National Food Institute
Technical University of Denmark
DTU

Opportunities and challenges with metagenomic based waste-water surveillance

(From standards to chaos?)

Frank M. Aarestrup

fmaa@food.dtu.dk, www.genomicepidemiology, www.globalsurveillance.eu

No matter how we look at it

- Surveillance is the basis of everything and what the world needs is:

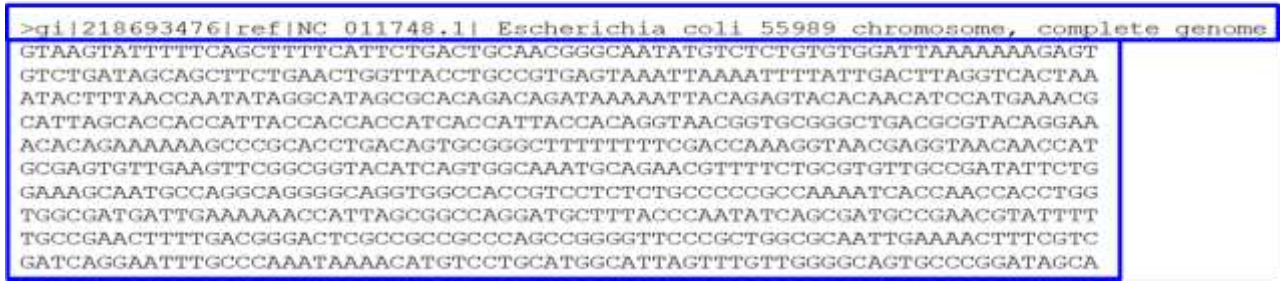
- Real-time data on occurences of all infectious agents and AMR everywhere
  - Geography, reservoir and pathogen independent

  - Observe trends and rapidly compare between data
  - Transfer of information to those who need to:
    - Take public health respose
    - Develop tests and treatments
    - Take clinical decisions

Advantages of Next Generation Sequencing (NGS)

- DNA/RNA are common across pathogens
- NGS provides a universal language
- Raw data are shared allowing for QC and re-analyses
- Less equipped labs may leapfrog

Sequence data:

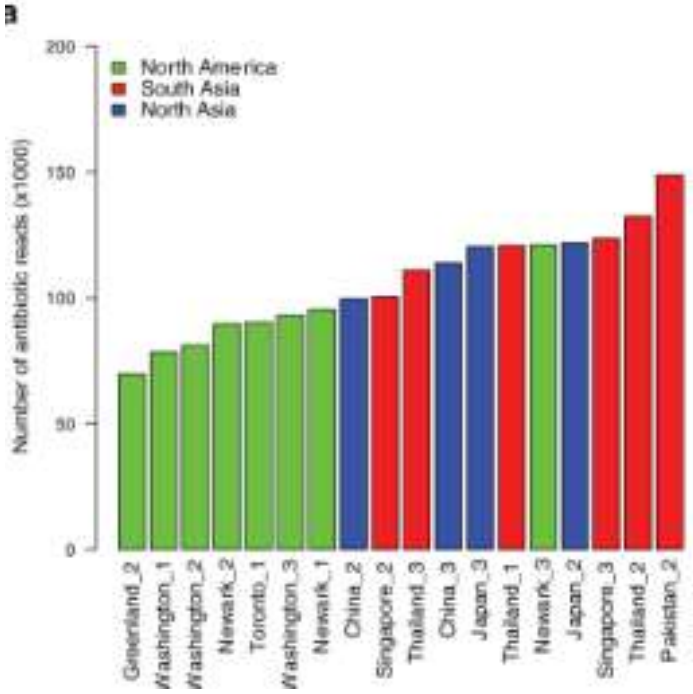


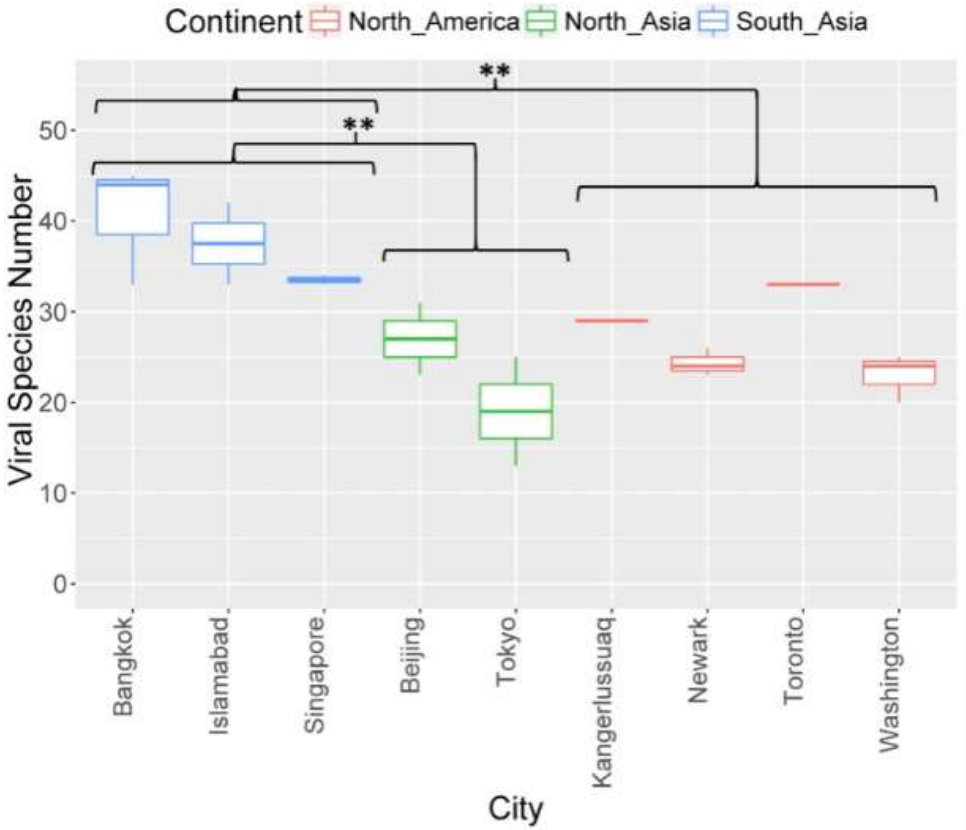
Metagenomics – One technology that takes all

Metagenomics is defined as the **sequencing-based analysis of genomes** contained within an **environmental sample**

Sewage
By-passing the problems of getting clinical isolates

Global hot-spots





Pathogens and AMR

Pedersen et al. 2015, Hjelmsø et al. 2019

Hendriksen et al. Pathogen surveillance in the informal settlement, Kibera, Kenya, using a metagenomics approach. PLoS One. 2019;14:e0222531.

Cluster 9

Population size 1,845

Density (People per 100 m²) 8.4/ m²

Sanitation facility status (1 poor -5 good) 1

Environment type:

- Evidence of fecal matter
- Presence of water puddles/pool
- Closer to the river which drains the eastern side of the study area

Sewage connections and direction

- Drains the eastern part of the study area

Cluster 10

Population size 3,727

Density (People per 100 m²) 8.8/ m²

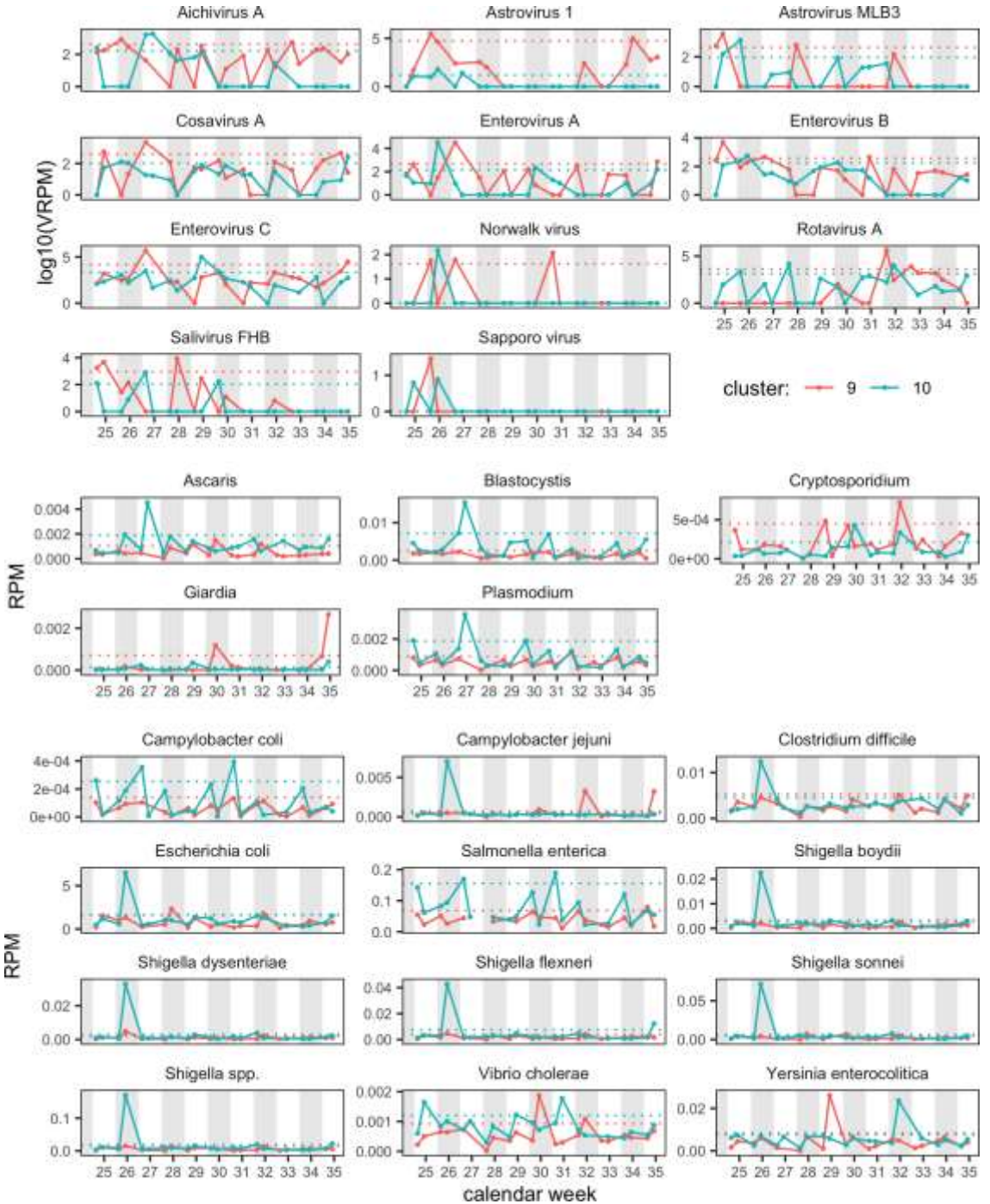
Sanitation facility status (1 poor -5 good) 2

Environment type

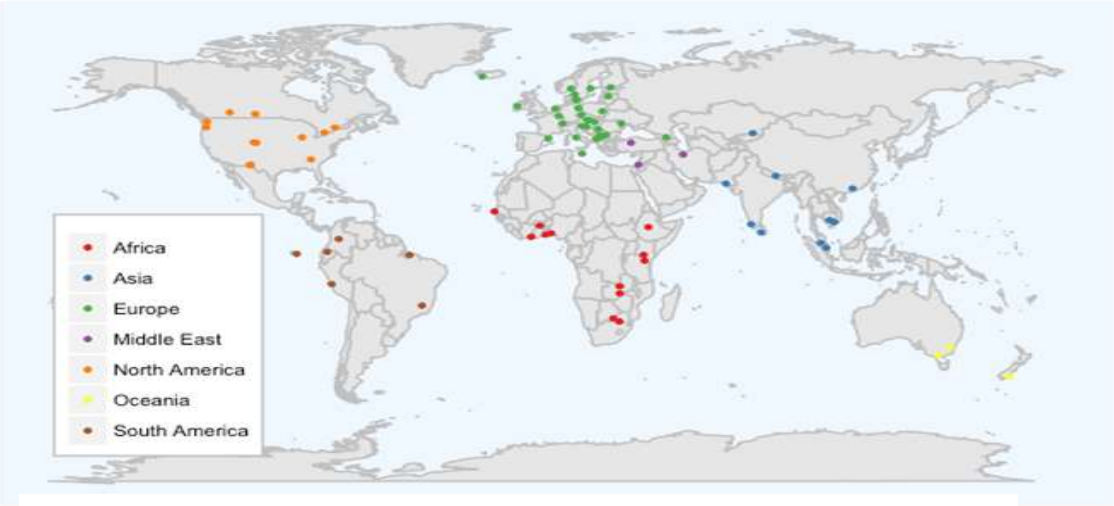
- Evidence of fecal matter
- Presence of water puddles/pool
- Closer to the confluence of rivers draining the eastern and southern parts of the study area

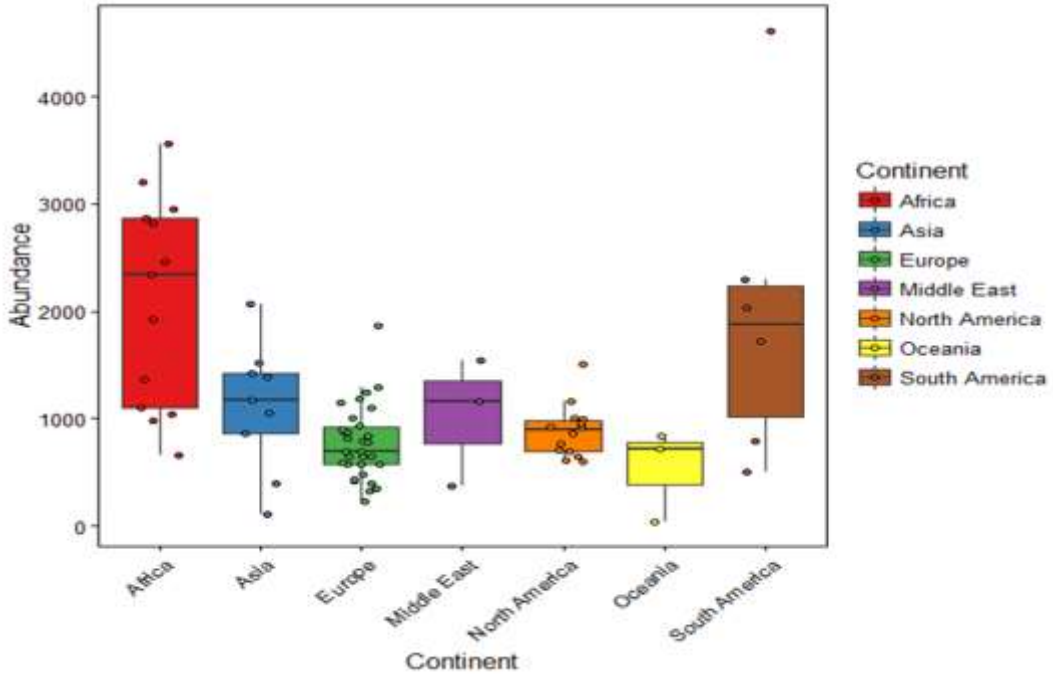
Sewage connections and direction

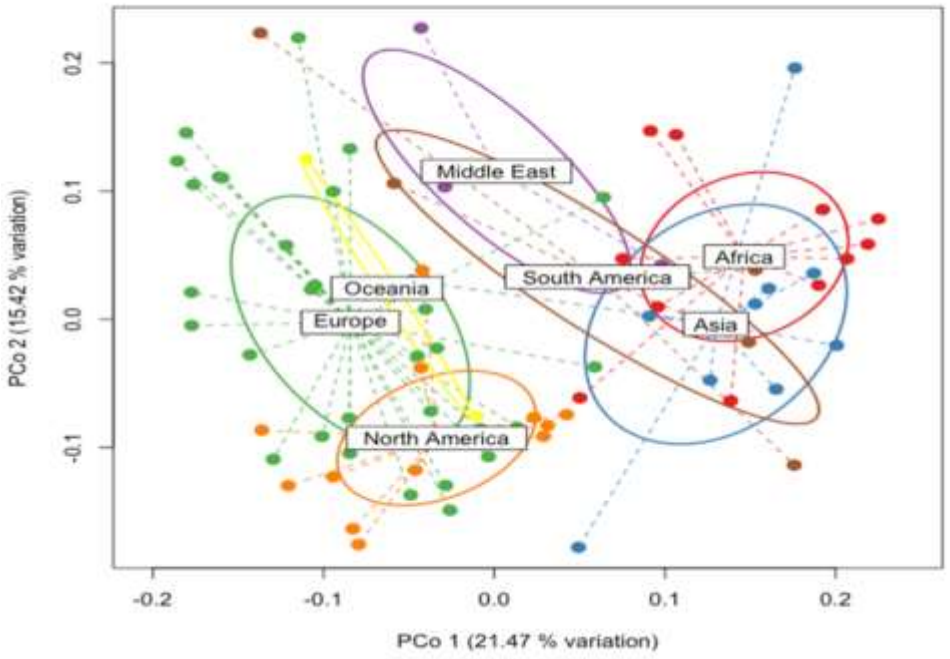
- Drains the south eastern part of the study area
- Accumulates flows from part of cluster 9



Global sewage surveillance - 2016







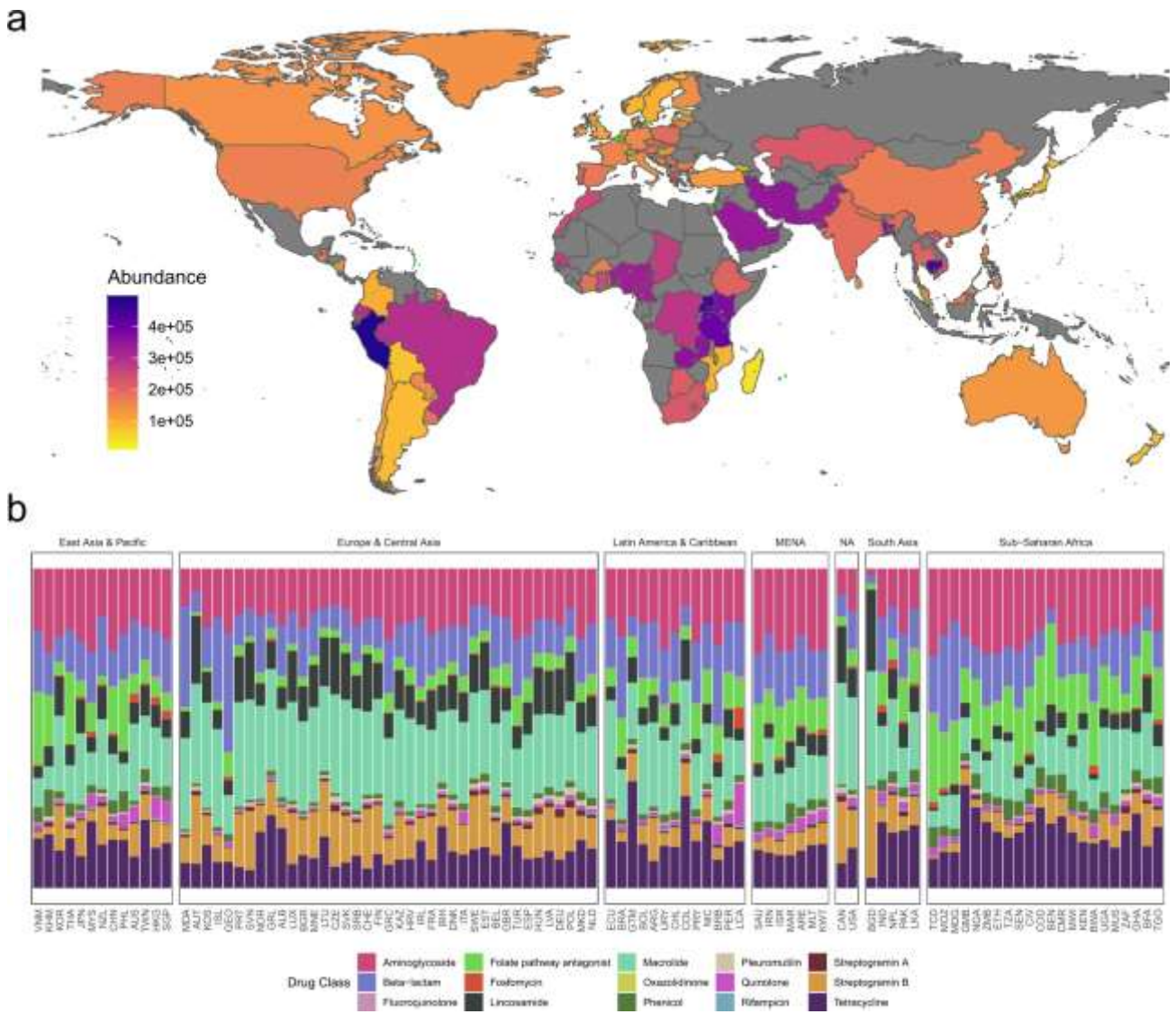
79 cities – 63 countries

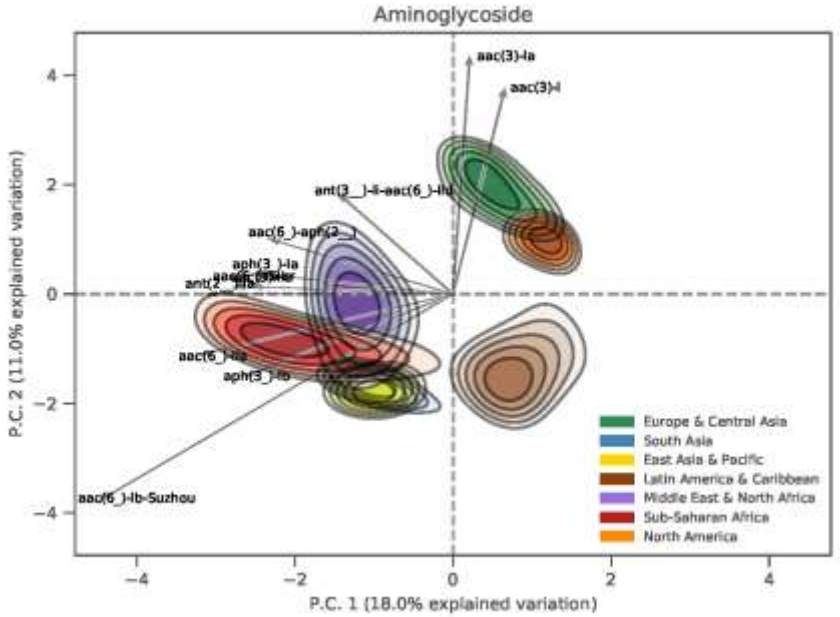
1.500 Gb, large diversity + 30 million genes

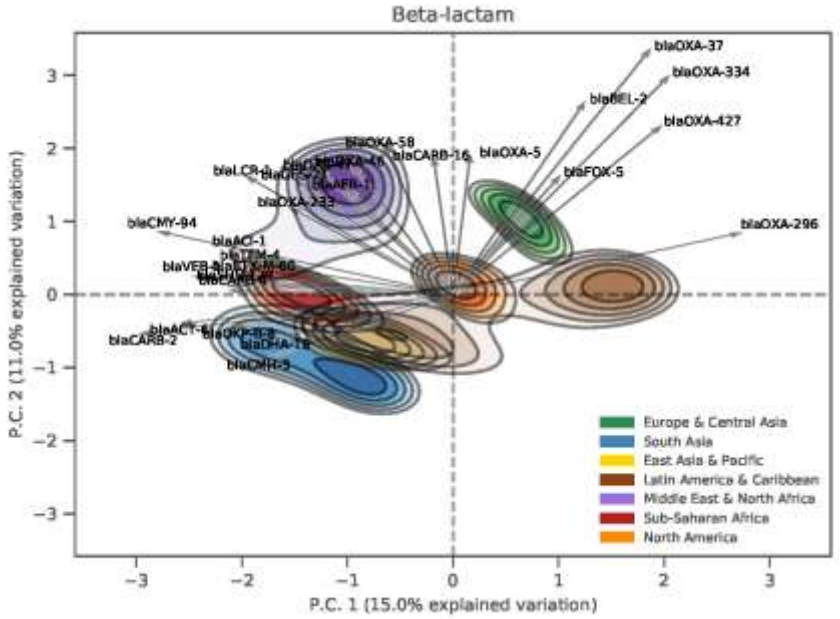
Clear regional separation

Hendriksen et al. Global monitoring of antimicrobial resistance based on metagenomics analyses of urban sewage. Nat Commun. 2019;10:1124.

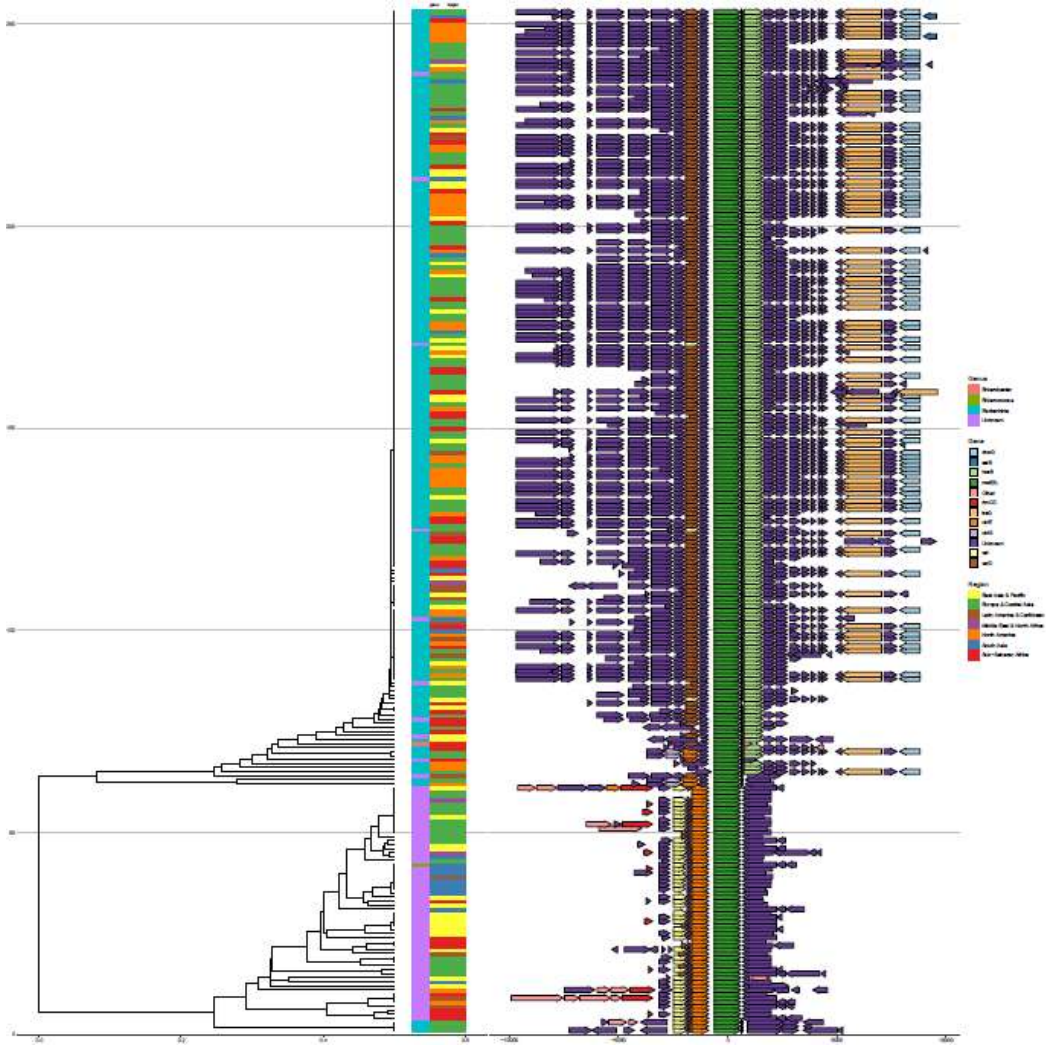
Continued global sewage surveillance



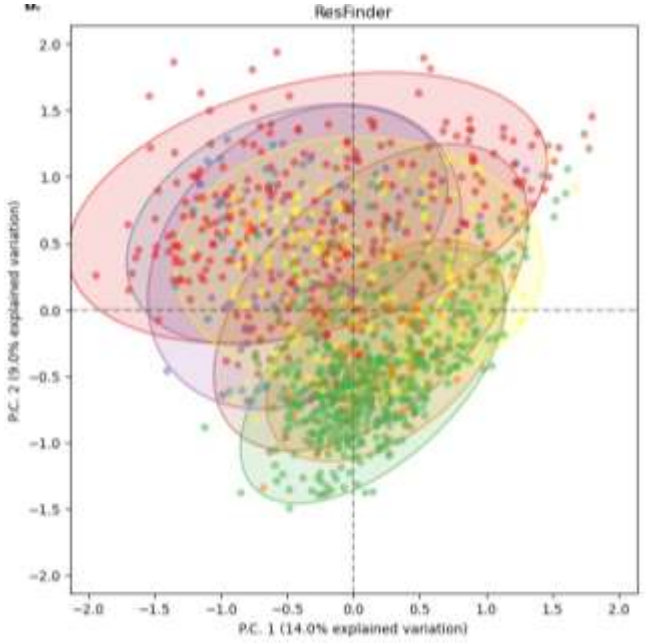




The different classes separate the world in different ways



Munk et al. Genomic analysis of sewage from 101 countries reveals global landscape of antimicrobial resistance. Nat Commun. 2022;13:7251.



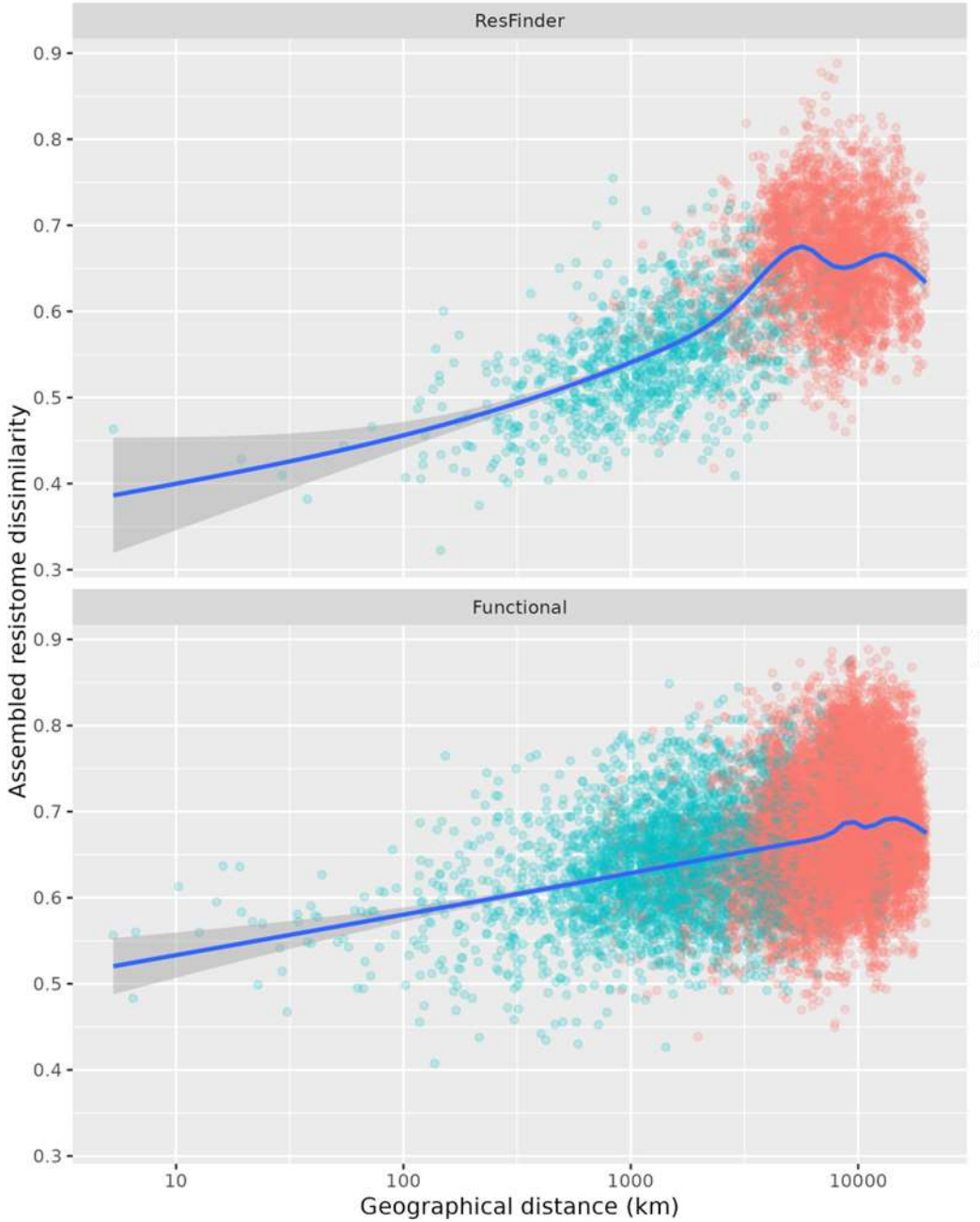
1240 samples

351 cities

111 countries

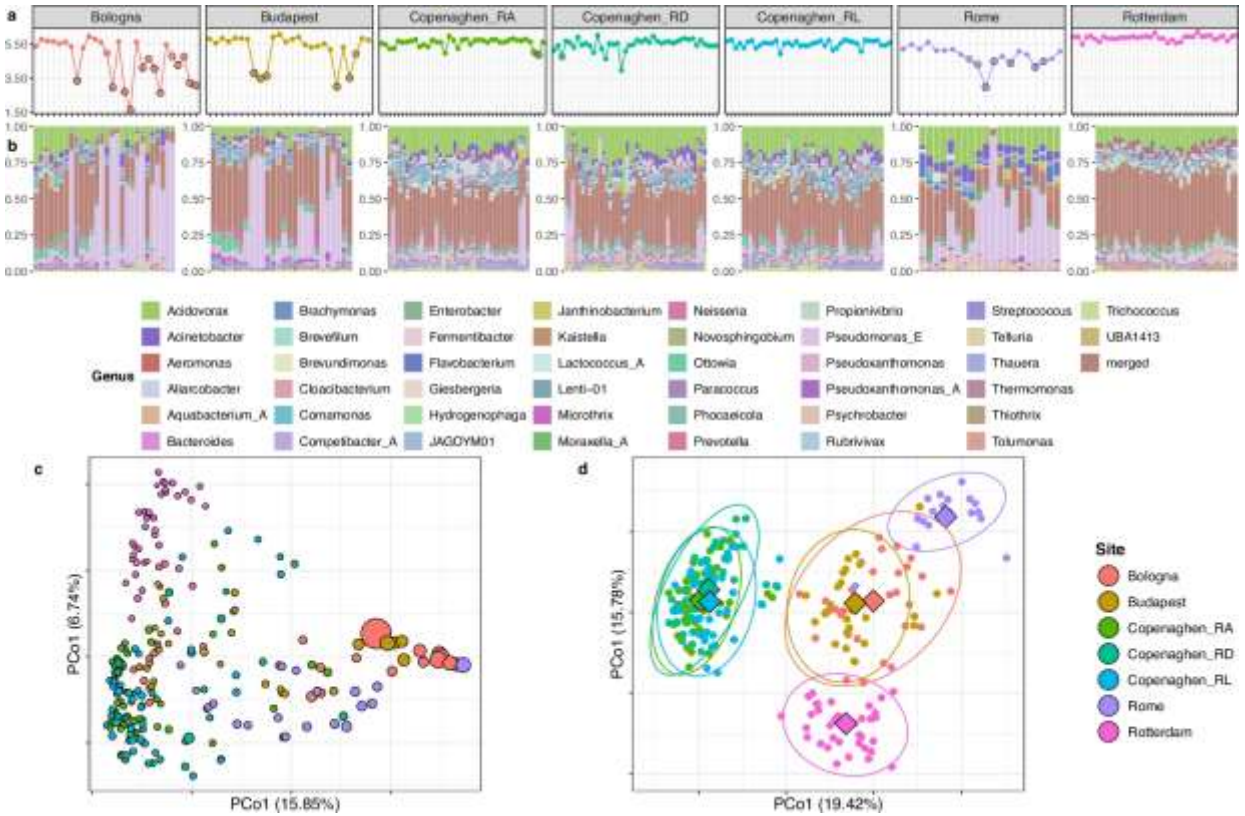
- Still clear separation, but less so for functionally identified resistance genes
- Using assemblies and looking at variants suggest distance-decay

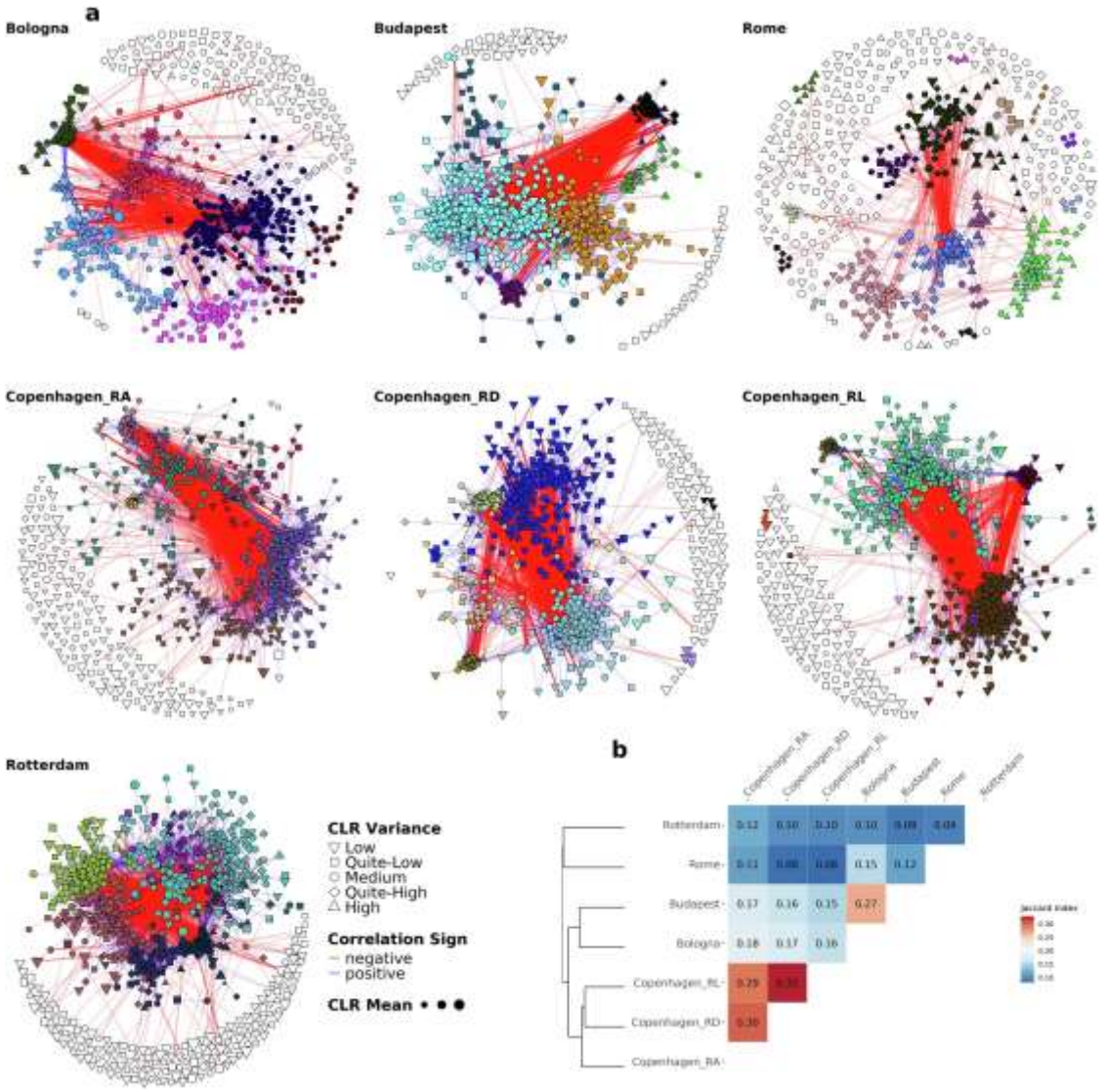
Munk, Martiny et al. in prep



Five european cities

- Very clear site-specific signatures
- Temporal changes
- Network analyses potentially all9ow identification of the fecal fraction





Becsei et al. Time-series sewage metagenomics distinguishes seasonal, human-derived and environmental microbial communities potentially allowing source-attributed surveillance. Nat Commun. 2024;15:7551.

Standardizing wet-lab methods

Sample storage. Poulsen et al. (2022) Microb. Spec.

DNA extraction from sample
BE Knudsen et al. (2016) mSystems

DNA sheared to 3-400 bp

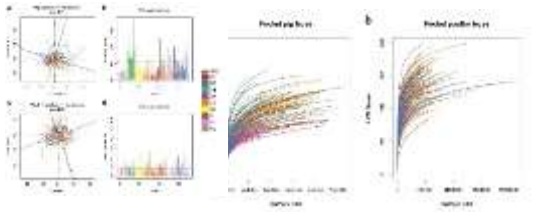
DNA sequencing. Poulsen et al. (2021) Microb. Spec.

KAPA Hyper PCR-free library preparation

HiSeq4000 (Pilot). Then NovaSeq6000

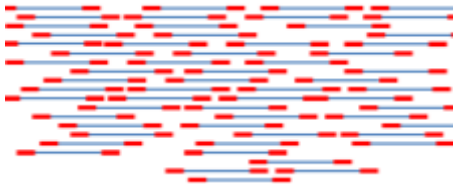
2 x 35M 150bp fragments = 10 Gbp

Dry-lab methods

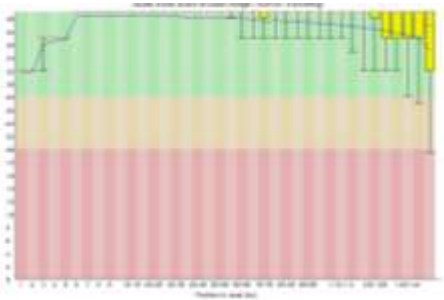


Quantitative analyses

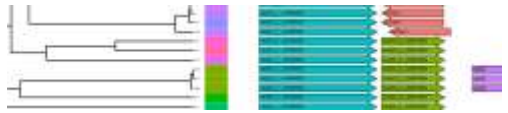
Alignment with **KMA** against **ResFinder** and **Silva**



PE DNA reads >35M/ sample

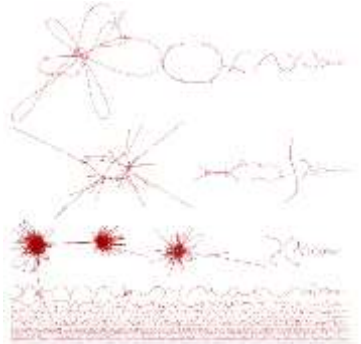


QC, quality- and adapter-trimming **bbduk2**



Flank-based analyses

Identifying ARGs with **ResFinder**



De novo metagenomic assembly using **metaSpades**

Genome binning with **MetaBat2**

Taxonomic assignment with **Kraken2**

Plasmid identification with **Kraken2, BLAST** and **PPR-Meta**

Finding and annotating genes with **Prokka**

Munk et al. Nat. Comm. 2022
Aytan-Aktug et al. mSystems. 2022
Bortolaia et al. JAC. 2020
Szarvas et al. Commun Biol. 2020
Hendriksen et al. Nat. Comm. 2019
Clausen et al. BMC Bioinformatics. 2018
Munk et al. Nat. Micro. 2018
Zankari et al. JAC. 2017
Kaas et al. PlosOne. 2014
Zankari et al. JAC. 2012
Larsen et al. J Clin Microbiol. 2012

Importance of standardisation

- Standardisation is important especially for determination of abundances
- However, un-standardised data might still be useful for network analyses and work with assemblies
- Sharing raw data will allow for multiple analyses

Concluding remarks

- Surveillance is the basis of everything, and without it we are flying blind
  - Easy to say – difficult to conduct
- Metagenomics and sewage offer several novel options for global surveillance
- Standardisation is important, but data can still be used
- (raw) Data-sharing is essential and should be a mandatory part of any funding

European Commission

novo nordisk fonden