Global genomic epidemiology

Frank M. Aarestrup
www.compare-europe.eu
www.genomicepidemiology.org
Identification of Russian commanders is limited to those who played a major role in the fighting. The entire Russian command system is not depicted, nor was it entirely clear exactly which units were subordinated to which major force commanders.
Strategy to win a war
Putting up the right defense

Knowing:
- Where
- What
- When
- How

Communicate & Prioritise
Infectious disease situation 2015

• 22% of all global deaths are directly caused by infections

• Dynamics of common infectious diseases are changing
  – Demographic change, population density, anti vaccine, AMR, etc.

• New diseases emerge frequently
  – Deforestation, population growth, health system inequalities, travel, trade, climate change

• Effects are difficult to predict due to complexity of problems
  – Rapid flexible response

• Public health and clinical response depend on global capacity for disease surveillance
  – Rapid sharing, comparison and analysis of data from multiple sources and using multiple methodologies
Response to ID outbreaks usually fragmented and too late
Response to ID outbreaks with improved detection and sharing of data

- Public Health response
- Preclinical research response
- Clinical research response
Global influenza surveillance network

The WHO Global Influenza Surveillance Network (GISN), July 2008
Cookie dough E. coli outbreak grows, CDC says

FDA: Salmonella outbreak linked to papayas from Mexico

Peanut butter distributor recall salmonella scare

Salmonella found on pepper; FDA strengthens warning

Tomatoes Pulled Off Shelves, Menus Amid Salmonella Scare

food recall: 36 million pounds ground turkey – Salmonella Heidelberg strain

Half a billion eggs have been recalled
What the world needs

• Real-time sharing of data on occurrences of all infectious agents
  – Specifically for AMR also selection in the microbiome

• Detection of related clusters in time and space

• Possibility to observe trends in clones and species as well as virulence and resistance

• Ability to rapidly compare between all types of data

There can be no real-time surveillance without real-time data sharing
Old School –
Current conventional bacterial diagnostics
New School –
Culture based diagnostics using NGS

Pipelines for multiple species
Workflow with WGS at the clinical laboratory

Modified from Didelot et al., 2012.
NGS advantages

• Laboratory diagnostics increasingly rely on (pathogen) genomic information
• RNA / DNA are common across pathogens, therefore, methods to analyse pathogen genomes are potentially universal
• Next generation sequencing capacity is developing fast, and costs are becoming competitive
• Data are easy to share electronically and are in a standardized format

➤ Capturing NGS developments may provide a universal language that can be harnessed for early detection of outbreaks across disciplines and domains
➤ If the technology keeps developing, less equipped labs may leapfrog
Our very simple strategy
Online bioinformatic tool

Resfinder

Resistance gene profile

Assembly pipeline

Center for Genomic Epidemiology

ResFinder usage, August 2012 - December 2015
Total use: 74,600 runs

Distribution among countries 2015
69 different countries in total

United States, 18%

Netherlands, 13%

United Kingdom, 10%

Australia, 7%

Germany, 6%

Denmark, 5%

France, 5%

Japan, 4%

Italy, 3%

Canada, 2%

Taiwan, 2%

Russia, 2%

China, 2%

Norway, 2%

Sweden, 1%

Spain, 1%

Malaysia, 1%

Singapore, 1%

Belgium, 1%

Portugal, 1%

Finland, 1%

Other, 9%

ResFinder-1.3 Server - Results

Center for Genomic Epidemiology

Home | Services | Institutions

ResFinder usage, August 2012 - December 2015
Total use: 74,600 runs

Distribution among countries 2015
69 different countries in total

United States, 18%

Netherlands, 13%

United Kingdom, 10%

Australia, 7%

Germany, 6%

Denmark, 5%

France, 5%

Japan, 4%

Italy, 3%

Canada, 2%

Taiwan, 2%

Russia, 2%

China, 2%

Norway, 2%

Sweden, 1%

Spain, 1%

Malaysia, 1%

Singapore, 1%

Belgium, 1%

Portugal, 1%

Finland, 1%

Other, 9%
Identification of acquired antimicrobial resistance genes

Ea Zankari¹,²*, Henrik Hasman¹, Salvatore Cosentino², Martin Vestergaard¹, Simon Rasmussen², Ole Lund², Frank M. Aarestrup¹ and Mette Voldby Larsen²

¹National Food Institute, Technical University of Denmark, 2800 Kgs. Lyngby, Denmark; ²Center for Biological Sequence Analysis, Department of Systems Biology, Technical University of Denmark, 2800 Kgs. Lyngby, Denmark

*Corresponding author. Tel: +45-35887183; E-mail: east@food.dtu.dk


Genotyping using whole-genome sequencing is a realistic alternative to surveillance based on phenotypic antimicrobial susceptibility testing

Ea Zankari¹,²*, Henrik Hasman¹, Rolf Sommer Kaas¹,², Anne Mette Seyfarth¹, Yvonne Agersø¹, Ole Lund², Mette Voldby Larsen² and Frank M. Aarestrup¹

¹National Food Institute, Technical University of Denmark, 2800 Kgs. Lyngby, Denmark; ²Center for Biological Sequence Analysis, Department of Systems Biology, Technical University of Denmark, 2800 Kgs. Lyngby, Denmark

Rapid Whole-Genome Sequencing for Detection and Characterization of Microorganisms Directly from Clinical Samples

Henrik Hasman, Dhany Saputra, Thomas Sichertz-Ponten, Ole Lund, Christina Aaby Svendsen, Niels Frimodt-Møller, Frank M. Aarestrup

National Food Institute, Technical University of Denmark, Lyngby, Denmark; Systems Biology, Technical University of Denmark, Lyngby, Denmark; Hvidovre Hospital, Hvidovre, Denmark
More than 450,000 jobs executed
Simple upload

Bacterial Analysis Pipeline - Batch Upload

The CGE Bacterial Analysis Pipeline executes a workflow of services with predefined parameters and stores the submitted data and result in the database at the user's disposal.

View the version history of this server.

STEP 1: Download Metadata Template
STEP 2: Fill out template
STEP 3: Upload Metadata File
STEP 4: Select Files
STEP 5: Submit

Progress Overview

<table>
<thead>
<tr>
<th>Name</th>
<th>Size</th>
<th>Progress</th>
<th>Status</th>
</tr>
</thead>
</table>

Remove all

Supported browsers

<table>
<thead>
<tr>
<th>Browser</th>
<th>Compatibility</th>
</tr>
</thead>
<tbody>
<tr>
<td>Linux</td>
<td>✔️</td>
</tr>
<tr>
<td>macOS</td>
<td>✔️</td>
</tr>
<tr>
<td>Windows</td>
<td>✔️</td>
</tr>
<tr>
<td>Edge</td>
<td>✔️</td>
</tr>
<tr>
<td>Firefox</td>
<td>✔️</td>
</tr>
<tr>
<td>Chrome</td>
<td>✔️</td>
</tr>
</tbody>
</table>

Support

Scientific problems

Technical problems

Copyright DTU 2011 / All rights reserved
Center for Genomic Epidemiology, DTU, Kemitorvet, Building 304, 2800 Kgs. Lyngby, Denmark
Funded by The Danish Council for Strategic Research
Last modified August 25, 2016 13:32:18 GMT
<table>
<thead>
<tr>
<th>Field Name</th>
<th>Mandatory</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>sample_name</td>
<td>No</td>
<td>The name of the isolate the user uses to identify the sample</td>
</tr>
<tr>
<td>group_name</td>
<td>No</td>
<td>The name of the group. Adding a group name will incorporate the sample in the analysis pipeline execution of that group</td>
</tr>
<tr>
<td>file_names</td>
<td>Yes</td>
<td>Name of all files associated to this sample. Multiple filenames should be separated by a space</td>
</tr>
<tr>
<td>sequencing_platform</td>
<td>Yes</td>
<td>Choose between: LS454, Illumina, Ion Torrent, ABI SOLiD and unknown</td>
</tr>
<tr>
<td>sequencing_type</td>
<td>Yes</td>
<td>Choose between: single, paired, mate-paired, and unknown</td>
</tr>
<tr>
<td>pre_assembled</td>
<td>Yes</td>
<td>Has the uploaded sample data been assembled? yes / no</td>
</tr>
<tr>
<td>sample_type</td>
<td>Yes</td>
<td>Has the uploaded sample data been assembled? yes / no</td>
</tr>
<tr>
<td>organism</td>
<td>Yes</td>
<td>Write 'unknown' if the organism name is unknown</td>
</tr>
<tr>
<td>strain</td>
<td>No</td>
<td>Strain type ID</td>
</tr>
<tr>
<td>subtype</td>
<td>No</td>
<td>Subspecific genetic lineage, i.e. MLST, serovar and biotype</td>
</tr>
<tr>
<td>country</td>
<td>Yes</td>
<td></td>
</tr>
<tr>
<td>region</td>
<td>No</td>
<td></td>
</tr>
<tr>
<td>city</td>
<td>(No)</td>
<td></td>
</tr>
<tr>
<td>zip_code</td>
<td>(No)</td>
<td>Please provide as much information as possible. Low resolution locations reduce the usability. The minimal recommended option is to provide either city, zip code or longitude and latitude coordinates.</td>
</tr>
<tr>
<td>latitude</td>
<td>(No)</td>
<td></td>
</tr>
<tr>
<td>longitude</td>
<td>(No)</td>
<td></td>
</tr>
<tr>
<td>location_note</td>
<td>No</td>
<td>The host from which the sample/organism has been taken. This should be a proper scientific name or one of the aliases found in the front of the document.</td>
</tr>
<tr>
<td>isolation_source</td>
<td>Yes</td>
<td>Is the organism deceased? yes / no / unknown</td>
</tr>
<tr>
<td>source_note</td>
<td>No</td>
<td>Additional relevant details about the isolation source i.e. blood, laboratory experiment or urine</td>
</tr>
<tr>
<td>pathogenic</td>
<td>Yes</td>
<td>Is the organism pathogenic? yes / no / unknown</td>
</tr>
<tr>
<td>pathogenicity_note</td>
<td>No</td>
<td>Additional relevant details about the organism's pathogenicity</td>
</tr>
<tr>
<td>collection_date</td>
<td>Yes</td>
<td>The date of the sample collection. Use one of the following format: YYYY-MM-DD or YYYY-MM-YY or YYYY</td>
</tr>
<tr>
<td>collected_by</td>
<td>No</td>
<td>Name of the institute or person who took the sample</td>
</tr>
<tr>
<td>usage_restrictions</td>
<td>Yes</td>
<td>Choose either 'private' or 'public'. Note that private data will be deleted after some time to free disk space on the server.</td>
</tr>
<tr>
<td>release_date</td>
<td>Not Yet Used</td>
<td>Write the date from when the data and results should be public available. Format: YYYY-MM-DD</td>
</tr>
<tr>
<td>email_address</td>
<td>Not Yet Used</td>
<td>Email address of the uploader</td>
</tr>
<tr>
<td>notes</td>
<td>No</td>
<td>Any additional information can be added</td>
</tr>
</tbody>
</table>

Limited mandatory epidemiological information
Uploaded data and analysis easy to search and download
A global platform for the sequence-based rapid identification of pathogens
- Start date December 1st, 2014

Coordinated by
Frank M. Aarestrup (Technical University of Denmark)
Marion Koopmans (Erasmus Medical Center, the Netherlands)
Stakeholder Consultations

Cost Effectiveness studies

Dissemination and Training

Supporting activities

Studies into Barriers to open data sharing

Dissemination and Training

Cost Effectiveness studies
Why a central public repository?

Besides the language and altruistic issues

• The data comparison problem
• Allowing easy transfer between levels of access including public
• Allow access to bioinformatics for the frontline
• Allowing for constantly improving the analytic pipelines
Data comparison problem

Global repositories
> 1-1000 Tb data

Client
~1-100 Gb data

Internet
~1Gb/hour

Bring the tools to the data
mcr-1

- Mcr-1 gene added to ResFinder database on Nov. 24
- Nov. 25-26: initiation of screening with high-throughput methods

**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 Hasman¹, AM Hammerum¹, F Hansen¹, RS Hendriksen², B Olesen³, Y Agersø², E Zankari², P Leekitcharoenphon², M Stegger¹, RS Kaas², LM Cavaco³, DS Hansen³, FM Aarestrup², RL Skov¹

1. 1. Department of Microbiology and Infection Control, Statens Serum Institut, Copenhagen, Denmark
2. 2. National Food Institute, Technical University of Denmark, Lyngby, Denmark
3. 3. Department of Clinical Microbiology, Herlev and Gentofte Hospital, Copenhagen University Hospital, Herlev, Denmark
4. 4. Pathogen Genomics Division, Translational Genomics Research Institute (TGen), Flagstaff, Arizona, USA

**Correspondence:** Henrik Hasman (henh@ssi.dk)
Improving the methods

• Benchmarking of methods for identification of AMR genes in microbial WGS data
ResFinder 99.74% accuracy

Studies indicates that information might be lost under the assembly.

SRST2: Rapid genomic surveillance for public health and hospital microbiology labs

Bias the data

Set threshold according to estimated depth

Density

Depth

Estimate the depth of the sample
Test data

143 isolates from Oxford University Hospital, comprising 858 phenotypic susceptibility tests, most on beta-lactams.

193 isolates from Danish pig farms, comprising 2,547 phenotypic susceptibility tests, covering a broad spectrum of antibiotic classes.
Sequence quality and resistance determination

<table>
<thead>
<tr>
<th></th>
<th>SRST2</th>
<th>ResFinder</th>
<th>Kmer</th>
</tr>
</thead>
<tbody>
<tr>
<td>SE</td>
<td>95</td>
<td>94</td>
<td>97</td>
</tr>
<tr>
<td>SP</td>
<td>96</td>
<td>96</td>
<td>97</td>
</tr>
<tr>
<td><strong>Down sampled</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SE</td>
<td>29</td>
<td>17</td>
<td>95</td>
</tr>
<tr>
<td>SP</td>
<td>98</td>
<td>98</td>
<td>97</td>
</tr>
</tbody>
</table>

JAC - accepted
The Surveillance Pyramid

Reported to health unit
Culture-confirmed case
Lab tests for organism
Specimen obtained
Person seeks care
Person becomes ill
Population exposures
• What if people will not share?

• How do we then get a global surveillance?
ONE SPOT GLOBAL RESISTANCE GENE AND PATHOGEN SURVEILLANCE

Human waste from long-distance airplanes is an attractive material for monitoring the occurrence, prevalence and dissemination of antibiotic resistance genes and pathogens.

1. Samples were collected from 18 international flights, which landed in Copenhagen, Denmark. This system can easily be upscaled to cover the entire globe.

2. Each airplane contains roughly 400 litres of human waste. Samples were collected from the storage tank and brought to a research facility for DNA extraction and analysis.

3. In recent years it has become technically and economically feasible to perform complete DNA sequencing of large samples as well as analyse the data computationally. 20 GB was sequenced from each sample, and analyzed.

4. The method offers a way of having global surveillance of all known pathogens close to real-time. In the future it can help monitoring and preventing outbreaks and dissemination of deadly diseases.
Meta-genomic analysis of toilet waste from long distance flights; a step towards global surveillance of infectious diseases and antimicrobial resistance

Thomas Nordahl Petersen¹, Simon Rasmussen¹, Henrik Hasman², Christian Carøe¹, Jacob Bælum¹, Anna Charlotte Schultz², Lasse Bergmark², Christina A. Svendsen², Ole Lund¹, Thomas Sicheritz-Pontén¹ & Frank M. Aarestrup³
Farm surveillance based on manure
Meta-genomic

![Graph showing log RPM tet resistance genes vs. log lifetime exposure ADD/animal](image)
**Figure 1:**

(a) Heatmap showing the correlation between CFU E. coli, MIC E. coli, CFU aerobic, and Metagenomic. The heatmap uses a color gradient from red to blue to represent the strength of correlation. 

(b) Dendrogram illustrating the Ward distance between the categories: CFU E. coli, MIC E. coli, CFU aerobic, Metagenomic, and Exp. res. The distances are labeled as follows: 91, 93, 97.
Global sewage surveillance - 2016
Copenhagen according to sewage
Our next very simple strategy
Building an epidemiological infrastructure – making data assessable

Data:
- World bank (demographic/economic)
- WHO (health)
- FAO (food production and trade)
- Flight connections
- Antimicrobial use

Simple scripts
- R-codes for regression

Search and download

Own analysis

Bioinformatic or measured data

Epidemiology for dummies
Conclusions

• WGS/NGS is rapidly entering diagnostic and public health, with near real time data generation

• Meta-genomic sequencing is superior to conventional and other genomic methods for quantification of AMR and pathogens

• Bottleneck at level of bioinformatics and data sharing, particularly for intergroup comparison, national, international

• Need for infrastructure and ICT to meet the coming demand

• Novel sites should be explored for sampling for example healthy populations

• Combining advanced mathematical modelling with global sampling, meta-genomics and open source research and epidemiology is the next frontier
Our vision: to build one system that serves all

Across locations

across sectors
- Human health
- Animal Health
- Food safety

cross-time

pathogen-independent
- viruses
- bacteria
- protozoa

Across data types
Next logical step

Platform
<table>
<thead>
<tr>
<th>Kingdom</th>
<th>Phylum</th>
<th>Class</th>
<th>Family</th>
<th>Genus</th>
<th>Species</th>
<th>IKEA-BigMac</th>
<th>Lyngby-Whopper</th>
</tr>
</thead>
<tbody>
<tr>
<td>Eukaryota</td>
<td>Chordata</td>
<td>Mammalia</td>
<td>Bovidae</td>
<td>Bos</td>
<td>Bos mutus</td>
<td>47,728767</td>
<td>55,138006</td>
</tr>
<tr>
<td>Eukaryota</td>
<td>Chordata</td>
<td>Mammalia</td>
<td>Bovidae</td>
<td>Bubalus</td>
<td>Bubalus bubalis</td>
<td>16,571103</td>
<td>16,052984</td>
</tr>
<tr>
<td>Eukaryota</td>
<td>Chordata</td>
<td>Mammalia</td>
<td>Bovidae</td>
<td>Capra</td>
<td>Capra hircus</td>
<td>0,779368</td>
<td>0,858108</td>
</tr>
<tr>
<td>Eukaryota</td>
<td>Chordata</td>
<td>Mammalia</td>
<td>Bovidae</td>
<td>Ovis</td>
<td>Ovis aries</td>
<td>0,362244</td>
<td>0,592654</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Vespertilionidae</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Eukaryota</td>
<td>Chordata</td>
<td>Mammalia</td>
<td>Muridae</td>
<td>Rattus</td>
<td>Rattus norvegicus</td>
<td>0,165231</td>
<td>0,151688</td>
</tr>
<tr>
<td>Eukaryota</td>
<td>Chordata</td>
<td>Mammalia</td>
<td>Macropodidae</td>
<td>Macropus</td>
<td>Macropus eugenii</td>
<td>0,153902</td>
<td>0,189751</td>
</tr>
<tr>
<td>Eukaryota</td>
<td>Chordata</td>
<td>Mammalia</td>
<td>Bovidae</td>
<td>Pantholops</td>
<td>Pantholops hodgsonii</td>
<td>0,129919</td>
<td>0,102347</td>
</tr>
<tr>
<td>Eukaryota</td>
<td>Chordata</td>
<td>Mammalia</td>
<td>Bovidae</td>
<td>Bos</td>
<td>Bos taurus</td>
<td>0,088133</td>
<td>0,203566</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Balaenopteridae</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Eukaryota</td>
<td>Chordata</td>
<td>Mammalia</td>
<td>Ursidae</td>
<td>Ailuropoda</td>
<td>Ailuropoda melanoleuca</td>
<td>0,012359</td>
<td>0,008599</td>
</tr>
<tr>
<td>Eukaryota</td>
<td>Chordata</td>
<td>Mammalia</td>
<td>Cebidae</td>
<td>Callithrix</td>
<td>Callithrix jacchus</td>
<td>0,010888</td>
<td>0,009586</td>
</tr>
</tbody>
</table>