Bacterial Next Generation Sequencing - nur mehr Daten oder auch mehr Wissen?

Dag Harmsen
Univ. Münster, Germany
dharmsen@uni-muenster.de

Commercial Disclosure

Dag Harmsen is co-founder and partial owner of a bioinformatics company (Ridom GmbH, Münster, Germany) that develops software for DNA sequence analysis. Recently Ridom and Life Technologies (Carlsbad, CA, USA) partnered and released SeqSphere+ software to speed and simplify bacterial typing.

The Well-Traveled Salad.

Do You Know Where Your Food Has Been?

As consumers, many of us fail to recognize that even our domestic and local food supplies are part of a global network. The daily activity of consuming food directly links our health as humans to the health of crops and produce, food animals, and the environments in which they are produced.

LETTUCE

Canada, Chile, Dominican Republic, Mexico, Peru, USA

CUCUMBERS

Canada, Honduras, India, Mexico, Spain, USA

FETA CHEESE

Canada, Denmark, Egypt, Germany, Greece, Israel, Italy, Turkey, UK, USA

VINAIGRETTE

Argentina, Brazil, Canada, Chile, China, France, Germany, Greece, India, Indonesia, Italy, Mexico, Morocco, Peru, Portugal, Spain, Thailand, Tunisia, Turkey, USA, Vietnam

OLIVES

Greece, Israel, Mexico, Spain, USA

SPROUTS

Argentina, Australia, Bangladesh, Canada, China, Egypt, France, India, Morocco, Nepal, Pakistan, South Africa, Spain, Turkey, USA

MANDARIN ORANGES

Israel, Mexico, Morocco, South Africa, Spain

CROUTONS

Argentina, Australia, Brazil, Canada, China, France, India, Mexico, Netherlands, Poland, Russia, Switzerland, Uruguay, USA, Vietnam

TOMATOES

Canada, Dominican Republic, Holland, Israel, Italy, Mexico, USA

ONIONS

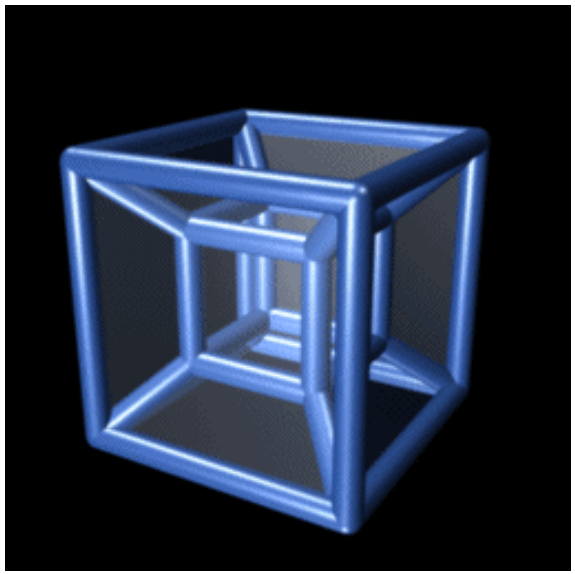
Canada, China, Germany, India, USA

A "One Health" approach to food safety—bringing together expertise and resources from the clinical, veterinary, wildlife health, and ecology communities—has the potential to reveal the sources, pathways, and factors driving the outbreaks of foodborne illness and possibly prevent them from occurring in the first place.

NOTE: Countries are listed in alphabetical order and not by volume of export.

INSTITUTE OF MEDICINE
OF THE NATIONAL ACADEMIES
www.iom.edu

Fourth Dimension Needed for More Specific Surveillance



Place, Time, ‘Person’ …Type!

Fourth Dimension Reloaded

Next Generation Sequencing - Bench-top Machines

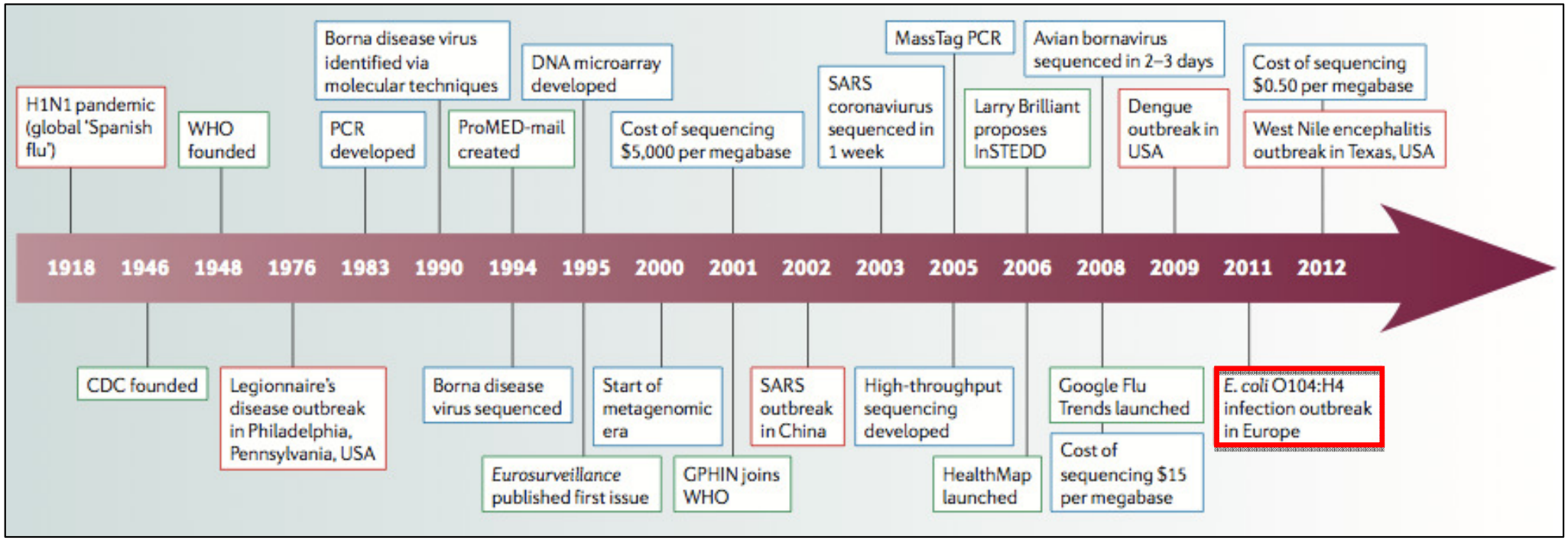
Ion Torrent Personal Genome Machine (**PGM**)

Roche/454 GS **Junior**

illumina **MiSeq** Personal Sequencing System

- Affordable
- **Speed**
- Simple workflow

Timeline – Microbial Surveillance and Disease



Red boxes indicate disease outbreaks, blue boxes indicate technological advances, and green boxes indicate events relating to surveillance. *E. coli*, *Escherichia coli*; GPHIN: Global Public Health Intelligence Network; InSTEDD: Innovative Support to Emergencies, Diseases and Disasters; ProMED-mail: Program for Monitoring Emerging Infectious Diseases; SARS: severe acute respiratory syndrome.

Lipkin (2013). *Nature Rev. Microbiol.* **11:** 133 [PubMed].

Epidemiology: food supply chain analysis, etc.

Bioinformatics: new tool developments and evaluation data-sets

Laboratory: nearly real-time prospective microbial genomics & platform comparisons

Prospective Genomic 'Ad Hoc' Epidemiology

Phylogenetic Analysis of STEC/EHEC 0104:H4

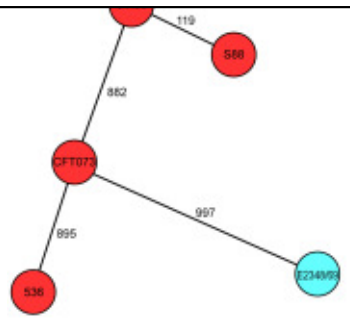
OPEN ACCESS Freely available online

PLoS one

Prospective Genomic Characterization of the German Enterohemorrhagic *Escherichia coli* O104:H4 Outbreak by Rapid Next Generation Sequencing Technology

Alexander Mellmann[1☯], Dag Harmsen[2*☯], Craig A. Cummings[3☯], Emily B. Zentz[4], Shana R. Leopold[1], Alain Rico[5], Karola Prior[2], Rafael Szczepanowski[2], Yongmei Ji[3], Wenlan Zhang[1], Stephen F. McLaughlin[3], John K. Henkhaus[4], Benjamin Leopold[1], Martina Bielaszewska[1], Rita Prager[6], Pius M. Brzoska[3], Richard L. Moore[4], Simone Guenther[5], Jonathan M. Rothberg[7], Helge Karch[1]

1 Institute of Hygiene, University Münster, Münster, Germany, **2** Department of Periodontology, University Münster, Münster, Germany, **3** Life Technologies, Foster City, California, United States of America, **4** OpGen, Gaithersburg, Maryland, United States of America, **5** Life Technologies, Darmstadt, Germany, **6** Robert Koch Institute, Wernigerode Branch, Wernigerode, Germany, **7** Ion Torrent by Life Technologies, Guilford, Connecticut, United States of America



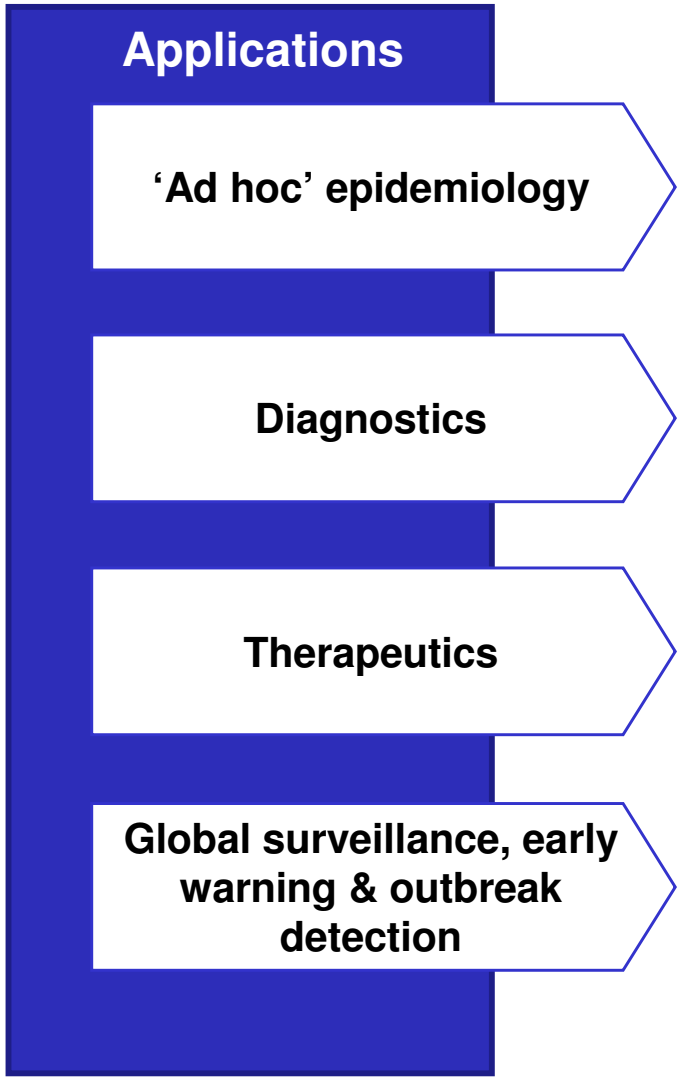
ExPEC
EPEC
ETEC
Commensal
Hypothetical O104:H4 Progenitor

EAEC strain 55989 share the same MLST ST 678. However, the three strains are clearly separated by MLST+.

Mellmann *et al.* (2011). *PLoS One.* **6:** e22751 [PubMed]

F1000 FACULTY of 1000

Rapid NGS Applications in Clinical & Public Health Microbiology



Details

- Introduction of benchtop Next Generation Sequencing (NGS) machines, enables small- and medium-sized laboratories ('democratizing of NGS') to perform 'ad hoc' genomic prospective epidemiology

- Speciation / identification & pathogenicity profiling
- Molecular diagnostic screening tests
- Ultra-deep sequencing for pathogen discovery from human tissues (e.g., hemorrhagic viruses)

- Susceptibility profiling
- Vaccine preventability
- Reverse vaccinology (rationale vaccine design)
- Non-targeted new drug detection

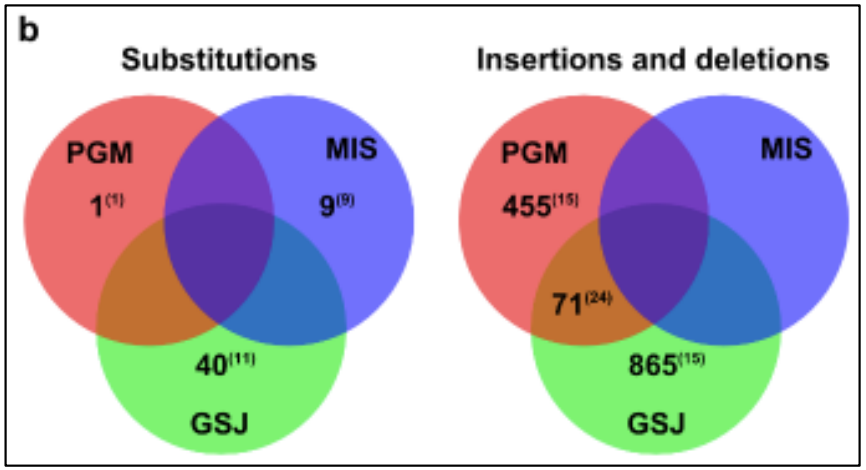
- Standardized Whole Genome Shotgun [WGS] NGS for detection of transmission between individuals
- Outbreak detection, i.e., establishing the spread of particular strains locally or regionally
- Longer-term and evolutionary studies to identify the emergence of particularly pathogenic or virulent variants

Laboratory Improvements (since June 2011)

Loman *et al.* (2012). *Nature Biotechnology* **30:** 562 [PubMed].

It's the Consensus
Genome-wide Gene by Gene *de novo* Consensus Accuracy

Venn diagram of *de novo* consensus accuracy for PGM, MiSeq and GSJ



PGM, Ion Torrent Personal Genome Machine **300bp**; **MiSeq**, Illumina MiSeq **2x 250bp PE**;
GSJ, 454 GS Junior with **GSJ Titanium** chemistry;
bp, base pairs

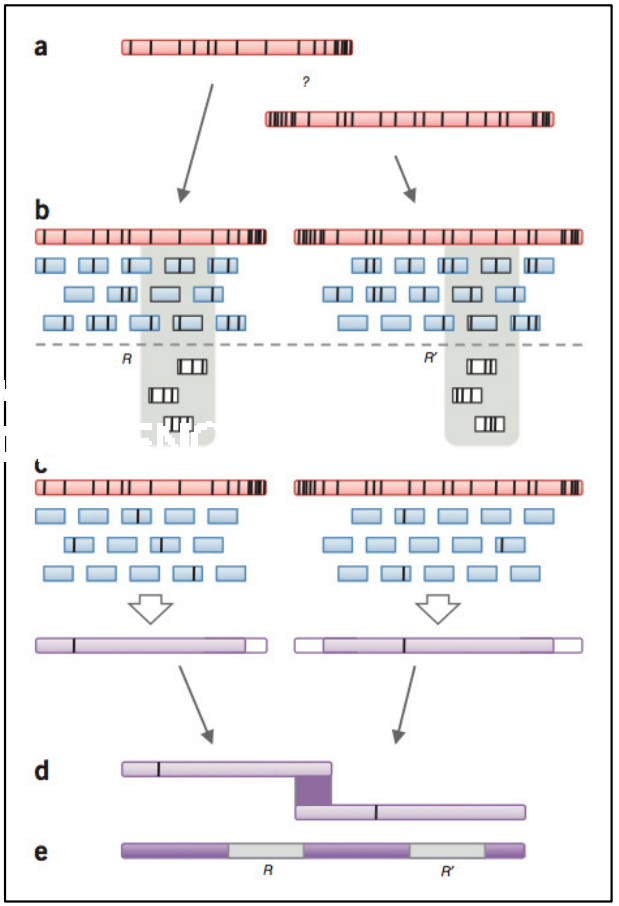
Details

- Consensus errors were analyzed for **4,632 coding NCBI Sakai reference genes** retrieved from **MIRA *de novo* assemblies** using **SeqSphere⁺** for all 3 platforms
- Number of variants confirmed by **bidirectional Sanger sequencing** indicated in parentheses
- Validation of the **8 substitution** and **15 indel** variants identified using all 3 NGS platforms, suggested that either the Sakai strain experienced micro-evolutionary changes or the genome sequence deposited in 2001 contains sequencing errors

Jünemann *et al.* (2013). *Nature Biotechnology* **31:** 294 [PubMed].

Laboratory: Towards Finished Genomes (chromosome & plasmids)

- Fast & easy **mate-pair** protocols (insert > 5kb)
- **Hybrid** 2nd- & 3rd generation **assemblies**

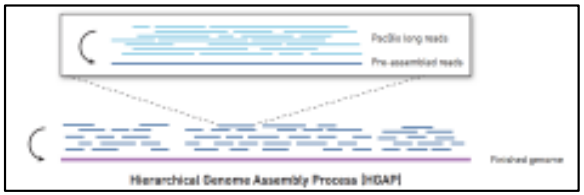


Koren *et al.* (2012). *Nature Biotechnology* **30:** 693 [PubMed].

genia

NANOPORE

- **HGAP** – Hierarchical Genome-Assembly Process (10kb seeds)



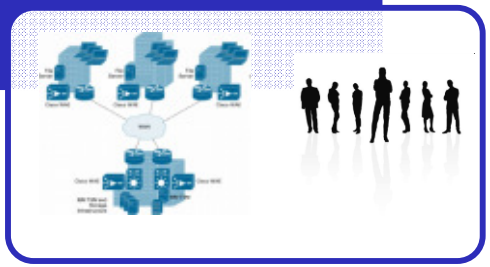
Chin *et al.* (2013). *Nature Methods* **10:** 563 [PubMed].

Current NGS Bottlenecks

Sample Processing

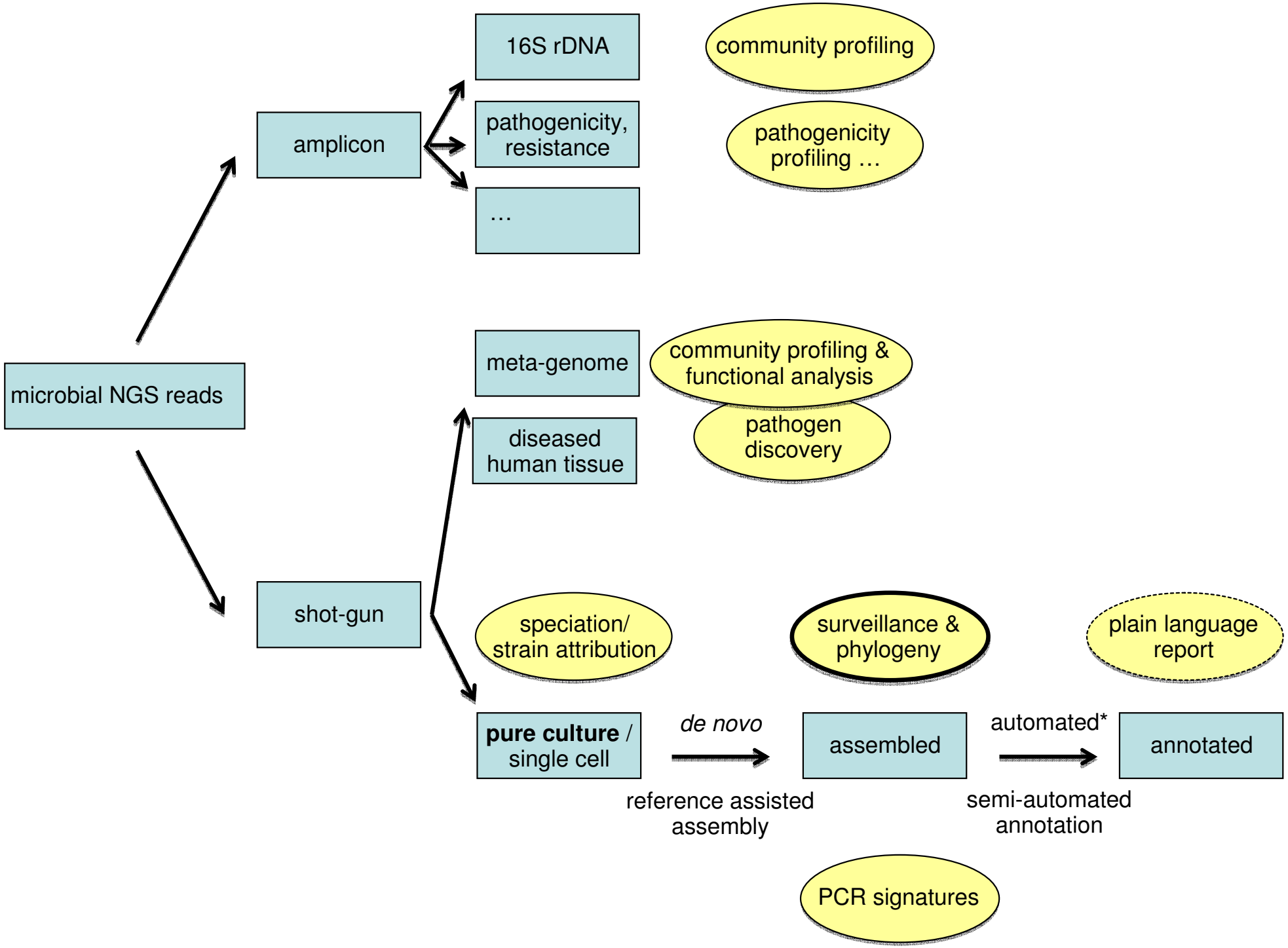
NGS Platforms

Bioinformatics, IT infrastructure



Bioinformatics Challenges

Microbial Bioinformatics Data & Workflow



Microbial NGS Sequence Typing

Pairwise Multiple Genome Alignment (e.g., Mauve)

- Impossible with draft genomes
- Problematic with rearrangement / recombination events
- Not additive expandable

K-mer

+ No assembly needed, quickly find closest matching genome
- Whole genome reduced to a single number of similarity
- Not additive expandable

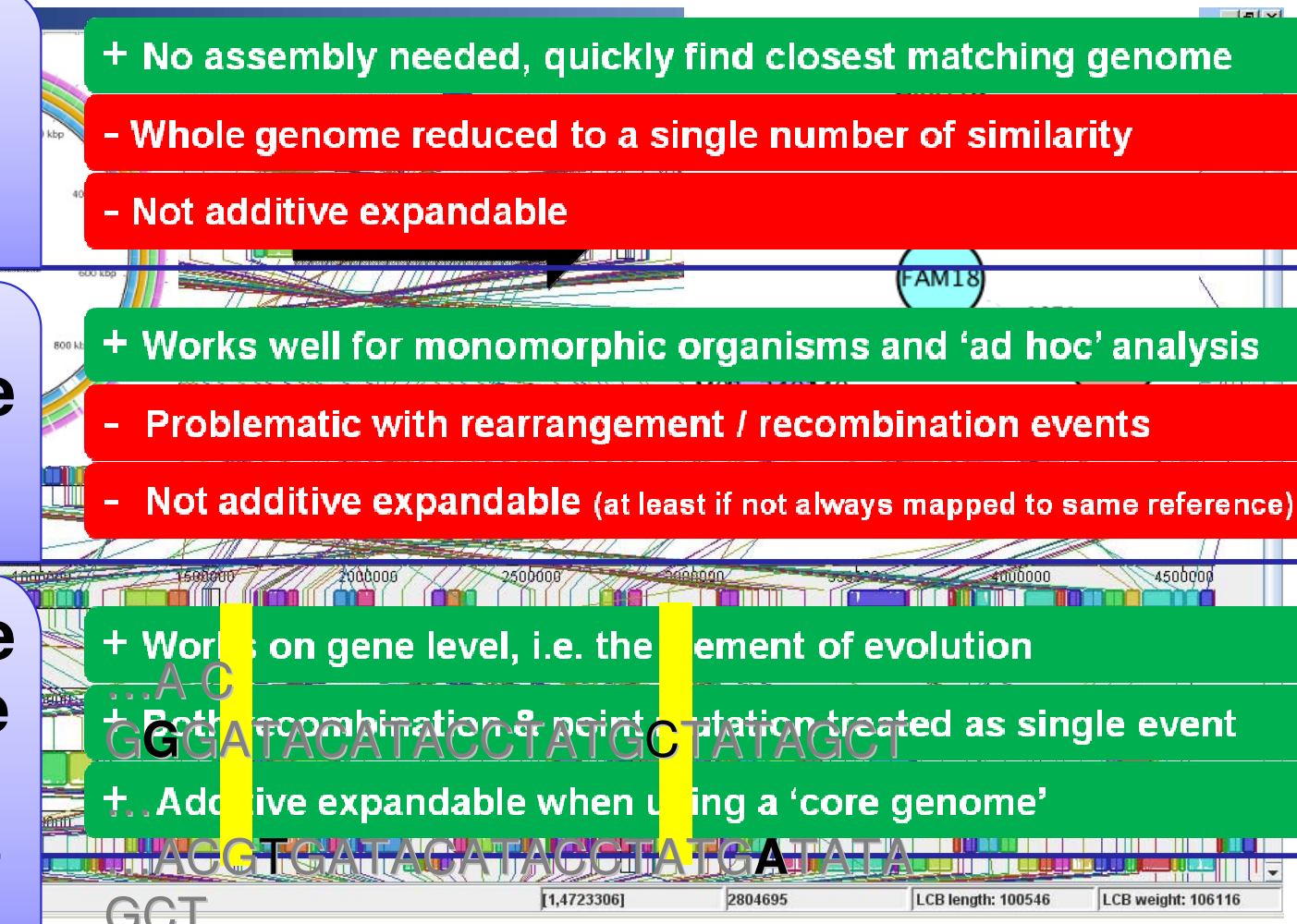
Genome-wide SNP

+ Works well for monomorphic organisms and 'ad hoc' analysis
- Problematic with rearrangement / recombination events
- Not additive expandable (at least if not always mapped to same reference)

Genome-wide gene by gene allele typing

(core genome MLST or MLST+)

+ Works on gene level, i.e. the element of evolution
+ Both recombination & point mutation treated as single event
+ Additive expandable when using a 'core genome'



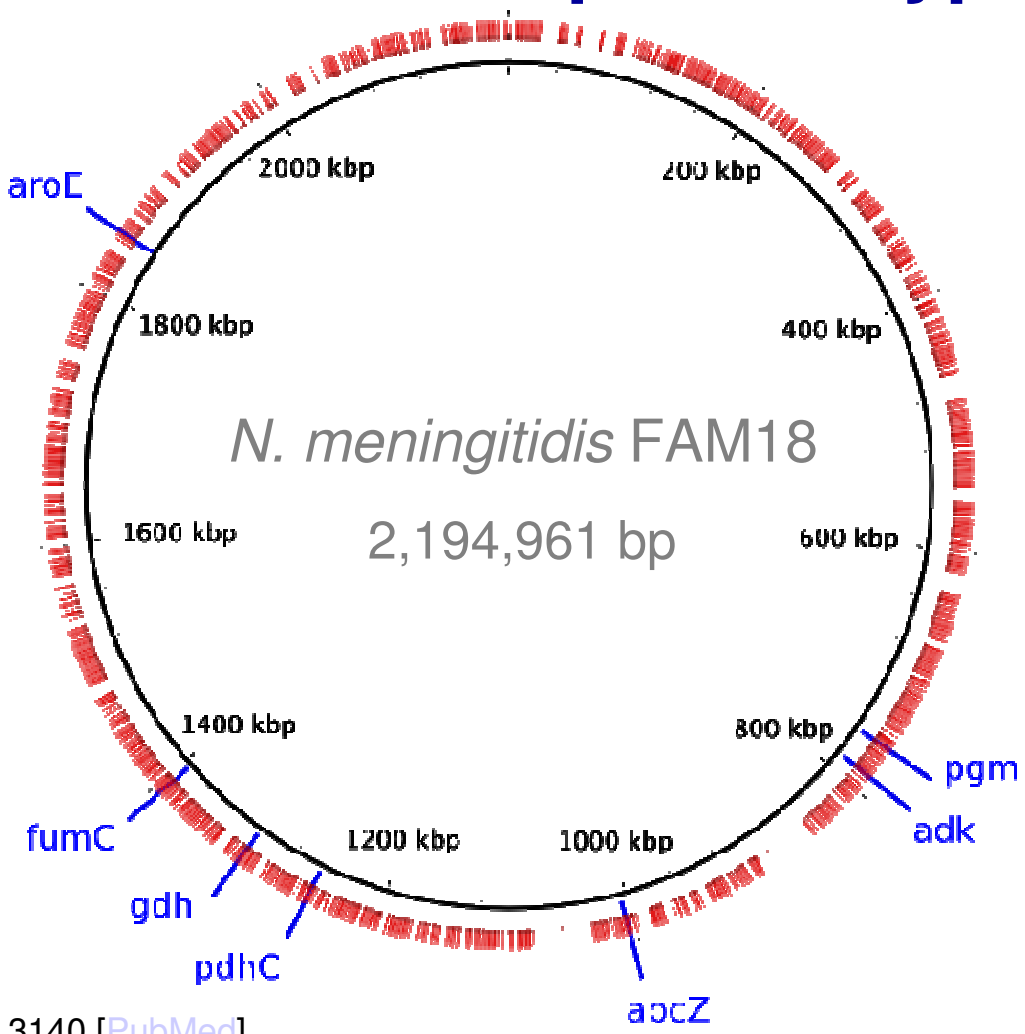
…A C
G GGATACATACCTATGCTATAGCT
…ACGTGATAGATACCTATGATATA
GCT…

The Multi Gene Sequence Typing MLST+

N. meningitidis MLST

7 genes

0.1 %
of FAM18 genome



N. meningitidis MLST+

1241 genes

54.5 %
of FAM18 genome

Maiden *et al.* (1998). *PNAS* **95:** 3140 [PubMed]

MLST

- **5-7 housekeeping genes**
- **Sequence type (ST) and Clonal complex (CC)**
- **Public nomenclature**

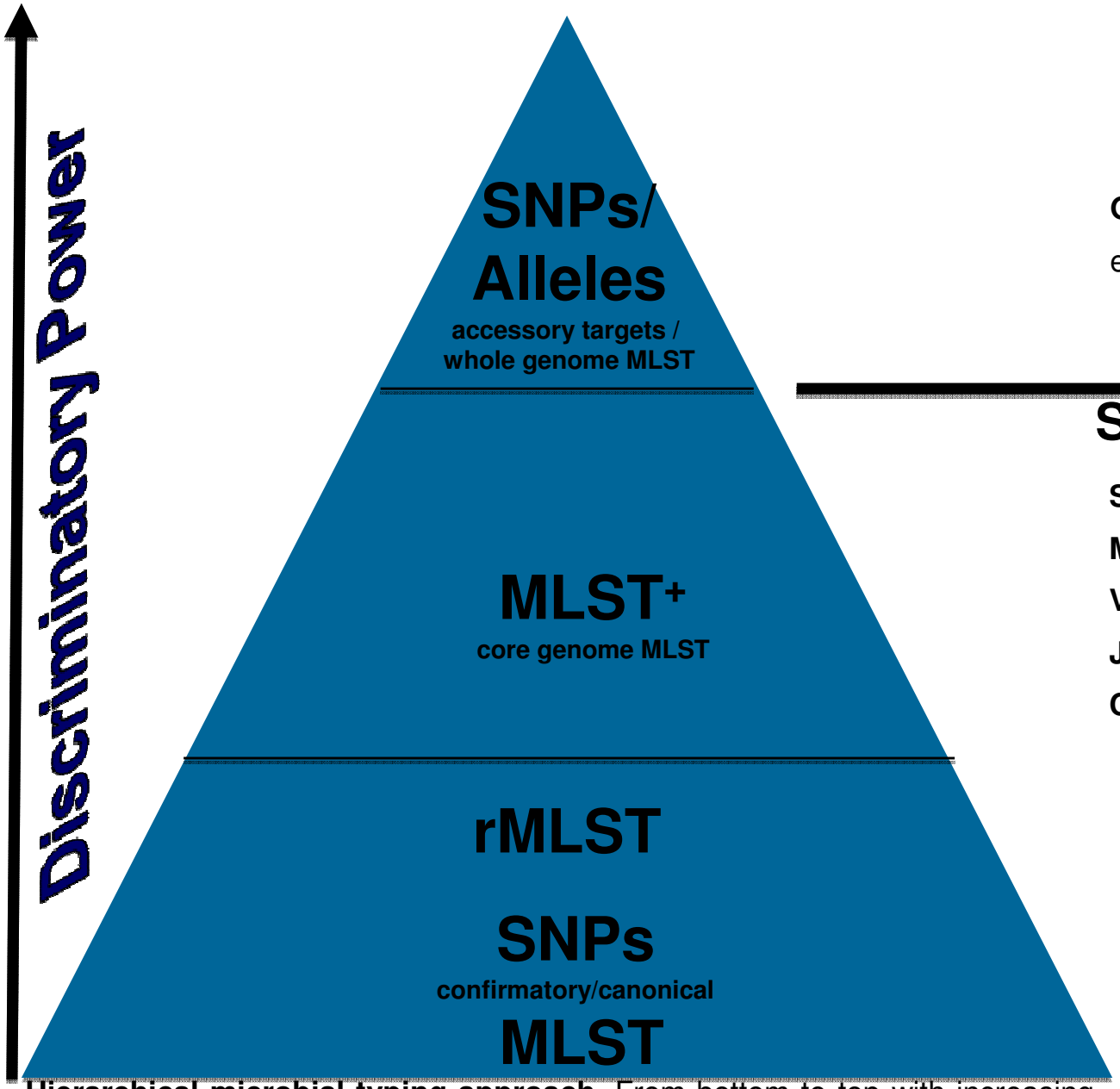
Used mainly for **population genetics & evolutionary studies**

MLST+/cgMLST

- **Hundreds/thousands of 'core genome' genes**
- **Scalable, portable and undstandable**
- **Public, additive expandable nomenclature**

Higher discrimination power for **outbreak investigation**

Standardized Hierarchical Microbial Typing



Hierarchical microbial typing approach. From bottom to top with increasing discriminatory power. MLST, multi locus sequence typing; rMLST, ribosomal MLST; SNP, single nucleotide polymorphism.

Outbreak / Lineage specific

e.g., **Köser** *et al* (2012). *NEJM* **366:** 2267 [PubMed]

Standardized

Species specific

Mellmann *et al.* (2011). *PLoS One*. **6:** e22751 [PubMed]

Vogel *et al.* (2012). *JCM* **50:** 1889 [PubMed]

Jolley *et al.* (2012). *JCM*. **50:** 3046 [PubMed]

Cody *et al.* (2013). *JCM*. **51:** 2526 [PubMed]

Pan-bacterial specific (also suited for speciation)
Jolley *et al.* (2012). *Microbiology* **158:** 1005 [PubMed]

Species specific
e.g., **Van Ert** *et al.* (2007). *JCM* **45:** 47 [PubMed]

Maiden *et al.* (1998). *PNAS* **95:** 3140 [PubMed]
also **needed for backward compatibility**

Maiden *et al.* (2013). *Nature Rev. Microbiol.* **11:** 728 [PubMed].

Antibiotic Resistance by NGS

Antibiotic Resistance Genes Databases

Published online 2 October 2008

Nucleic Acids Research, 2009, Vol. 37, Database issue ***D443–D447***
doi:10.1093/nar/gkn656

ARDB—Antibiotic Resistance Genes Database

Bo Liu[1] and Mihai Pop[1,2,*]

[1]Center for Bioinformatics and Computational Biology and [2]Department of Computer Science, University of Maryland, College Park, MD 20742, USA

ARDB: http://ardb.cbcb.umd.edu/

Liu *et al* (2008). *NAR* **37:** D443 [PubMed]

J Antimicrob Chemother 2012; **67**: 2640–2644
doi:10.1093/jac/dks261 Advance Access publication 10 July 2012

Journal of Antimicrobial Chemotherapy

Identification of acquired antimicrobial resistance genes

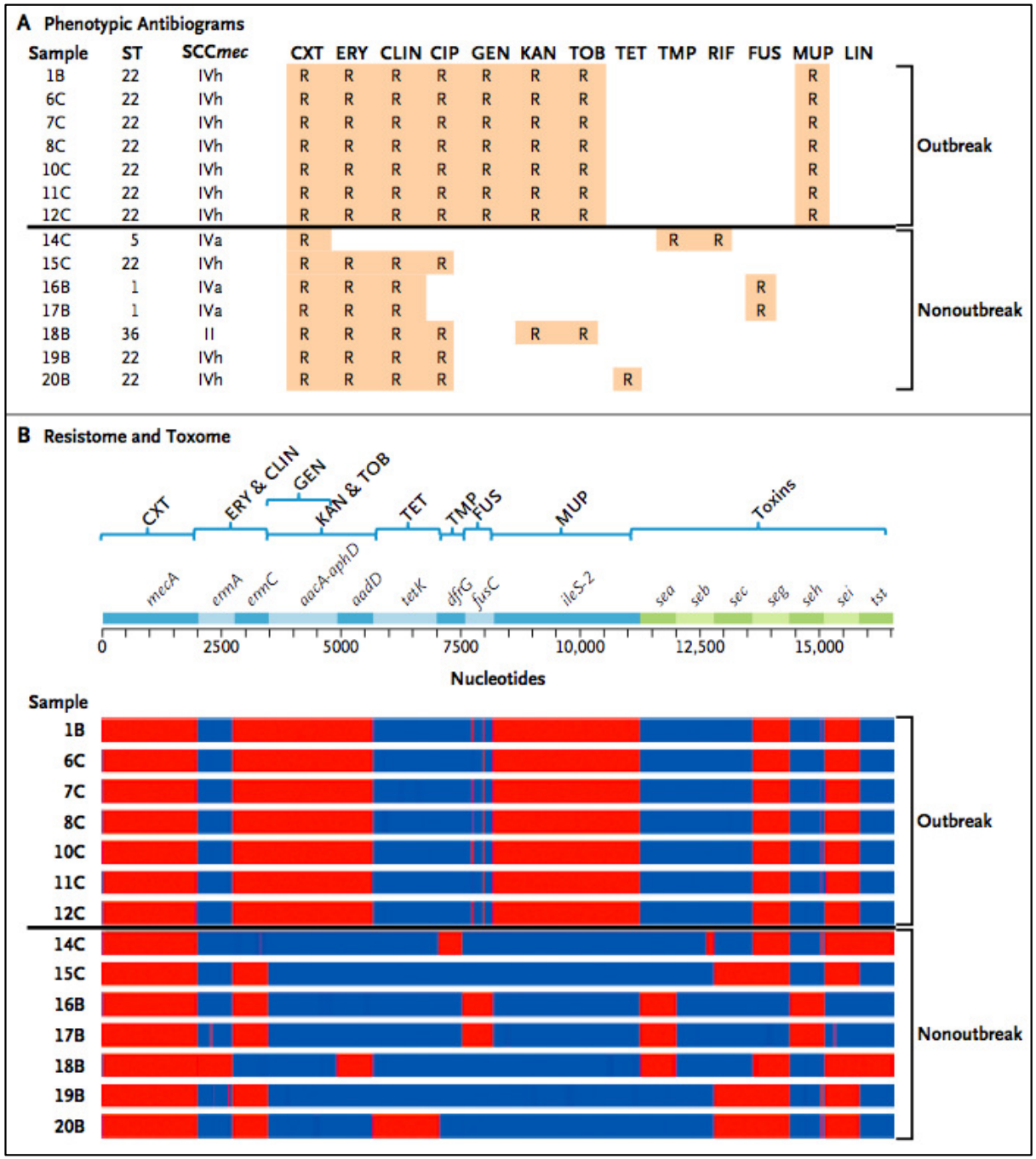
Ea Zankari[1,2*], Henrik Hasman[1], Salvatore Cosentino[2], Martin Vestergaard[1], Simon Rasmussen[2], Ole Lund[2], Frank M. Aarestrup[1] and Mette Voldby Larsen[2]

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

Zankari *et al* (2012). *J Antimicrob Chemother* **67:** 2640 [PubMed]

ResFinder: http://cge.cbs.dtu.dk/services/ResFinder/

Absence/Presence - a Sense of ‚Plain Language Report‘



Phenotypic anti-biograms, resistome, and toxome of **MRSA** hospital outbreak

Köser *et al* (2012). *NEJM* **366:** 2267 [PubMed]

Advanced Antibiotic Resistance NGS Demonstrations

Table 1. Comparison of WGS and Reference Laboratory Testing of Carbapenem-Resistant Gram-Negative Bacteria

Organism	Isolate No.	Phenotypic Resistance to Carbapenems and Third-Generation Cephalosporins	Attributable Resistance Mechanism According to Reference Laboratory[a]	Dominant Resistance Mechanism Based on WGS[b]
Acinetobacter baumannii	AB223	MEM, IPM[c]	OXA-23 carbapenemase	OXA-23 carbapenemase
Enterobacter cloacae	EC1a[d]	ETP, MEM, IPM, CTX, CAZ	IMP-1 carbapenemase	IMP-1 carbapenemase
E cloacae	EC302	ETP, CTX, CAZ	No carbapenemase genes detected. AmpC activity present	No carbapenemase genes detected. OmpF porin loss
Klebsiella pneumoniae	KP652	ETP, CTX, CAZ	No carbapenemase genes detected. ESBL activity consistent with CTX-M. ETP resistance consistent with porin loss	No carbapenemase genes detected. CTX-M-15 ESBL with OmpK36 porin loss
Escherichia coli	Eco216	ETP, CTX, CAZ	No carbapenemase genes detected. ESBL activity present. ETP resistance consistent with porin loss	No carbapenemase genes detected. CTX-M-15 ESBL with OmpF porin loss

Abbreviations: CAZ, ceftazidime; CTX, cefotaxime; ESBL, extended-spectrum β-lactamase; IPM, imipenem; ETP, ertapenem; MEM, meropenem; WGS, whole-genome sequencing.

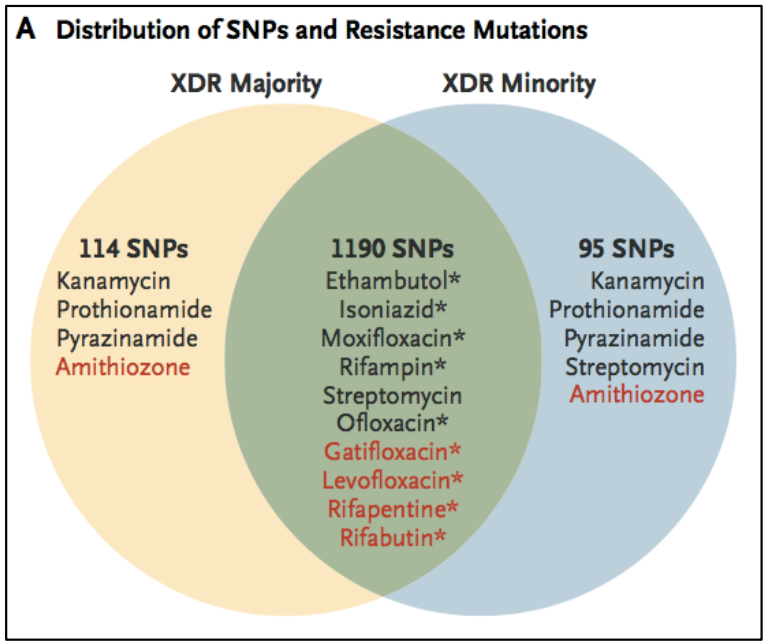
[a] See eMethods in the Supplement.

[b] For a more comprehensive list see eMethods and eTable 2 in the Supplement.

[c] Ertapenem has no activity against A baumannii; no EUCAST (European Committee on Antimicrobial Susceptibility Testing) susceptibility breakpoints available for cefotaxime and ceftazidime.

[d] Isolate from patient EC1 described in *E cloacae* outbreak.

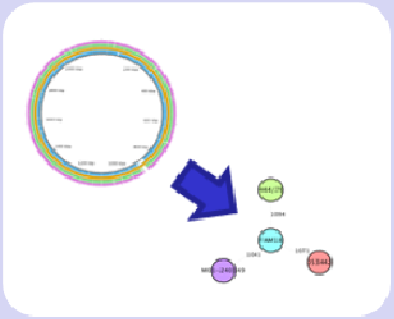
Reuter *et al* (2013). *JAMA Intern Med.* **173:** 1397 [PubMed]

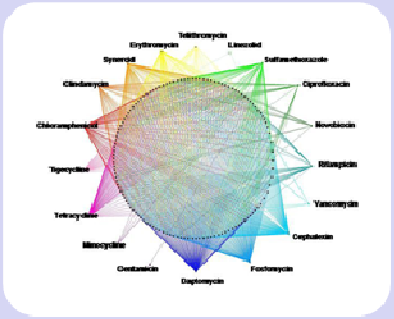


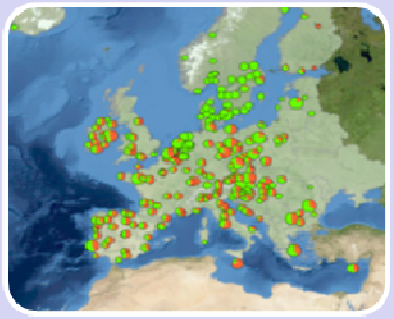
The numbers refer to the number of single nucleotide polymorphisms (SNPs) that were shared by the two strains of extensively drug-resistant (XDR) ***M. tuberculosis*** isolated from a single patient (mixed infection) or were unique to one strain as compared with the *M. tuberculosis* H37Rv reference genome. The reference laboratory had reported resistance to nine antibiotics (black type).

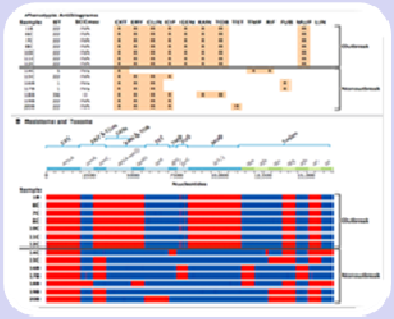
Köser *et al* (2013). *NEJM* **369:** 290 [PubMed]

Future NGS Bioinformatic Developments









Standardize WGS Typing with MLST+	From genotype to phenotype (resistome, pathogenome & toxome analysis)	Early warning & GIS	Plain language report

PATHOGENTrace

http://patho-ngen-trace.eu/

Bacterial Genome Detectives

Base pairs 2 Bedside

Outlook

Assuring Quality of NGS in Clinical Laboratory Practice

Table 1 Selected workgroup recommendations for establishing NGS test systems for clinical use

Requirements for test establishment	Objective	NGS-specific recommendations[a]
Validation	Document reliability of the platform, test, and informatics pipeline before testing of patient specimens	•Platform validation: establish that the system provides reliable sequence analysis across the genomic regions targeted by the test. •Test validation: establish that the system correctly identifies disease-associated (and other) variants in targeted regions of the genome (**Supplementary Guidelines**, section 4). •Informatics pipeline validation: establish that the algorithm(s) reliably analyze platform data to produce an accurate sequence. •Establish and validate alternate methods (for example, Sanger sequencing) to derive high-quality sequence data for problematic genomic re
Quality control	Document reliability of the sequence analysis during patient testing	•Utilize a co in, that mir test is desig •During pati scores, dep bias and tra compared t •Clinically ac analysis usi
Proficiency testing	The independent assessment of test performance	•PT challeng associated genomic re sequence a •Electronic s and variant additional c •PT program by each rec laboratory p
Reference materials	The use of materials for quality management of the analytical phase of testing	•RMs with both naturally occurring and disease-associated sequence variations are needed for test validation, QC procedures and the independent assessment of test performance. •Synthetic DNA and electronic reference data files may serve as RMs for rare or challenging sequence variations. •Efforts should be undertaken to establish a suitable NGS RM and the sequence of the RM should be refined as the technology changes. Such a RM should be annotated to indicate regions of high and low sequence reliability.

[a]See **Supplementary Guidelines** for complete recommendations. RM, reference material.

http://www.globalmicrobialidentifier.org/

genomeweb

CAP Publishes Accreditation Checklist for NGS in Clinical Labs

August 01, 2012

CAP Publishes Accreditation Checklist for NGS in Clinical Labs

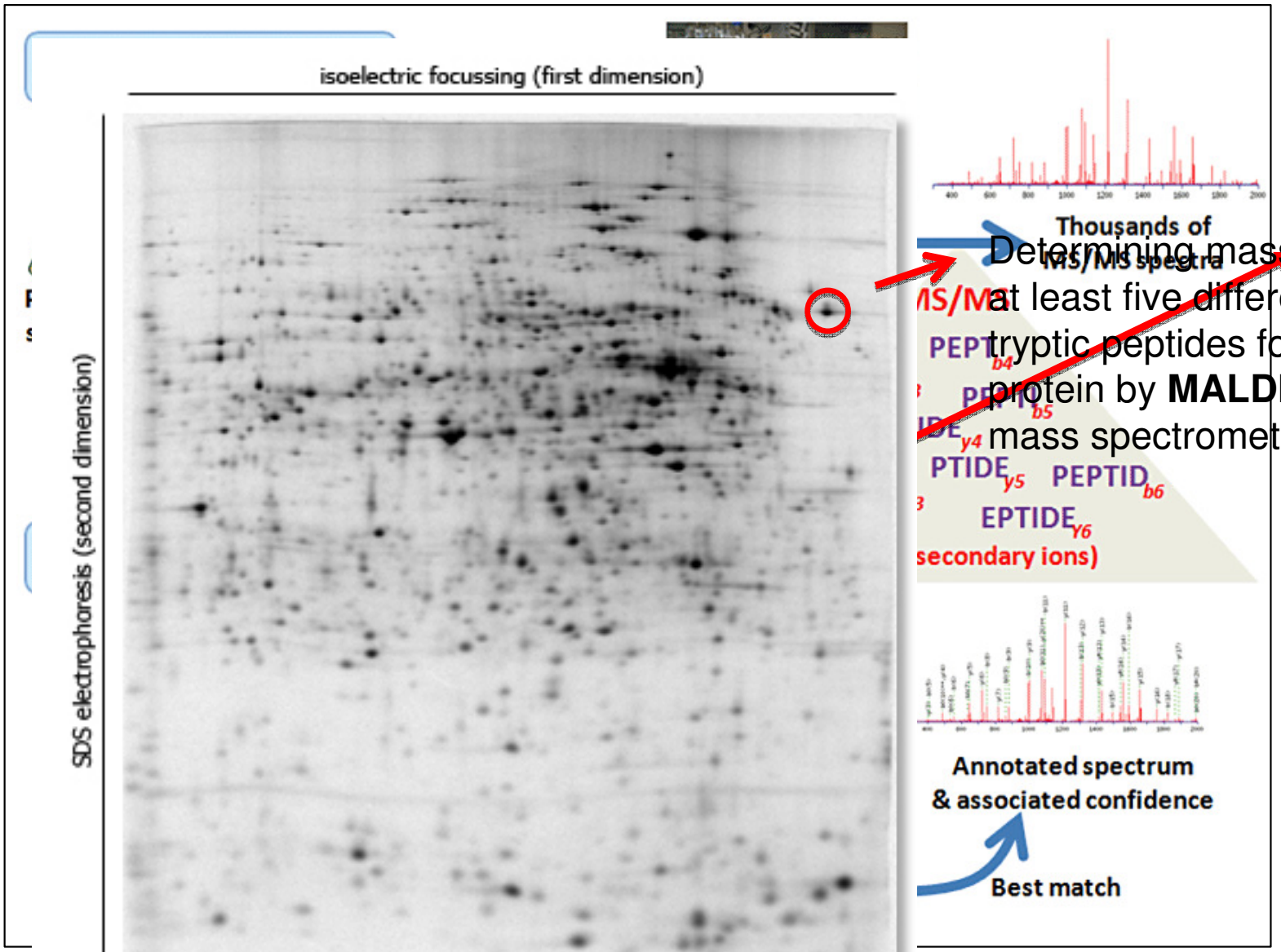
By Monica Heger

The College of American Pathologists has published a checklist specific to next-generation sequencing for clinical lab accreditation. The NGS-specific checklist is part of CAP's revised molecular pathology checklist for accrediting clinical laboratories, released this week.

http://www.cap.org/

Gargis *et al.* (2012). *Nature Biotechnology* **30 (11):** 1033 [PubMed].

Transition from Geno- to Phenotype: Proteomics & Transcriptomics



Determining mass of at least five different tryptic peptides form one protein by **MALDI ToF** mass spectrometry

NGS makes each bacterial isolate quickly amenable to proteomics

- WGS NGS, un-annotated draft genome (translation in all six possible reading frames)

e.g.: **Greub** (2009). *PLoS ONE* **4:** e8423 [PubMed]

- RNA-seq (reverse transcription of mRNA; Whole Transcriptome Shotgun Sequencing)

e.g.: **Bräutigam** (2008). *J Biotechnol.* **136:** 44 [PubMed]

N have pushed out 2D-PAGE Two-dimensional gel electrophoresis (**2D-PAGE**)

Armengaud (2013). *Environ Microbiol.* **15:** 12 [PubMed]

Bacterial Next Generation Sequencing - nur mehr Daten oder auch mehr Wissen?

Dag Harmsen
Univ. Münster, Germany
dharmsen@uni-muenster.de