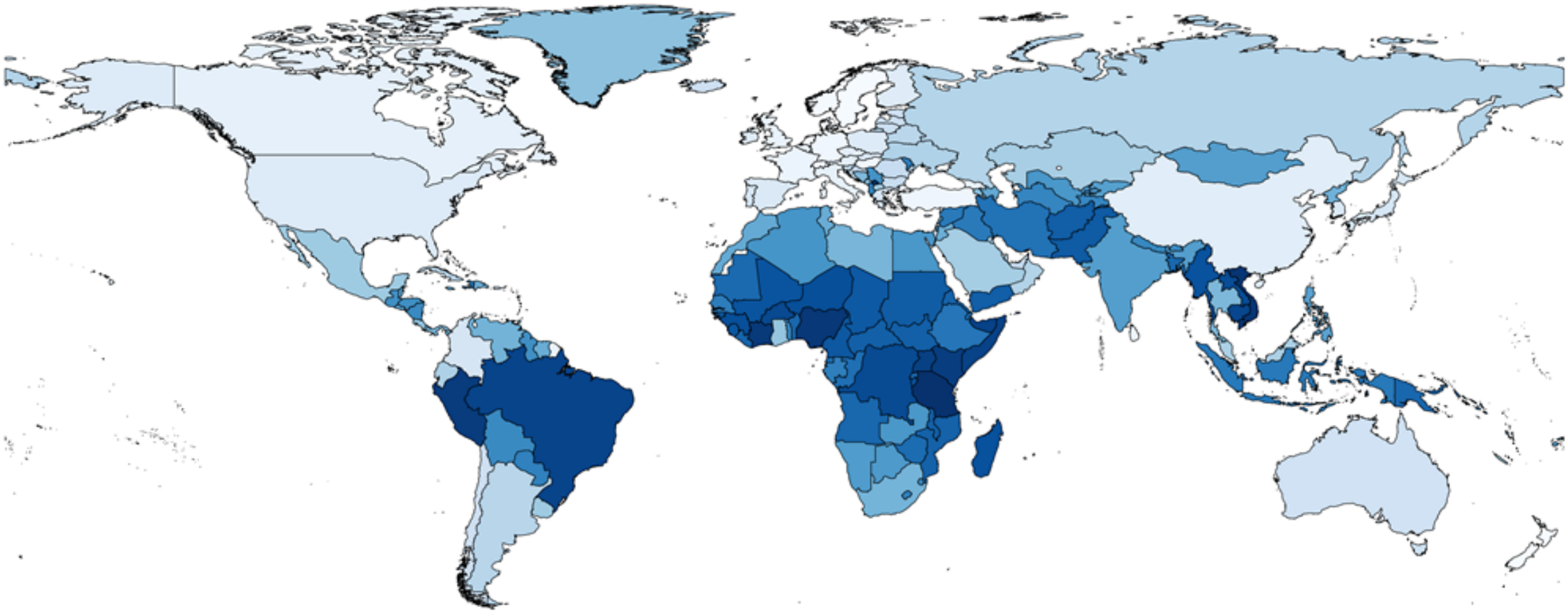
Predictors of higher AMR

- 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



Hendriksen RS, et al. Nature Communications, 2019

Samples

Pilot (2016)

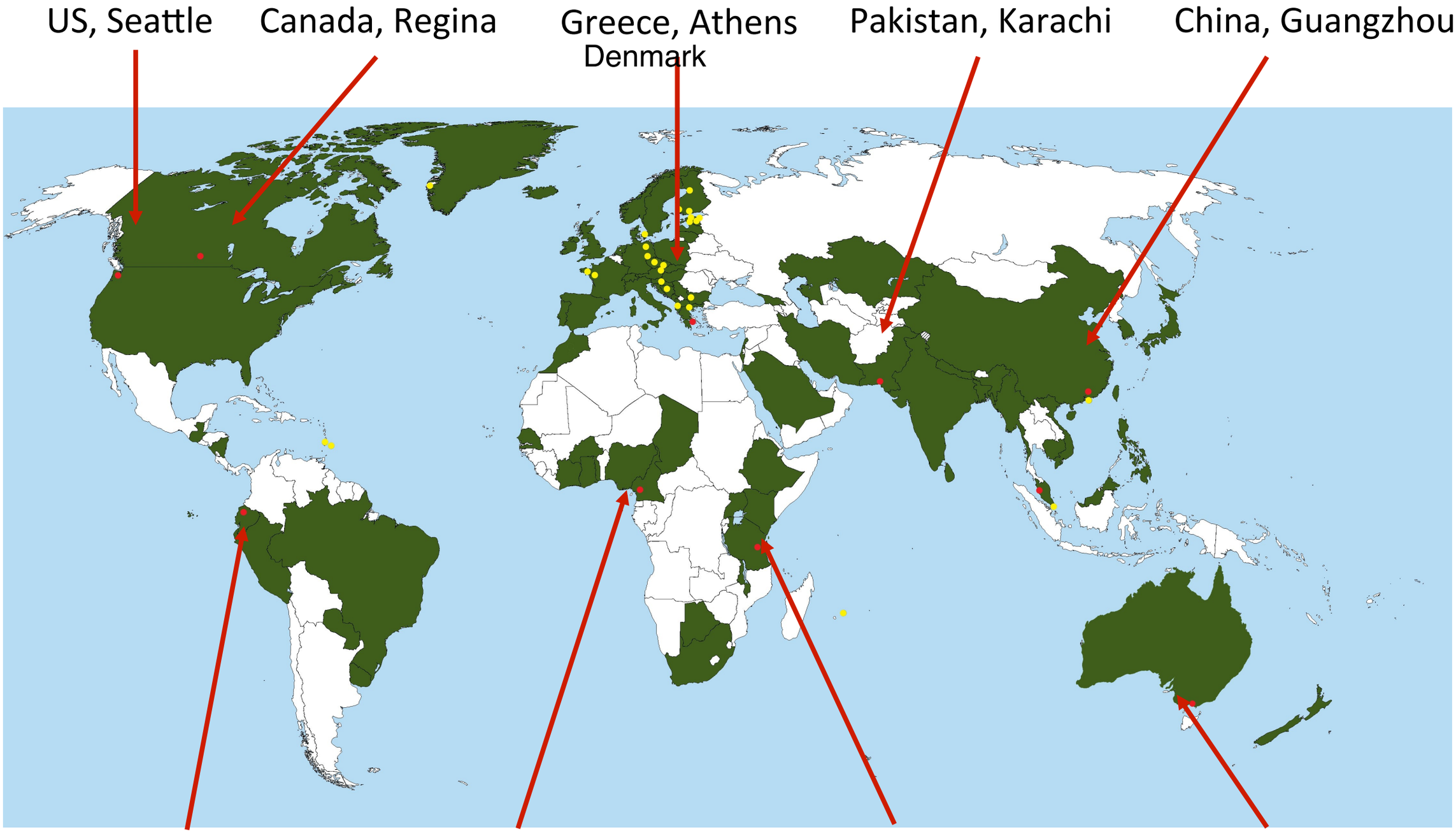
June 2017

November 2017

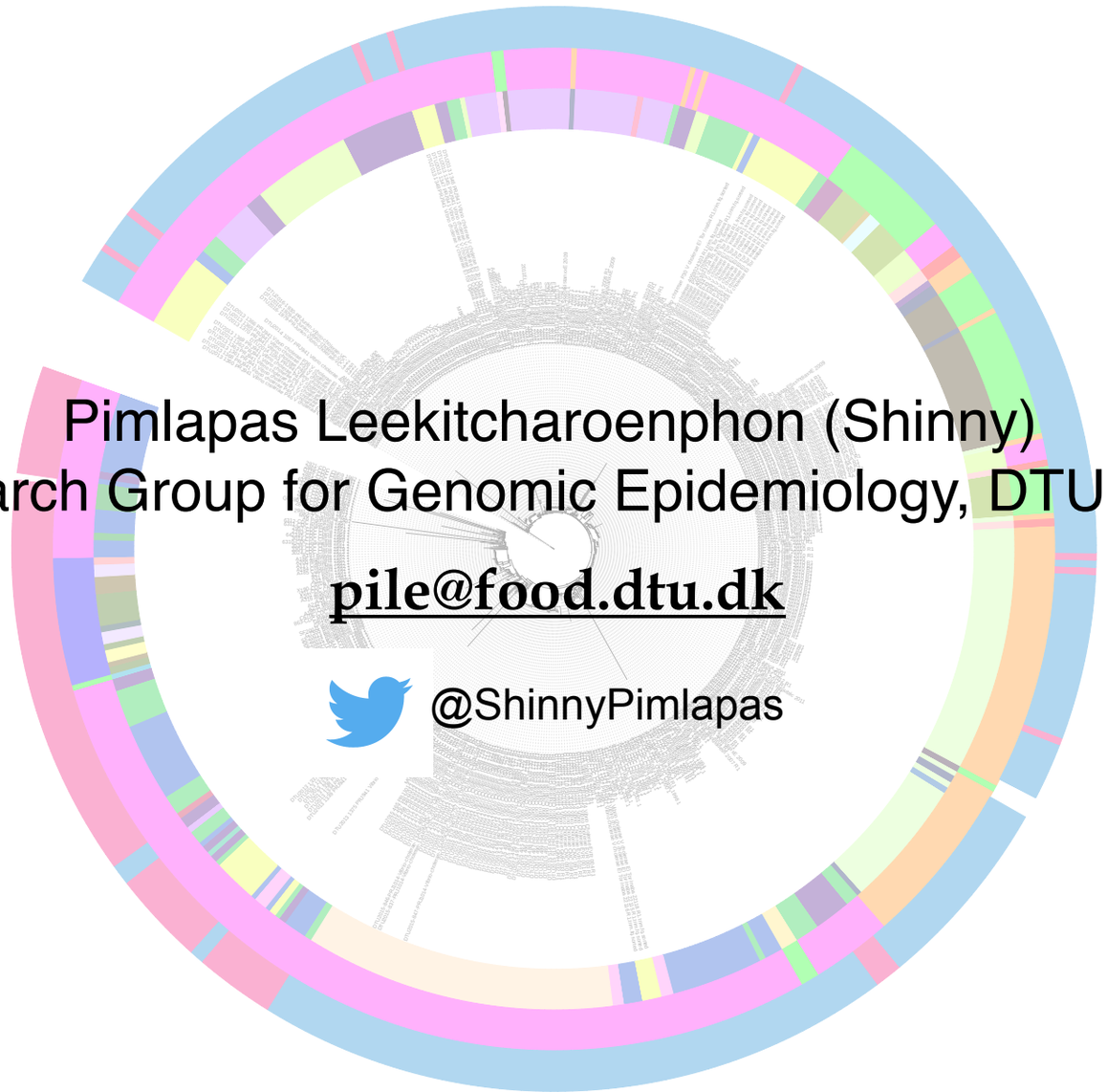
June 2018

November 2018

Sample collection – Longitudinal
Monthly samples in one year



X



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

pile@food.dtu.dk

@ShinnyPimlapas