Genome Based Surveillance and Implications for Public Health

Lothar H. Wieler, Torsten Semmler, Guido Werner, Tim Eckmanns, Norbert Bannert, Andreas Nitsche, Bernhard Renard

Hamburg, 09.06.2016
Outline

- Robert Koch Institute History
- Tasks and Infrastructure
- Data and Facts
- National Reference and Consultation Laboratory Network
- Genomic Surveillance
  - Viral Pathogens
  - Bacterial Pathogens
- Genomic Diagnostics
  - Viral Pathogens
  - Bacterial Pathogens
- Genomic research
- Bioinformatics
Robert Koch: his work on epidemiology and etiology of TB

Source: Robert Koch, Epidemiology of tuberculosis; Presentation at the Academy of Sciences in Berlin on April 7, 1910
The Robert Koch Institute: From 1891 to 2016

Robert Koch 1843 - 1910
Co-Founder of Modern Bacteriology and Hygiene

„Triangle“, 1891 - 1900
1891 Royal Prussian Institute for Infectious Diseases. Robert Koch founding director until 1904.

1952 Institute as part of the newly founded Bundesgesundheitsamt

1994 Disolution of the Bundesgesundheitsamt independent Federal Institute (Bundesoberbehörde) in the Ministry of Health

Robert Koch Institute 1900
The Robert Koch Institute: Locations
Legally based networks for Alerts in Germany, EU and WHO

- Local Health Authorities
- Regional Health Authorities
- Robert Koch Institute

Infection Protection Act (2001)

IHR
GMLZ
WHO

EU
- EWRS
- TESSy
- Reports

European Commission
ECDC
EFSA

Decision 1082/2013 EU
Directive 2003/99/EG

IHR (2005)

RaMi-NGS
The Robert Koch Institute: Data, Facts

- Employees: approx. 1.100, more than 400 Scientists
- Budget: approx. 88 Mio. Euro (12.9 Mio for Investments)
- More than 120 third party funded research projects (ca. 13 Mio Euro / year)
- Peer-reviewed publications (2012 - 2015): between 438 and 551
- 90 different professions, 50 different academics, 48 trainees
- 74 PhD students (Robert Koch Doktoranden-Kolleg, RoKoDoKo)
- Numerous research/lecture links to Universities in Berlin and Braunschweig in particular
- RKI website highly accessed by general and professional public (40.0 Mio. page impressions / year; 500.000 accesses to Epi. Bull. / month)
Outline

- Robert Koch Institute History
- Tasks and Infrastructure
- Data and Facts
- **National Reference and Consultation Laboratory Network**
- Genomic Surveillance
  - Viral Pathogens
  - Bacterial Pathogens
- Genomic Diagnostics
  - Viral Pathogens
  - Bacterial Pathogens
- Genomic research
- Bioinformatics
National Reference Centres (NRZ) and Consultation Laboratories (KL)

- 19 NRZ (5 at RKI)
- 40 KL (10 at RKI)
National Reference Centres (n = 19)

- **External (n = 14)**
  - Borrelien, Oberschleißheim
  - *Helicobacter pylori, Freiburg*
  - Meningococci and H. influenzae, Würzburg
  - Mykobacteria, Borstel
  - Streptococci, Aachen
  - Invasive fungal infections, Jena
  - Hepatitis B- und D-Viruses, Gießen
  - Hepatitis C-Viruses, Essen
  - Papillom- und Polyomaviruses, Köln
  - *Retroviren, Frankfurt*
  - Tropical Infectious agents, Hamburg
  - Gram-negative nosokomial agents, Bochum
  - Surveillance Transmissibler Spongiformer Enzephalopathien, Göttingen
  - Surveillance von nosokomialen Infektionen, Berlin

- **RKI Internal (n = 5)**
  - Salmonella and other bacterial Enteritis-causing agents
  - Staphylococci and Enterococci
  - Influenza
  - Measles, Mumps, Rubella
  - Poliomyelitis and Enteroviruses

*(10 at Universities, 4 at other Institutions (z. B. Regional Health Authorities, private laboratories))

*Italics = new submission in 2016*
Consultant Laboratories (n = 40)

- **External (n = 30)**
  - Adenoviruses
  - Anaerobic bacteria
  - Bartonella
  - Bordetella pertussis
  - Brucella
  - *Chlamydia*
  - Clostridium difficile
  - Coxiella burnettii
  - Coronavirus
  - Dermatophytes
  - Diphtheria
  - Echinococci
  - Filoviruses
  - FSME
  - Gonococci
  - Hantaviruses
  - Hepatitis-A & Hepatitis-E-Virus
  - Herpes-simplex-Virus & Varizella-zoster-Virus
  - HUS
  - Legionella
  - Leptospiro
  - Mukoviszidosis-bacteriology
  - Mykoplasma
  - Parvoviruses
  - Rabies
  - Toxoplasma
  - Treponema
  - Tropheryma whipplei
  - Yersinia pestis
  - Zytomegalieivirus

- **RKI Internal (n = 10)**
  - Anthrax
  - Clostridium botulinum
  - Electron microscopic
  - Ddagnostics of infectious agents
  - *Kryptococcus and imported Systemmycoses*
  - Listeria
  - *Noroviruses*
  - Poxviruses
  - *Rotaviren*
  - RSV, PIV & hMPV
  - Tularemia

*Italics* = new submission in 2016
Outline

- Robert Koch Institute
  - History
  - Tasks and Infrastructure
  - Data and Facts
- National Reference and Consultation Laboratory Network
- Genomic Surveillance
  - Viral Pathogens
  - Bacterial Pathogens
- Genomic Diagnostics
  - Viral Pathogens
  - Bacterial Pathogens
- Genomic research
- Bioinformatics
Public Health Surveillance

Public health surveillance is defined by the World Health Organization (WHO) as a systematic ongoing collection, collation, analysis, and interpretation of data and a timely dissemination of information (http://www.who.int/topics/public_health_surveillance/en/).

Two main pillars of surveillance are existing:
- Indicator based
- Event based surveillance.

Indicator based surveillance can be reporting of physicians and laboratories. If syndromic surveillance is organized in a structured way it is also indicator based. For event based surveillance often media and/or rumors and alerts are monitored and assessed for relevance.

Integrated surveillance: covering various diseases - ideally also animal diseases.
Bioinformatic methodology
NGS-based surveillance
Sentinel Surveillance of recently acquired HIV infections: Sampling strategy

- Serum of 60% of new HIV diagnoses
- 3200 DSS filters/year
- 82 laboratories

- Gender
- Transmission group
- Country of origin
- Country of infection
- Other data
Sentinel Surveillance of recently acquired HIV infections:
Workflow in the Lab

- DSS sample receipt (3200 per year)
- Recency Assay
  - BED IgG capture ELISA
- Recent infections (1000 samples/year)
  - RNA isolation
- RT-PCR of sequences associated with resistance
- Population sequencing + HIV-genotyping
  - (Resistance and Subtype)
  - Analysis using the linked socio demographic & clinical data from HIV-report
Sentinel Surveillance of recently acquired HIV infections: Transmitted drug resistance (TDR) in recent HIV infections (Germany 2013-2015)

- Proportion (\%): 3.9%, 3.0%, 2.8%, 1.0%, 0.3%
- PTrend > 0.05 (CI 95%: 9.3-12.6)

- Stable TDR rate

NRTI, PI, NNRTI, dual, multi
Contact investigation: tracing and interrupting transmission as a priority for TB control

- Source case investigation
- Contact tracing
- Molecular typing of pathogens

Classical epidemiological contact tracing

Latent TB infection
Bact. negative active TB
Bact. confirmed active TB
A joint cross-border investigation of a cluster of MDR-TB in Austria, Romania and Germany using classic, genotyping and whole-genome-sequencing.
Outline

- Robert Koch Institute
  - History
  - Tasks and Infrastructure
  - Data and Facts
- National Reference and Consultation Laboratory Network
- Genomic Surveillance
  - Viral Pathogens
  - Bacterial Pathogens
- **Genomic Diagnostics**
  - Viral Pathogens
  - Bacterial Pathogens
- Genomic research
- Bioinformatics
Examples Outbreak support (e.g. Ebola)  ★ Strengthening IHR-Systems  ★ Biosafety

09.-10. Juni 2016  RaMi-NGS
First proof of Monkeypox in wild-living monkeys

- Infant mangabey found dead in the Tai National Park (Fabian Leendertz)
- Multiple skin lesions typical of poxvirus infection
- Comparison of classical virus propagation plus NGS and direct NGS

Radonic A. et al., EID (2014)
First proof of Monkeypox in wild-living monkeys

- Direct sequencing revealed the virus type rapidly
- New virus belongs to the West African clade (Radonic A. et al. EID 2014)
  - Western African clade viruses are known to be of low virulence

Radonic A. et al., EID (2014)
NGS-based virus diagnostics: Challenges

- Viral genomes are comparably small
  - 5 kb to 200 kb
- Viruses require cells for replication
  - Cellular genome 3.3 x10^9 bp
- Unfavorable ratio of viral to cellular nucleic acid in clinical specimen
  - Read numbers attributable to viruses in clinical specimens usually low

- Viral genomes can consist of DNA or RNA
  - Different nucleic acid preparations needed
  - Ratio of viral to cellular RNA variable (transcriptional state etc.)
NGS-based virus diagnostics: Approaches

- Generation of a dramatically increased number of NGS-reads
  - At monetary and bioinformatics cost
- Enrichment of virus particles (e.g. TUViD, Kohl et al. EID 2015)
- Depletion of background (capture technology)
  - Previous knowledge of virus type helpful
  - Risk of losing the relevant viruses
- Pre-amplification of viral target sequences (Brinkmann et al. in preparation)
  - Loss of open view
- Routine application: HIV-resistance-testing and subtype determination on plasma/serum viruses following RNA isolation (with or without pre-amplification (RT-PCR) of viral target sequences
- Combination of different approaches
Healthcare associated outbreak 1: *Klebsiella pneumonia* (ESBL)

- Bremen 2009 to 2012
- Neonatal Intensive Care Unit (NICU)
- Root: 2.5 years before outbreak was recognised

Healthcare associated outbreak 2: *Klebsiella pneumonia* (ESBL)

Cluster 1
(2015/11)

putative new Cluster (Cluster 3)

Cluster 2
**K. pneumoniae: Deducing a diagnostic MP-PCR from NGS data**

Primer pair 1 - product: **606bp**
4160kb-F: GTTAGAATCAAGATGCACAGTACGC
4160kb-R: CCATGGATTGAACCTTGGTGAG

Primer pair 2 - product: **379bp**
Hem-F: GGGTTTGTTGTATAAATGCCACG
Hem-R: CCCAATCGCTTTATTTTCTGACG

Primer pair 3 - product: **265bp**
Unique-F: CACAAATTCCATCTGGTCATG
Unique-R: CCACCAAGCTAAATCTTCGCTG

**Interpretation**

- Presence of all three bands indicates the outbreak strain

<table>
<thead>
<tr>
<th>S</th>
<th>1</th>
<th>2</th>
<th>1</th>
<th>1</th>
<th>1</th>
<th>1</th>
<th>1</th>
<th>1</th>
<th>1</th>
<th>1</th>
<th>1</th>
<th>1</th>
<th>N</th>
<th>P</th>
</tr>
</thead>
</table>
1, outbreak strain
2, non-outbreak strain
N, negative control strain
P, positive control strain
S, size marker
Outline

- Robert Koch Institute
  History
  Tasks and Infrastructure
  Data and Facts
- National Reference and Consultation Laboratory Network
- Genomic Surveillance
  Viral Pathogens
  Bacterial Pathogens
- Genomic Diagnostics
  Viral Pathogens
  Bacterial Pathogens
- Genomic Research
- Bioinformatics
One Health: CTX-M-15 *E. coli* of ST410 crossing host barriers

(A)

<table>
<thead>
<tr>
<th>Isolate</th>
<th>Origin</th>
<th>Date</th>
</tr>
</thead>
<tbody>
<tr>
<td>IMT28707</td>
<td>mute swan cloacal isolate</td>
<td>01/12</td>
</tr>
<tr>
<td>IMT28764</td>
<td>mute swan cloacal isolate</td>
<td>01/12</td>
</tr>
<tr>
<td>IMT31352</td>
<td>dog feces isolate</td>
<td>06/13</td>
</tr>
<tr>
<td>IMT33180</td>
<td>human clinical isolate</td>
<td>10/09</td>
</tr>
<tr>
<td>IMT33204</td>
<td>human clinical isolate</td>
<td>03/10</td>
</tr>
<tr>
<td>IMT33181</td>
<td>human clinical isolate</td>
<td>07/09</td>
</tr>
<tr>
<td>IMT30467</td>
<td>bean goose cloacal isolate</td>
<td>11/12</td>
</tr>
<tr>
<td>IMT31359</td>
<td>dog feces isolate</td>
<td>06/13</td>
</tr>
<tr>
<td>IMT31487</td>
<td>dog clinical isolate</td>
<td>07/13</td>
</tr>
<tr>
<td>IMT31351</td>
<td>dog feces isolate</td>
<td>06/13</td>
</tr>
</tbody>
</table>

(B)

Cluster I
- max. 75 SNPs difference between isolates
- 0.2-14.4 SNPs/Mbp

Cluster II
- max. 87 SNPs difference between isolates
- 0-16.7 SNPs/Mbp

Phylogeny of ST131 (n=228) from various sources (Maximum likelihood, 9 non-ST131 outliers isolates to root the phylogeny)

McNally et al. (submitted)
Phylogeny of ST131 (n=228) from various sources

- multiple circulating clones of *E. coli* ST131
- each contains a fixed plasmid repertoire
- each has undergone clonal expansion and global dissemination

The core genome alignment (BAPS clusters, BratNextGen analysis), CTX-M gene type, accessory gene profile cluster (KPax2).
(The accessory gene profile phylogeny is colour coded by overlaying the accessory genome profile)
Pandemic ST131 (n=228) from various sources: Recombination

McNally et al. (submitted)
Outline

- Robert Koch Institute
  History
  Tasks and Infrastructure
  Data and Facts
  National Reference and Consultation Laboratory Network
- Genomic Surveillance
  Viral Pathogens
  Bacterial Pathogens
- Genomic Diagnostics
  Viral Pathogens
  Bacterial Pathogens
- Genomic Research
- Bioinformatics
Use of NGS for clinical microbiology

**Major Challenges**

- requires major changes in the organization, skill mix and infrastructure of diagnostic laboratories
- strengthening competence in bioinformatics and software development
- Advances are required in databases, efficient software and algorithms for analysis, software that automatically updates knowledge bases and sophisticated links between pathogen genomics databases and patient clinical record systems
**E. coli**: How big is the Core-Genome?

- Size of core- und pan-genome depends of the numbers and relatedness of genomes used
- Core-Genome: Seems to converge at around 1400 gene families
- Pan-genome: more than 53,000 genes in the Ecore set of 429 strains
- Each *E. coli* strain has only ~10% of all *E. coli* genes!
Finding homologous genes

- Clustering was used to group genes into gene families
- We used usearch/uclust, a fast heuristic clustering, and a clustering threshold of 70% protein sequence identity
- uclust performs a star-like clustering and calculates only the similarity to one reference sequence for each cluster
Defining the Maximum Common Genome (MCG)

• The MCG is defined as the set of conserved genes occurring in every of the considered genomes.

• The size of the MCG changes depending on the group looked at:
  ➢ The more related the genomes are, the bigger will be the MCG.

• Genes in the MCG don't need to be essential. Essential genes don't need to be in the MCG, but usually the overlap is quite big.

von Mentzer et al., Nature Genetics (2014)
Quality Control

- Assemblies
- Read Mapping
- Phylogenetic Rooting
- Quality Control

ContigA1
ContigA2
ContigA3
ContigA4

Kuhring et al., BMC Bioinformatics, 2015

Giese et al, Bioinformatics 2013

Calvignac-Spencer et al., PLoS Currents Outbreaks, 2014

09.-10. Juni 2016 RaMi-NGS
Emergence of Zaire Ebola Virus Disease in Guinea

Phylogenetic Analysis of Guinea 2014 EBOV Ebolavirus Outbreak

Clock Rooting Further Demonstrates that Guinea 2014 EBOV is a Member of the Zaire Lineage

12 September 2014
• Genome analysis of Ebolavirus from 78 patients in Sierra Leone
• Confirmation of phylogenetic placement based on new data
Outbreak lies within lineage of prior outbreaks => Import from Central Africa
Computational Metagenomics

- Strain level quantification

<table>
<thead>
<tr>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>B</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>C</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>D</td>
</tr>
</tbody>
</table>

Lindner & Renard, Nucleic Acids Research, 2013

Piro et al, Bioinformatics, 2016

- Characterizing known unknowns

Lindner et al., PLoS ONE, 2015

09.-10. Juni 2016 RaMi-NGS
Real-Time Sequencing

NGS

Conventional analysis

HiLive

computer idle time for conventional read mapping

-55:56 Start
-29:45 First data available HiLive starts
00:00 Sequencing finished bcl2fastq starts
00:10 HiLive finished
00:48 bcl2fastq finished Mapping starts
12:31 Mapping finished

Time [hh:mm]

BWA
Thanks indeed for your attention