24.001: Whole genome sequencing for diagnosing infectious diseases

PAUL D BROWN, PHD
THE UNIVERSITY OF THE WEST INDIES
MONA, JAMAICA

ICID2018, BUENOS AIRES, ARGENTINA
SESSION # 24, CLINICAL APPLICATIONS OF WHOLE GENOME SEQUENCING

Disclosures

- Nothing to declare

Main Goals of Clinical Microbiology

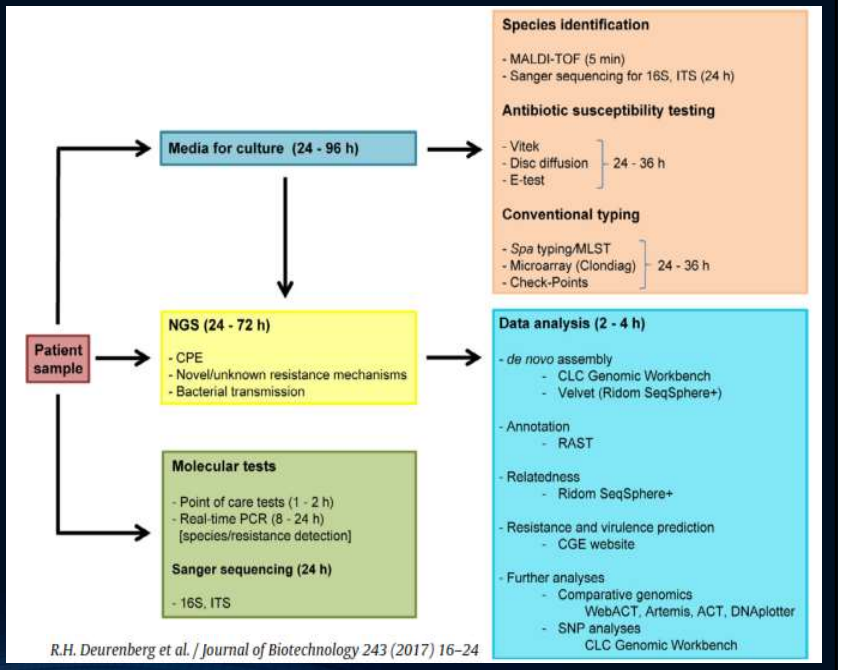
- What organism(s) is(are) causing the infection?
  - Species identification
  - Single or multiple aetiological agents
- What drugs (antimicrobials) can be used to treat it?
  - Phenotypic characterization (AMR and virulence)
- How is it related to similar infections?
  - Evolutionary traits and pathogen tracking
    - Part of outbreak investigations; Important for surveillance

Transforming Clinical Microbiology with WGS

- Allows sequencing of the whole genome of numerous pathogens in one sequence run, either from bacterial isolates of (different) patients, or from multiple species present in material from one individual (metagenomics)
- Both the investment- and the running costs have decreased dramatically during the last decade
- Single protocol can be used for multiple pathogens for identification and characterization (typing)

WGS Strategy

- A step further than molecular diagnostics



Platforms

- **WGS**
- **Whole-exome sequencing (WES)**
- **Metagenomic sequencing**
- **Targeted gene sequencing**

Properties of current NGS platforms.

Company	Equipment	Output/run (Gb)	Maximum read length (bp)	Reads (x10^6)	Running time
Illumina	MiniSeq	0.6–7.5	2 × 150	25	4–24 h
Illumina	Miseq	0.3–15	2 × 300	25	5–55 h
Illumina	NextSeq	20–120	2 × 150	130/400	12–30 h
Illumina	HiSeq 3000	125–700	2 × 150	2500	<1–3.5 days
ThermoFisher	Ion PGM™	0.03–2	200–400	0.4–5.5	2–7 h
ThermoFisher	Ion 5S™	0.6–15	200–400	3–80	2.5–4 h
ThermoFisher	Ion 5S™ XL	0.6–15	200–400	3–80	<24 h
Oxford Nanopore	MinION	21–42	230,000–300,000	2.2–4.4	1 min–48 h
Pacific Biosciences[a]	Sequel	0.75–1.25	>20,000	370,000	30 min–6 h
Pacific Biosciences[a]	RSII	0.5–1	>20,000	55,000	30 min–4 h

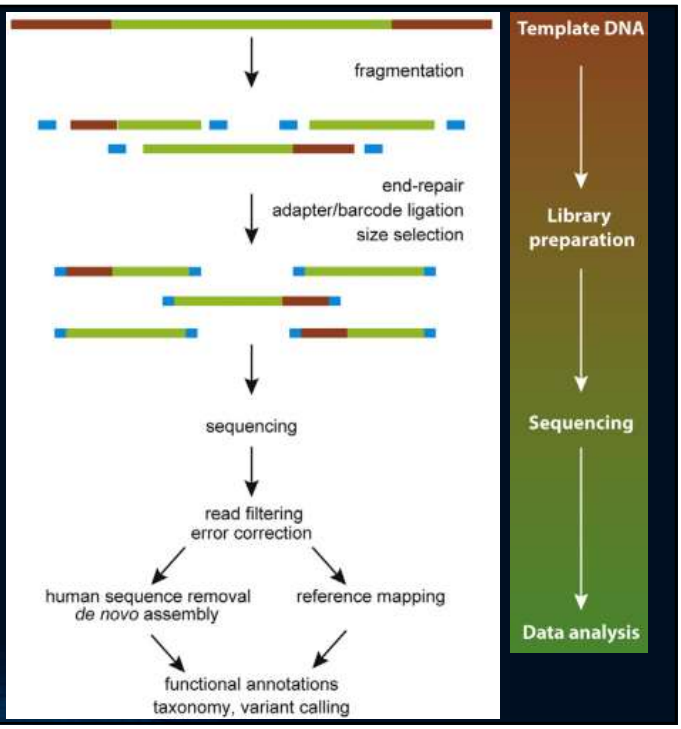
[a] The Pacific Biosciences data are per smart cell; both the Sequel and the RSII can run 1–16 smart cells in one run.

R.H. Deurenberg et al. / Journal of Biotechnology 243 (2017) 16–24

WGS Strategy

- Fragmentation of template DNA
- Library preparation
- Sequencing
- Processing and data analysis
- (Diagnosis and treatment/interventions)

Lefterova et al., J Mol Diagn 2015



Transforming Clinical Microbiology with WGS

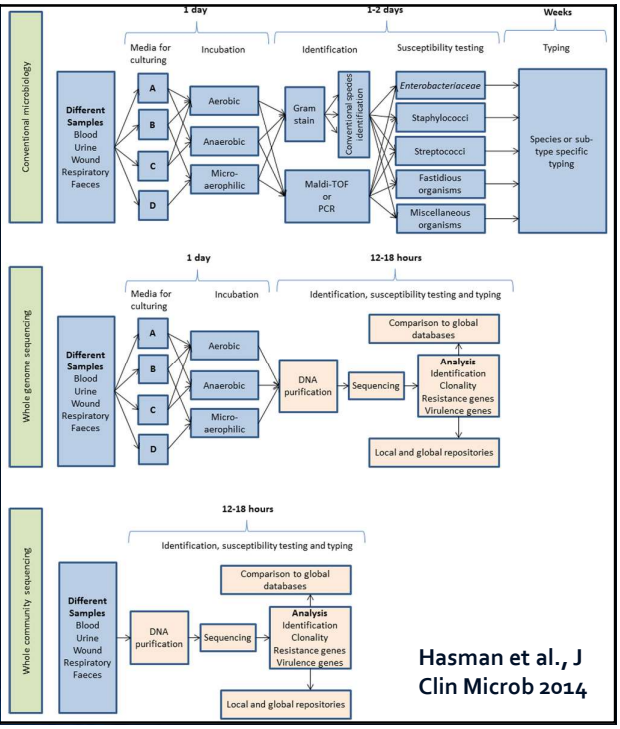
- Workflow anticipated after adoption of WGS of cultured isolates or directly on the clinical samples, with an expected time scale

Table 1. Summary of In Silico Validation Results, Accuracy[a]

Total	Taxa	Simulated Samples	Accuracy, %
RNA viruses	73	230	100
DNA viruses	29	70	100
Bacteria	82	329	99.1
Fungi	1	6	100

[a] The table shows numbers of species and viral types (taxa), number of different reference sequences used to generate simulated samples representing the genetic intraspecies or intratype variability (simulated samples), and percent of simulated samples that were correctly identified as positive (accuracy).

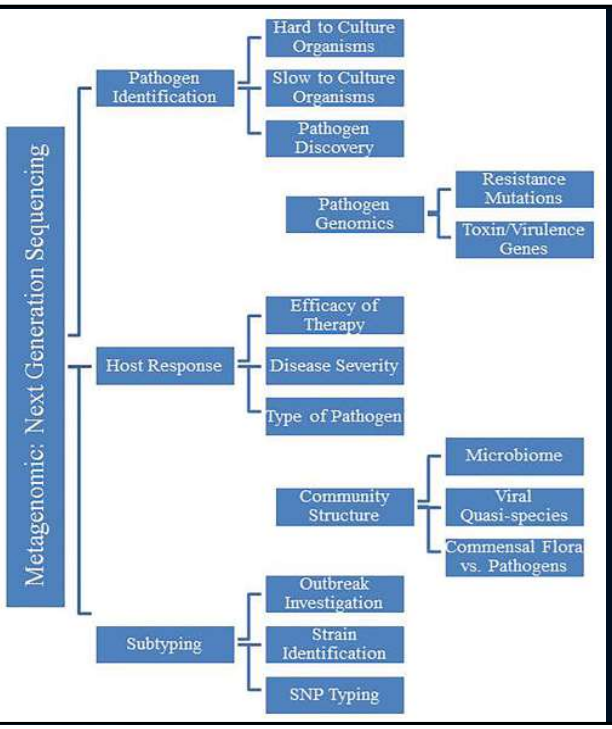
Schlaberg et al Arch Pathol Lab Med—Vol 141, June 2017



Hasman et al., J Clin Microb 2014

Transforming Clinical Microbiology with WGS

- Potential clinical applications for metagenomics sequencing



Goldberg et al., mBio 2015

Transforming Clinical Microbiology with WGS

- Assessment of the Performance Characteristics of WGS-Based Tests for Clinical Microbiology

Performance characteristic	Approach to evaluation
Accuracy	Use of specimens with known findings and confirmation of additional findings detected during validation by an orthogonal method.
Precision	Reproducibility (between-run precision): sequencing of the same samples on different runs. Repeatability (within-run precision): sequencing of the same samples in replicates within a run. Between-library precision: sequencing different library preparations of the same samples on the same sequencing run.[92]
Analytical sensitivity	Microbial variant detection: mixes of known variant strains and wild-type strains at different percentages and at low, medium, and high levels (eg, viral loads). Microbial identification: serial dilutions of samples in an appropriate matrix containing a known pathogen(s) coupled with an estimation of the minimum coverage needed to detect the pathogen.
Analytical specificity	Microbial variant detection and microbial identification: estimation of the false-positive rate at various read depths.

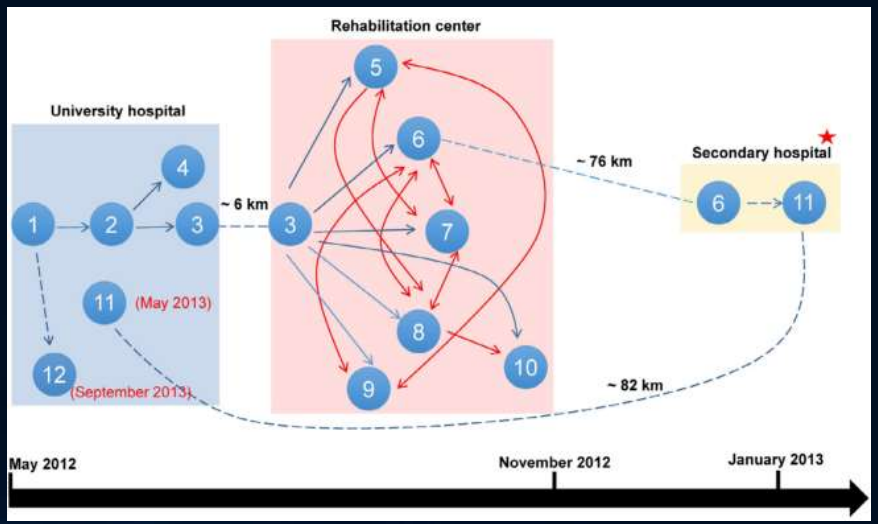
Lefterova et al., J Mol Diagn 2015

Transforming Clinical Microbiology with WGS

- Characterization and surveillance of pathogens, incl. identification in polymicrobial samples
- Correlation of genomic features with strains of clinical importance
- Generating regional phylogenomic sublineages
- Metagenomics approaches on clinical samples
- Outbreak management
- Determination of transmission of animal to human zoonoses
- Tracking of MDR organisms in hospitals and communities

Transforming Clinical Microbiology with WGS

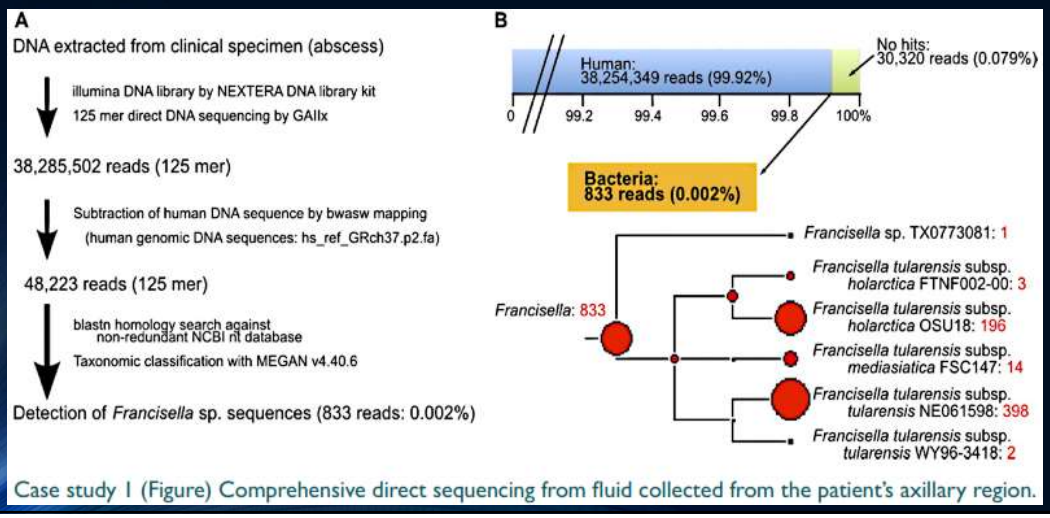
- Outbreak management



R.H. Deurenberg et al. / Journal of Biotechnology 243 (2017) 16–24

From the trenches: *pathogen identification* - *Francisella tularensis* subspecies differentiation

- Potential biosecurity issue; Preventing need for isolation

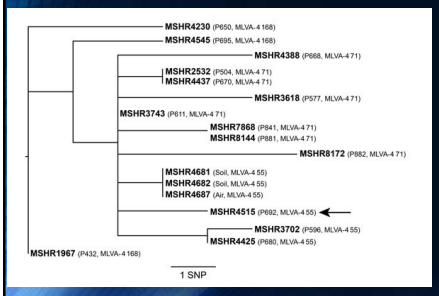


Case study 1 (Figure) Comprehensive direct sequencing from fluid collected from the patient's axillary region.

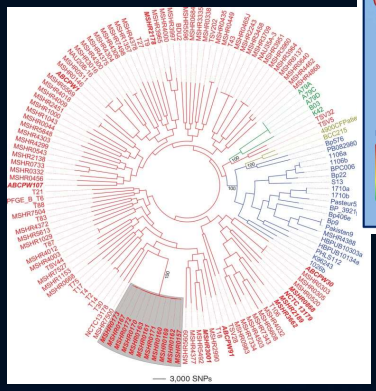
Koruda et al JCM 2012

From the trenches: *pathogen identification* – *Burholderia pseudomallei* association with environmental samples

- Potential biosecurity issue



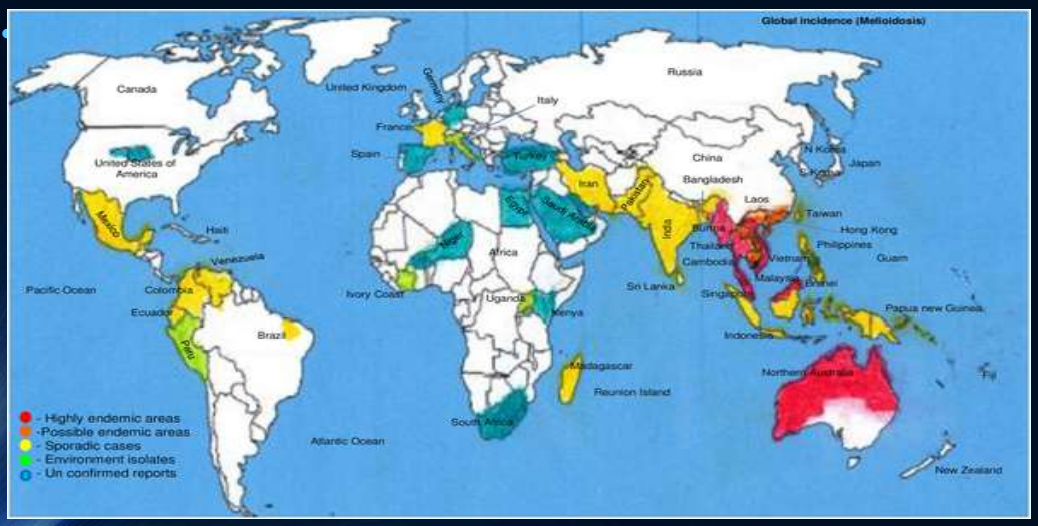
Currie et al. EID 2015



Chappelle et al. Microb Genom 2016

Melioidosis cases reported (1945–2017)
Sanchez-Villamil & Torres. Trop Med Inf Dis 2018

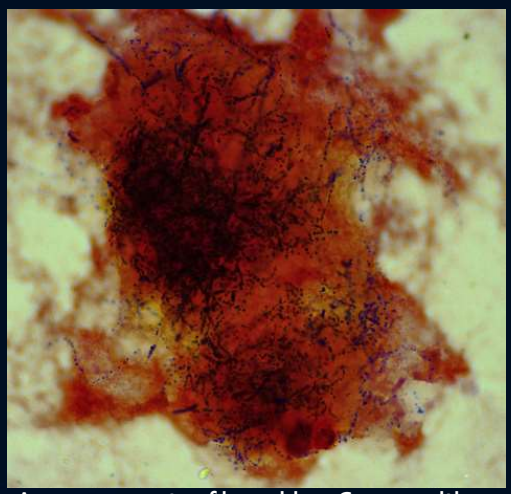
From the trenches: *pathogen identification – Melioidosis distribution in the tropics*



Perumal Samy et al. (2017) PLOS Negl Trop Dis

From the trenches: *pathogen identification - Mastoiditis caused by Fusobacterium nucleatum-Actinomyces israelii coinfection*

- Resolving discrepancies in conventional microbiology

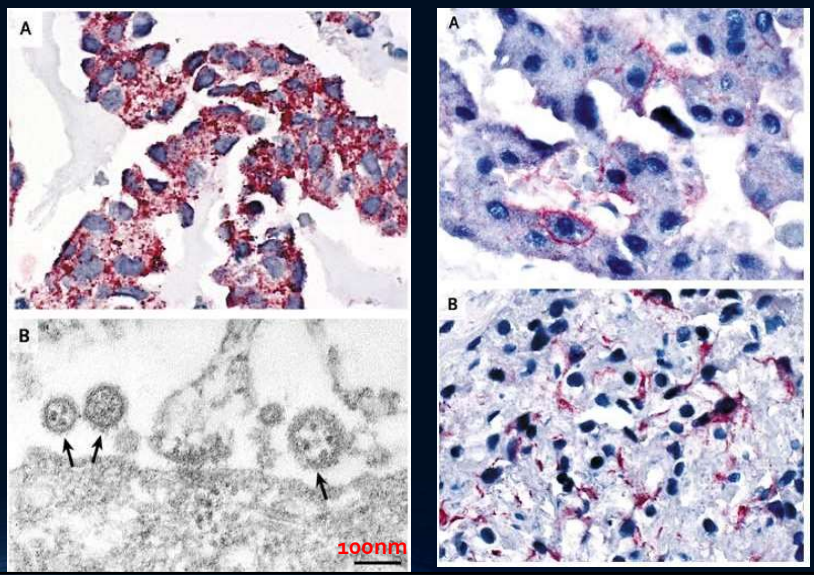


A rare aggregate of branching, Gram-positive rods is depicted, consistent with Nocardia or Actinomyces species

Salipante et al JCM 2014

From the trenches: *pathogen identification* – discovery of novel arenavirus

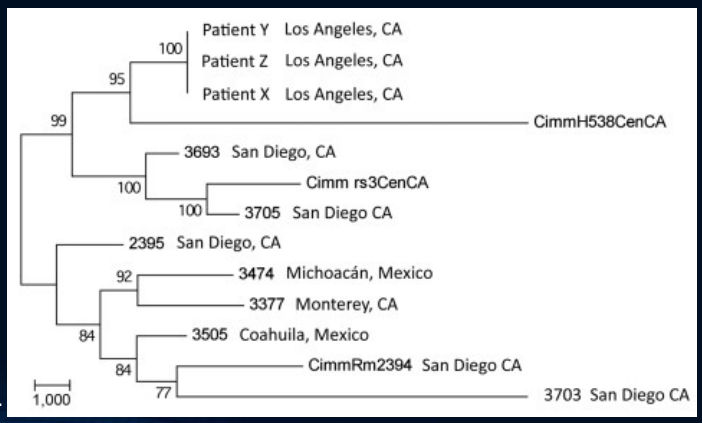
- First study for new pathogen detection and discovery



Palacios et al NEJM 2008

From the trenches: *pathogen identification*

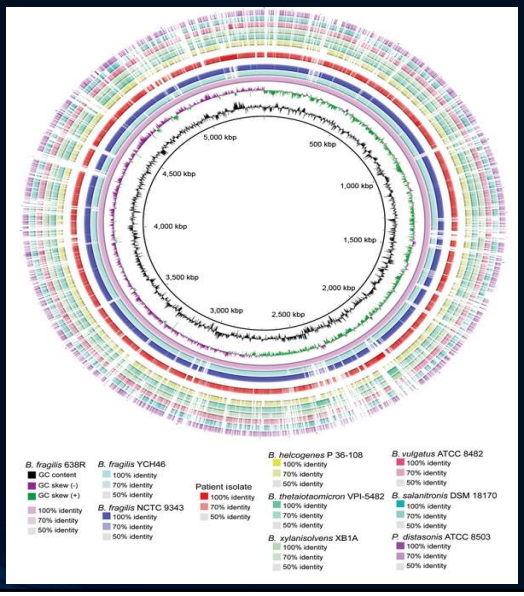
- Reveals *Coccidioides immitis* cluster in organ transplant patients



Engelthaler et al Emerg Infect Dis. 2011

From the trenches: *pathogen identification* - novel genomospecies of *Bacteroides*

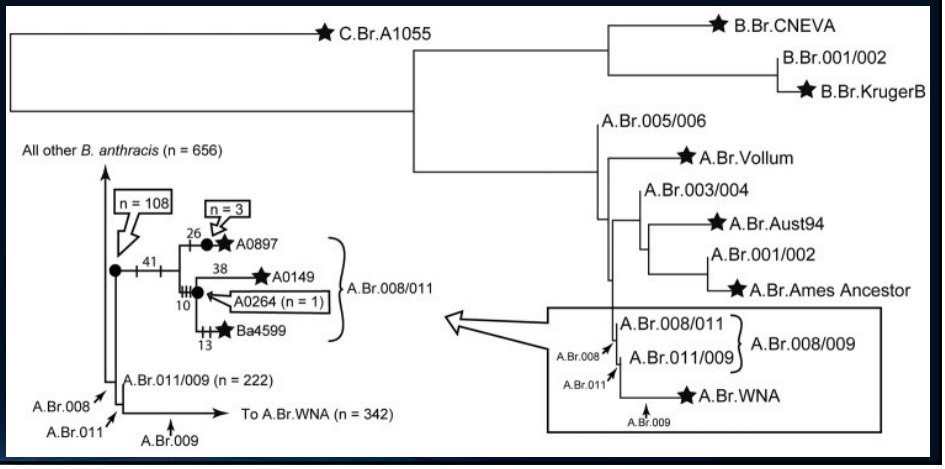
- Potential to enhance diagnostic accuracy in identifying clinical microbial isolates



Salipante et al Emerg inf Dis 2015

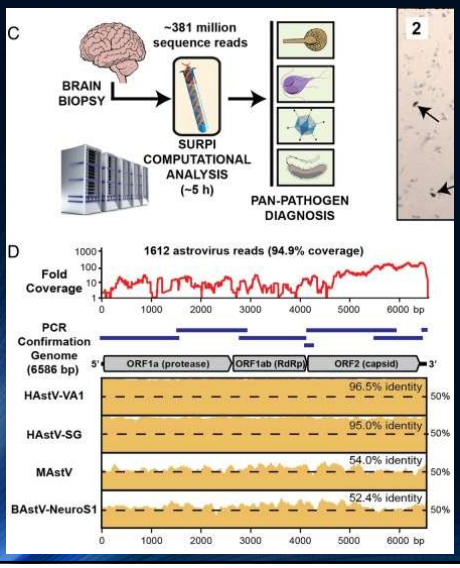
From the trenches: *pathogen identification* - Determination of anthrax outbreak among European heroin users

- Confirmation of accidental heroin contamination into supply along trafficking route

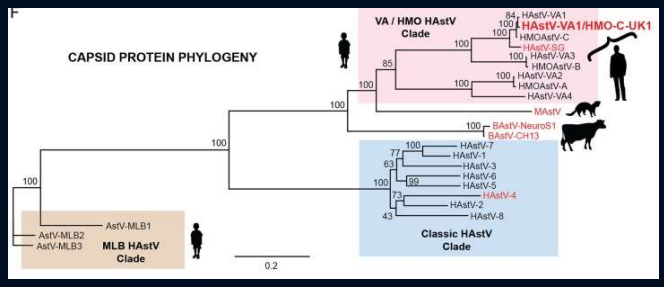


Price et al Emerg Inf Dis 2012

From the trenches: *pathogen identification* - Encephalitis by neuroinvasive astrovirus



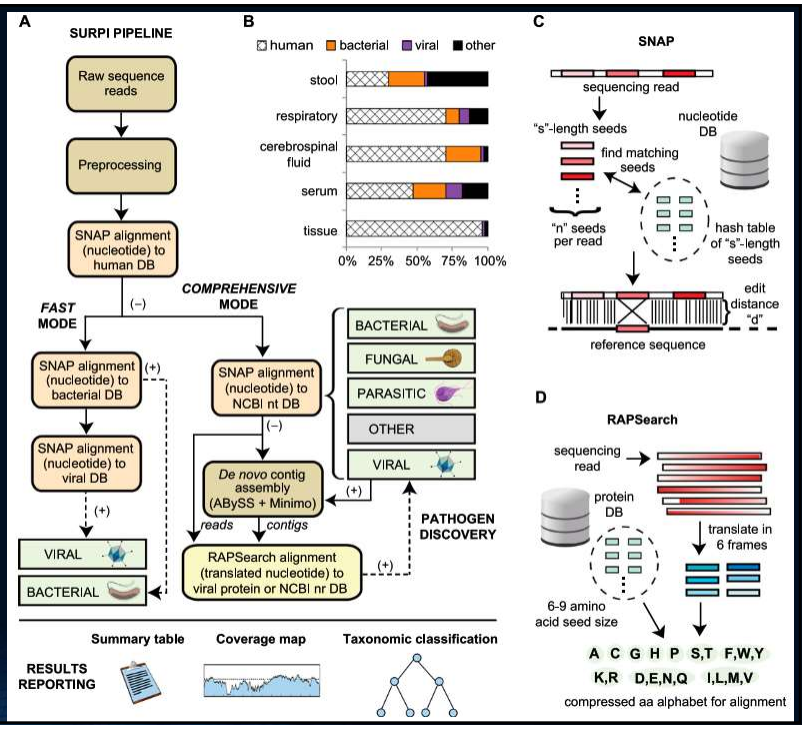
- Capability of detecting all potential pathogens simultaneously



Naccache et al. Clin Inf Dis 2012

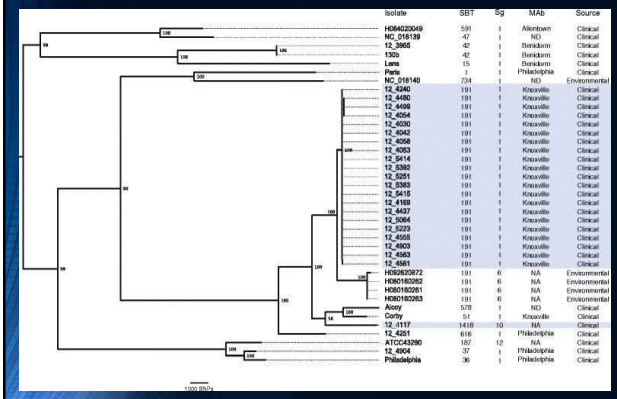
From the trenches: *pathogen identification*

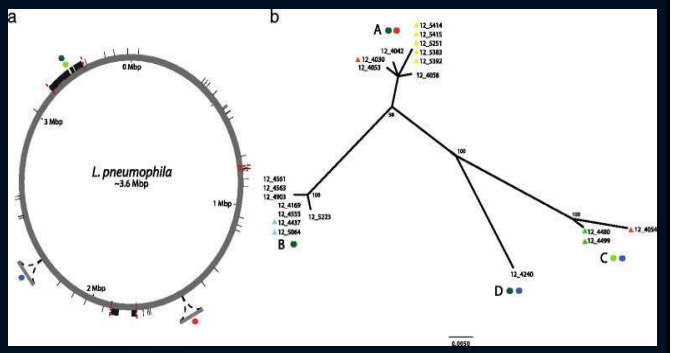
- SURPI (sequence-based ultra-rapid pathogen identification)



Naccache et al. Genome Res 2014

From the trenches: *pathogen identification* – gene flow and heterogeneity during Legionnaires disease outbreak

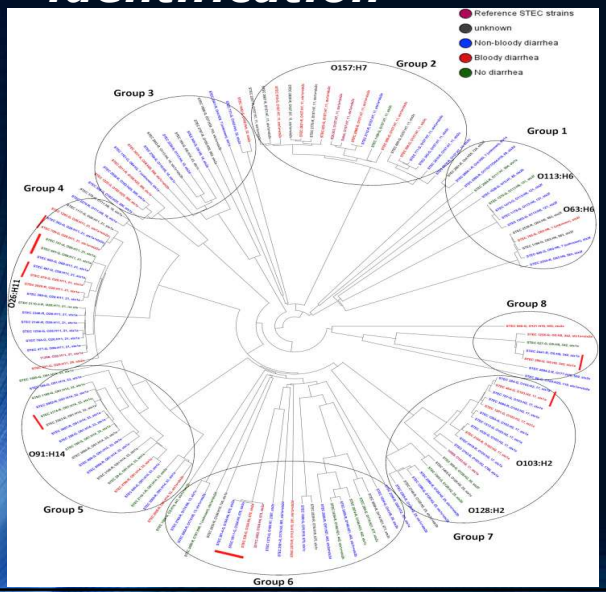


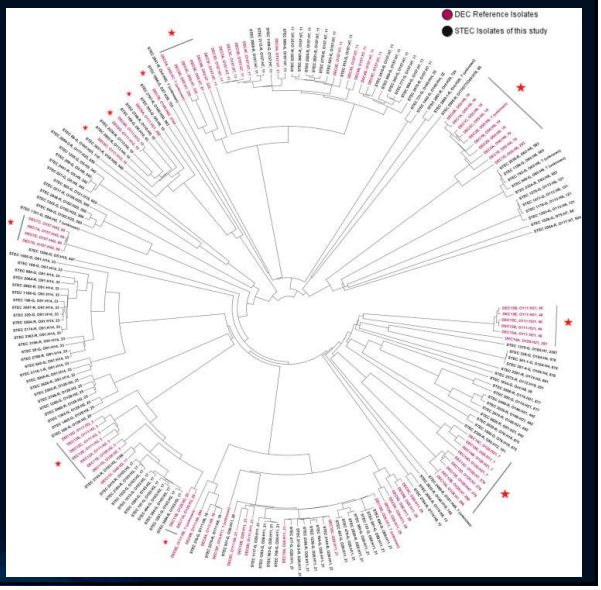


McAdam et al. BMC Genom Biol 2014. LD (ST191) outbreak in Edinburgh, UK

From the trenches: *pathogen identification*

Ferdous et al. CMI 2016; Dutch STEC isolates





From the trenches: *pathogen identification*

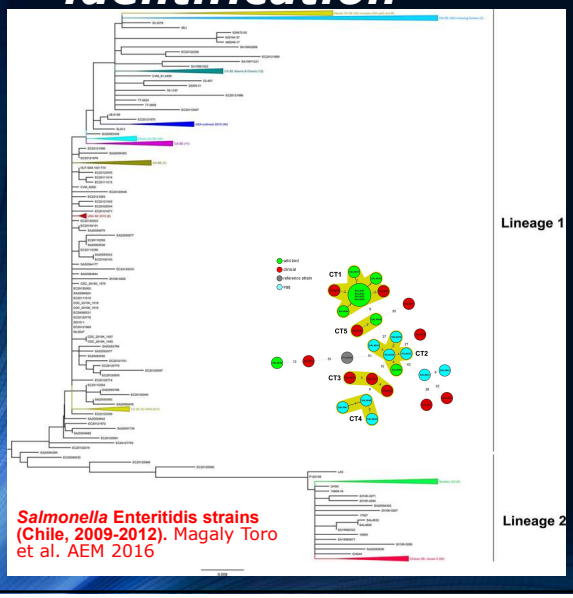
Plasmid
Volume 83, January 2016, Pages 8-11

Whole genome sequencing of diverse Shiga toxin-producing and non-producing *Escherichia coli* strains reveals a variety of virulence and novel antibiotic resistance plasmids

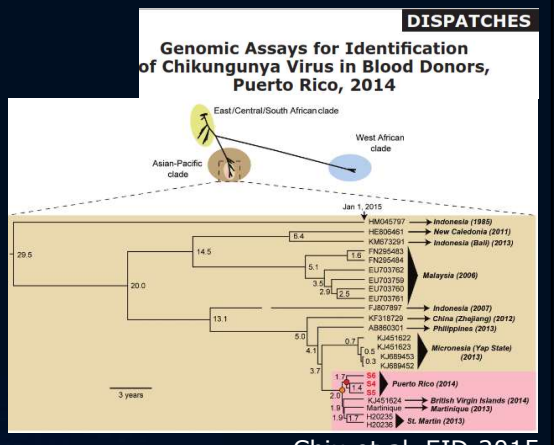
Liliana Losada [a], Chitrita DebRoy [b], Diana Radune [a], Maria Kim [a], Ravi Sanka [a], Lauren Brinkac [a], Subhashinie Kariyawasam [b], Daniel Shelton [c], Pina M. Fratamico [d], Vivek Kapur [b], Peter C.H. Feng [e]

- Plasmids sequenced from 26 Shiga toxin-producing and non-producing *E. coli* strains
- Identified 39 plasmids that carried various virulence and putative virulence genes; 8 with resistance to various antibiotics and some also had resistance genes for heavy metals.

From the trenches: *pathogen identification*



***Salmonella* Enteritidis strains (Chile, 2009-2012).** Magaly Toro et al. AEM 2016

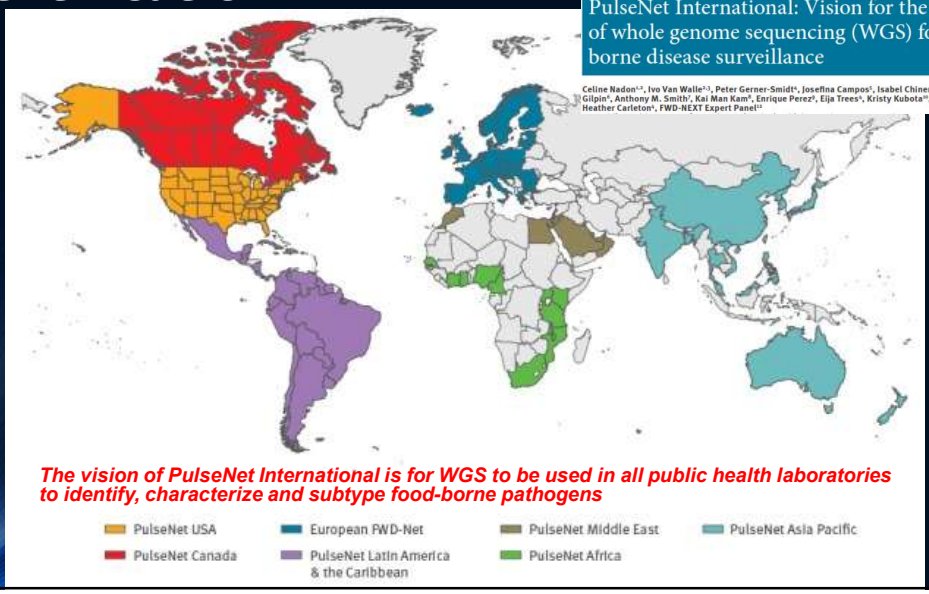


- Chiu et al. EID 2015

From the trenches: *pathogen identification*

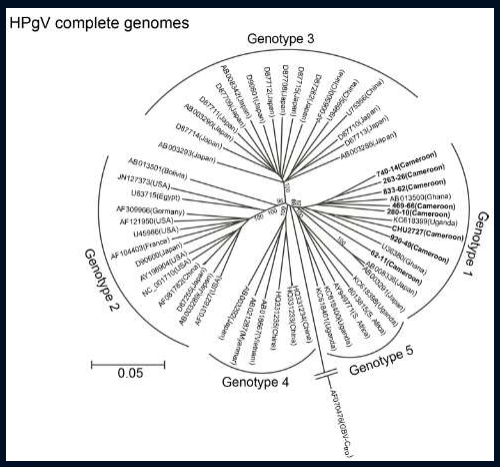
PulseNet International: Vision for the implementation of whole genome sequencing (WGS) for global food-borne disease surveillance

Celine Nadon[1,*], Ivo Van Walle[2,3], Peter Gerner-Smidt[4], Josefina Campos[5], Isabel Chinen[6], Jeniffer Concepcion-Acevedo[4], Brent Gilpin[7], Anthony M. Smith[7], Kai Man Kam[8], Enrique Perez[9], Eija Trees[4], Kristy Kubota[10], Johanna Takkinen[3], Eva Møller Nielsen[11], Heather Carleton[4], FWD-NEXT Expert Panel[12]



From the trenches: *pathogen identification*

- Pathogen diversity (HIV and Pegivirus)
  - HPgV has a worldwide distribution and is transmitted sexually, parenterally and by mother-to-child transmission, and as a consequence HIV patients are often co-infected
  - HPgV sequences were detected in 9 of 35 (26%) specimen libraries, with 8 yielding at least 99% of the viral genome.



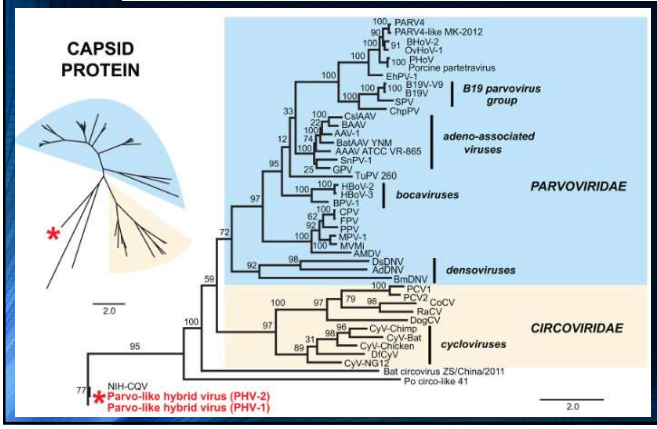
From the trenches: *pathogen identification*

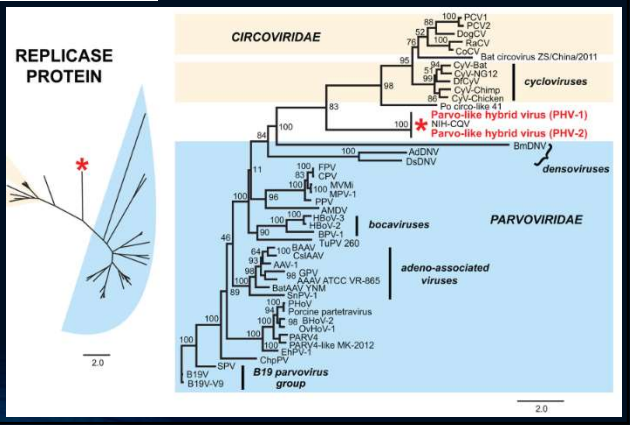
JVI
Journals.ASM.org

The Perils of Pathogen Discovery: Origin of a Novel Parvovirus-Like Hybrid Genome Traced to Nucleic Acid Extraction Spin Columns

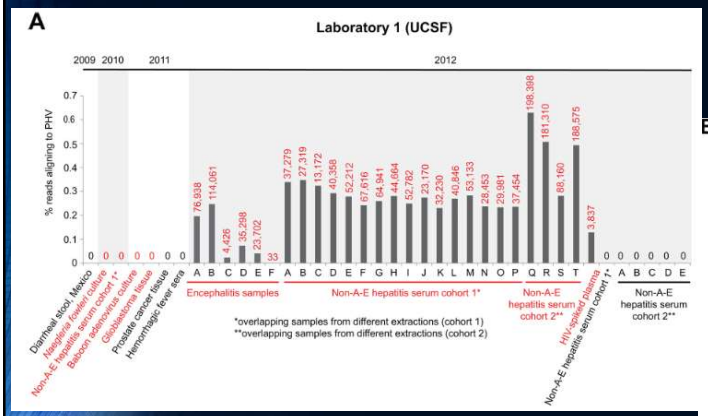
Samia N. Naccache,[a,b] Alexander L. Greninger,[a,b] Deanna Lee,[a,b] Lark L. Coffey,[c] Tung Phan,[c] Annie Rein-Weston,[a,b] Andrew Aronsohn,[d] John Hackett, Jr.,[e] Eric L. Delwart,[a,c] Charles Y. Chiu[a,b,f]

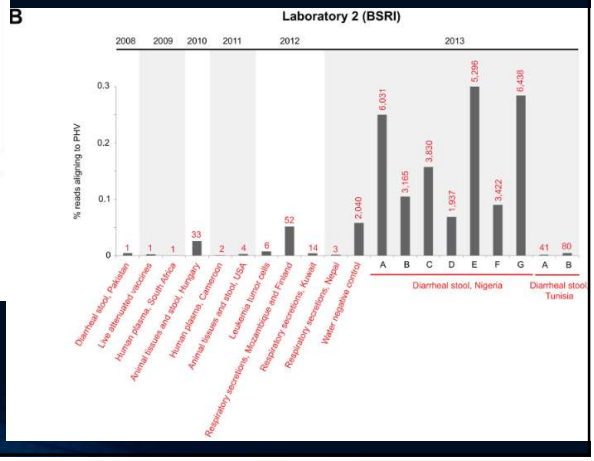
- Screening for PHV in clinical NGS data sets





From the trenches: *pathogen identification - Parvovirus*

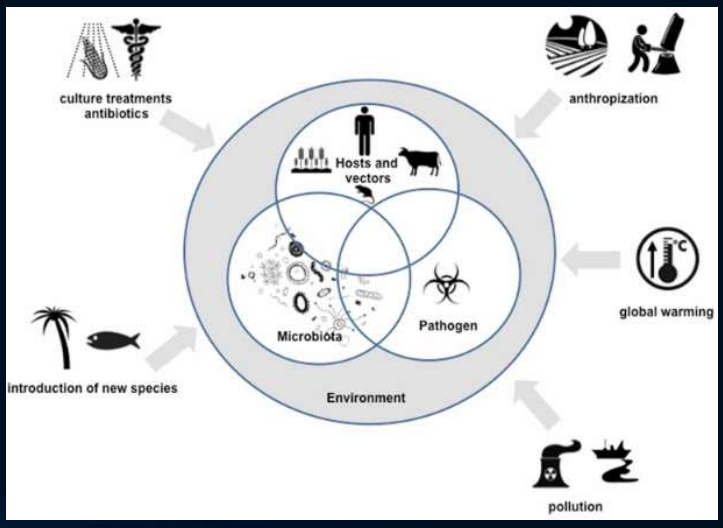




- NGS data sets processed from 2008 to 2013. Bar graphs show the percentage of PHV reads in each data set, with the number of PHV reads given above the bars.
- Sample sets extracted using Qiagen spin columns (red); those extracted using other methods (black)

From the trenches: *from pathogen to pathobiome*

- Pathobiome concept
- Objectives:
  - Accurately describe the biodiversity of relevant microbial communities
  - Shift from descriptive to functional metagenomics
  - Understand the impact of microbial communities on the persistence, transmission, and evolution of pathogens

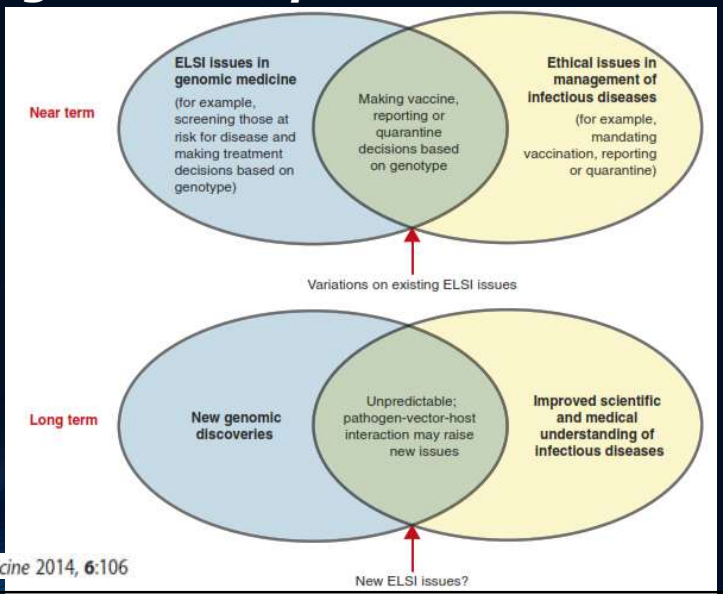


Vayssier-Taussat et al. 2014 Front Cell Infect Microbiol

From the trenches: *treatment and management implications*

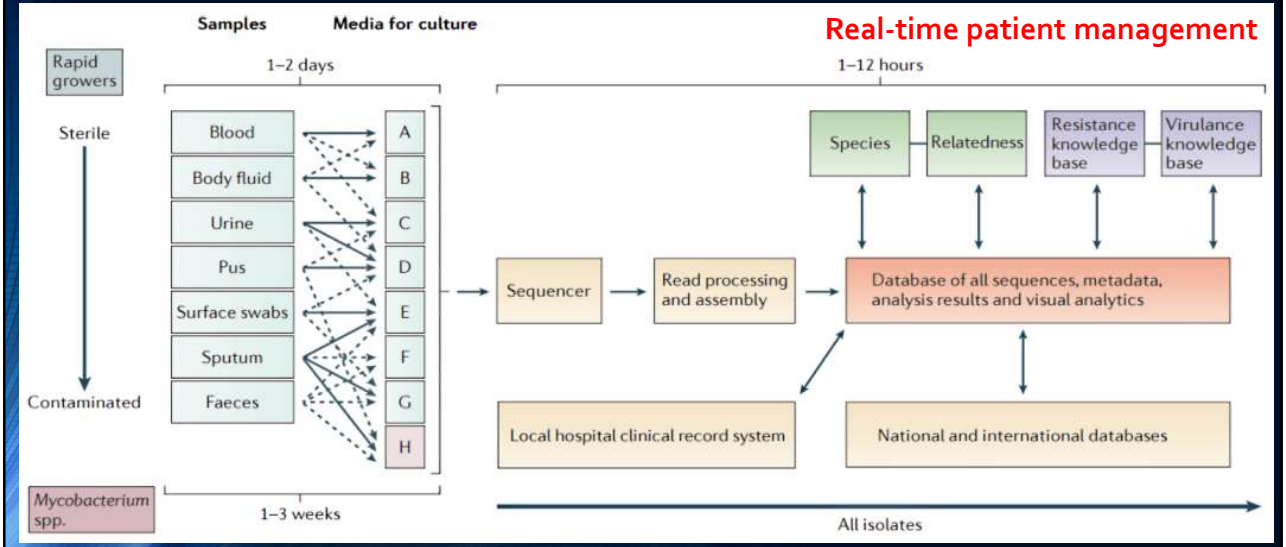
- For clinical WGS (where there is cohort studies biobanking), pertinent ELSI issues include the interpretation of data, data storage, data sharing, informed consent and identifiability/privacy
- For infectious disease WGS, issues include who (else) is at risk, and potential benefits/harms of healthcare policy that accrue to the entire population (source and contact tracing)

From the trenches: *treatment and management implications*



Geller *et al. Genome Medicine* 2014, **6**:106

From the trenches: *treatment and management implications*



AAM 2016

Is it feasible to routinely use WGS in the clinical microbiology laboratory?

- Every isolate that was recovered from culture on a single day at Houston Methodist Hospital was sequenced.
- 130 samples, including 107 aerobic cultures, 9 anaerobic cultures, and 14 acid-fast bacillus/mycology samples
- 115 isolates were correctly identified using WGS.
- 12 hours to extract DNA and prepare the sequencing libraries, 39 hours to perform the sequencing, and 2 to 4 hours for de novo assembly of contigs
- Isolates that were unable to be identified by sequencing were due to low read counts, insufficient sample preparation, or lack of a reference genome

To the future…

- Optimism for the future to deliver tangible clinical benefits to improve patient care, patient outcomes, and public health
- Could be proactive in identifying potential pathogenic outbreak species and lineages
- Pre-emptive use to monitor, detect, and control threatening agents locally, nationally, and internationally
- Will have profound implications in critical decision support systems and clinical care pathways
- Could become a cornerstone of clinical care