Whole Genome Sequencing (WGS) of Foodborne Bacteria at the FDA Field Laboratories

Okumu K’Aluoch, BVM, MSc, PHM
Microbiologist
USFDA San Francisco Laboratory

Presentation at the NCASM Spring Meeting, March 6th and 7th, 2015, Double Tree by Hilton Hotel, Pleasanton, CA
<table>
<thead>
<tr>
<th>Date</th>
<th>Advance</th>
<th>Applications</th>
</tr>
</thead>
<tbody>
<tr>
<td>1670s</td>
<td>Microscope invented by Leeuwenhoek</td>
<td>Visualize bacteria, protozoa</td>
</tr>
<tr>
<td>1850s</td>
<td>Puerperal fever identified as infectious and interventions implemented by Semmelweis [23]</td>
<td>Hospital infection control motivated by growing understanding of microbial etiology</td>
</tr>
<tr>
<td>1864</td>
<td>Cholera transmission by water proven by Snow</td>
<td>Risk factor (mode of transmission) and prevention measure for specific infectious syndrome</td>
</tr>
<tr>
<td>1890s</td>
<td>Proof of parasitic origin (Grass) and mosquito transmission (Ross) of malaria</td>
<td>Vector control</td>
</tr>
<tr>
<td>1890s</td>
<td>Identification of microbial etiologies for tuberculosis, anthrax, and so on; Koch's postulates</td>
<td>Targeted diagnostics, therapeutics, and move from syndromic diagnosis to pathogen identification</td>
</tr>
<tr>
<td>1900-1930s</td>
<td>Discovery of filterable animal viruses [24]</td>
<td>Influenza etiology settled (previously thought bacterial) [25]</td>
</tr>
<tr>
<td>1910s-1950s</td>
<td>Phenotypic subspecies taxonomy: serotyping [26,27], phage typing [28]</td>
<td>Association of particular types with prognosis [27,29], drug resistance</td>
</tr>
<tr>
<td>1944</td>
<td>Discovery of DNA as the genetic material [30]</td>
<td>Basis for genotyping tools for molecular epidemiology</td>
</tr>
<tr>
<td>1970</td>
<td>Restriction enzymes [31]</td>
<td>Basis for restriction fragment length polymorphism approaches, including pulsed field gel electrophoresis</td>
</tr>
<tr>
<td>1975-1985</td>
<td>Sanger DNA sequencing [32], PCR [33]</td>
<td>Basis for variable number tandem repeat (VNTR) and multilocus sequence typing (MLST) approaches to characterize microbes and their genetic relatedness</td>
</tr>
<tr>
<td>2000s-now</td>
<td>High-throughput rapid sequencing technologies</td>
<td>Microbial genome sequencing</td>
</tr>
</tbody>
</table>
Objectives

Establish a public database of Whole Genome Sequence for infectious organisms - Genome Trakr Program

Use the sequence data for epidemiological outbreak investigations

Determine the application of WGS for confirmation of bacterial isolates
Program Organization

• Collaboration with CFSAN-CVM and state public health labs for the GenomeTrackr program encompasses nine ORA field labs and our MOD-1 facility which is manned by CFSAN scientists:

Lab:  ARL, DEN, NRL, PRL-SW, PRL-NW, SAN, SRL, FCC, WEAC    MOD-1

• Sequence personnel are field laboratory microbiologists and ORA researchers.

• Sequence Salmonella, Listeria monocytogenes

• Labs are equipped with Illumina MiSeq genome sequencers, QIACube DNA purification systems, QUBIT fluorometers.
Establishment of WGS

MiSeq installation

Training - Industry

Proficiency Testing (PT) - Salmonella

Salmonella sequencing

Training FDA

Listeria PT

Routine sequencing
PRIORITIZATION OF SEQUENCING

Isolates related to outbreak samples or consumer complaint samples, both Salmonella and L. mono, should be sequenced as soon after receipt as is possible.

Regulatory sample isolates, not related to outbreaks or environmental swabs, sequencing should be initiated within one week of receipt of the isolate.

Environmental isolates should be sequenced after the PFGE results are available.

Archived isolates should be used to fill any unused flow cell capacity when running isolates from #1-3 above.
Laboratory Process

- Enrichment
- Screening and / or Selective enrichment
- Selective culture
- Confirmatory tests
Laboratory Process

Bacterial culture - DNA extraction
- Automated QIAcube
Laboratory Process

DNA Quantification

Qubit Fluorometer

- dsDNA, ssDNA, RNA, proteins
Nextera XT Library Prep Workflow

1. Sample Input
2. Tagmentation
3. PCR
4. AMPure
5. Normalization
Tagmentation

- **Protocol:**
  - This is a 5 min. incubation at 55°C
  - XT DNA: 5μL DNA (at 0.2ng/μL) + 10μL Buffer + 5μL transposome = 20μL total

---

**Sample Input** → **Tagmentation** → **PCR** → **AMPure** → **Normalization**
PCR Amplification

- **Protocol:**
  - This step is a 2 index primers primer and 12 cycle PCR reaction.
  - CRITICAL STEP: Two primers from the Index kit add the index and the P5 or P7 that attaches the library to the flow cell.
Illumina MiSeq sample loading
CURRENT DATA FLOW

Field lab isolates → Illumina MiSeq

Pending: direct network connection & DROBO data storage capacity

20TB external Drive

NCBI National Center for Biotechnology Information

CFSAN server
Basic Data Flow for Global WGS Public Access Databases

**DATA ACQUISITION**
Sequence and upload genomic and geographic data

**DATA ASSEMBLY, ANALYSIS, AND STORAGE**
- International Nucleotide Sequence Database Collaboration (INSDC)
- Shared Public Access Databases
  - NCBI – National Center for Biotechnology Information
  - EMBL – European Molecular Biology Laboratory
  - DDBJ – DNA Databank of Japan

**PUBLIC HEALTH APPLICATION AND INTERPRETATION OF DATA**
- Find clinical links
- Identify clusters
- Conduct traceback
- Develop rapid methods
- Develop culture independent tests
- Develop new analytical software

11/2014 | State, Local, Federal, and Foreign Public Health Agencies | Academia/Industry
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.

Mouse over the salad to the right to see the ingredient origins.
Epidemiology concerns

» is an outbreak present?

» how, where and when did the infectious pathogen enter the population?

» who is the target?

» what is the transmission rate?

» how is the pathogen transmitted through the population?

» what are the characteristics associated with the pathogen?
Why Database?

Genotypic characteristics data of a microbial pathogen can be used to better define the specific population responsible for an outbreak to facilitate public health and/or clinical intervention.

The complexity of tracing foodborne pathogens

- >160,000 domestic and >250,000 foreign registered food facilities in the USA
- >300 ports of entry and >130,000 importers for more than 10M import lines per year
- > 2 m farms in the USA
Food Safety News

Breaking news for everyone’s consumption

Whole-Genome Sequencing Expected to Revolutionize Outbreak Investigations

By James Andrews | July 30, 2014

For decades, food safety experts have lived with the reality that roughly one-third of foodborne illness outbreak investigations ended without finding the source.

But with the wide proliferation of new laboratory technology on the horizon, outbreak investigations could soon become more accurate, more efficient and more complete, according to researchers at Cornell University, the New York State Department of Health and the U.S. Food and Drug Administration.
Epidemiologic data and pathogen genome sequences: a powerful synergy for public health
Yonatan H Grad and Marc Lipsitch; Genome Biology 2014, 15:538 (review)

……genome sequence augment epidemiological inquiry to generate inferences about spread and evolution of pathogens thus help guide effort to reduce disease burden.

Rapid Whole-Genome Sequencing for Surveillance of Salmonella enterica Serovar Enteritidis

In a retrospective and prospective analyses, WGS identified additional isolates that could be attributed to the outbreak, but which differed from the outbreak-associated PFGE type. Additional putative outbreak clusters were also detected. The study demonstrates the practicality of implementing WGS for outbreak surveillance in a state public health laboratory.
Evaluation of Whole Genome Sequencing for Outbreak Detection of Salmonella enterica

Pan-genome tree, K-mer tree, Nucleotide differences tree and SNP tree phylogenetic analyses compared to PFGE. The findings suggests that WGS and data analysis using SNP and/or nucleotide difference approaches are superior methodologies for epidemiological typing of S. Typhimurium isolates and might be very successfully applied for outbreak detection in combination with epidemiological data.
WGS Applications

Whole-genome sequencing to control antimicrobial resistance

WGS has become an invaluable tool to combat antibiotic resistance. Has the ability to differentiate exogenous re-infection and relapsing primary infection for drug efficacy trials. Used in diagnostic test for drug susceptibility tests and can provide superior information compared to current methods by simplifying workflows.

Whole genome Multi-locus sequence typing (wg MLST): Real-time identification, serotyping, virulence and antimicrobial resistance profiling in one workflow. Carleton H. In Current methods and tools for analysis of foodborne pathogen genomes- workshop for Genome trakr participants, Dec. 8th, 2014
Real-Time Whole-Genome Sequencing for Routine Typing, Surveillance, and Outbreak Detection of Verotoxigenic \textit{Escherichia coli}


The study shows that WGS-based typing and surveillance using VirulenceFinder were able to detect verocytotoxin (vtx1 & 2), intimin (eae) and other additional virulence genes faster and cheaper. WGS delivers typing results that equal or even surpass the current typing methodologies in terms of microbiological information.
Detecting Emerging and Novel Antimicrobial Resistance Genes Using Whole Genome Sequencing

Shaohua Zhao DVM, MPVM, PhD

U.S. Food and Drug Administration
Center for Veterinary Medicine
Office of Research
Laurel, MD

Disclaimer
This communication is consistent with 21 CFR 10.85 (k) and constitutes an informal communication that represents my best judgment at this time but does not constitute an advisory opinion, does not necessarily represent the formal position of FDA, and does not bind or otherwise obligate or commit the agency to the views expressed.
Application of Next Generation Sequencing in NARMS Program

• Antibiotic resistance profiles - to a panel of 15 antimicrobials followed by a secondary panel of nine β-lactam antimicrobials

• Seven phenotypically positive ESBL E. coli isolates
  – Cattle (n=5) and chicken breast (n=2)

• Whole genome sequencing
• All ESBL producing *E. coli* isolates were multidrug resistant (resistance to ≥3 antimicrobial classes).

• ESBL producing *E. coli* isolates have carried CTX-M genes either on the plasmid or chromosome.

• This is the first report of CTX-M in *E. coli* isolates from NARMS retail meat program in the US.

• The finding of CTX-M producing *E. coli* in food animals and retail chicken breast coupled with the carriage of these genes on plasmids that are highly efficient at conjugal transfer are worrisome and may suggest the potential transmission of these genes to humans through the food chain.
Summary

- Based on current knowledge and technology, WGS predicts resistance very well
- 98-100% correlation for the drug classes beta-lactam, tetracycline, chloramphenicol, sulfonamide, trimethoprim/sulfamethoxazole, macrolides and quinolone
- 92-97% correlation for aminoglycoside, lincosamides and keolides
- A comprehensive and accurate database of ARG is critical
- Reasons for disconnect
  - AST interpretation standard
  - experimental and analytical error
  - variable gene expression level
  - unknown mechanisms
Benefits of a WGS Strategy in NARMS

WGS has potential to serve as a single assay of NARMS surveillance and supplant multiple methods

1. Classical serotyping
2. PFGE and other molecular typing methods
3. In vitro antimicrobial susceptibility testing
4. Multiple PCR assays to detect resistance genes and plasmid typing

And to provide:
1. Genome surveillance
2. Virulence profiles
3. Markers for source attribution
4. Better understanding of emerging resistance trends, origin, dissemination and selection pressure
5. Cost saving
Using Whole Genome Sequencing to Predict *E. coli* Antibiotic Resistance

Gregory Tyson, Ph.D.
Center for Veterinary Medicine
Office of Research
DAFM/NARMS

Disclaimer
This communication is consistent with 21 CFR 10.85 (k) and constitutes an informal communication that represents my best judgment at this time but does not constitute an advisory opinion, does not necessarily represent the formal position of FDA, and does not bind or otherwise obligate or commit the agency to the views expressed.
Phenotype/Genotype Correlation

- Aminoglycosides
- Cephems
- Folate synthesis inhibitors
- Macrolides
- Penicillins
- Phenicols
- Quinolones
- Tetracyclines

<table>
<thead>
<tr>
<th>Resistance prevalence (%)</th>
<th>GEN</th>
<th>KAN</th>
<th>STR</th>
<th>AMC</th>
<th>AXO</th>
<th>FOX</th>
<th>TIO</th>
<th>COT</th>
<th>FIS</th>
<th>AZI</th>
<th>AMP</th>
<th>CHL</th>
<th>CIP</th>
<th>NAL</th>
<th>TET</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>30</td>
<td></td>
<td>20</td>
<td></td>
<td>10</td>
<td></td>
<td>20</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Aminoglycosides</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>β-lactam/β-lactam inhibitors</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cephems</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>30</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Folate synthesis inhibitors</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>30</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Macrolides</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>30</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Penicillins</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>30</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Phenicols</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>30</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Quinolones</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>30</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tetracyclines</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>30</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Correlation of Genotype/Phenotype

- 99.6% sensitivity, 97.8% specificity of genotype-phenotype correlation
- Only 17 total discrepancies among 1140 phenotypic tests
  - 12 due to streptomycin
Weaknesses to WGS

• Can only identify known resistance genes/mutations
  – Novel genes or variants may not be detected if low homology to known ones
  – Does not work for new antibiotic classes
• Need highly curated, updated resistance gene database
• Expertise needed to analyze data
  – Automation making it easier
• Fragmented genomes
  – Complicates identification of resistance elements
  – Assembly methods may improve, raw data always available
Conclusions

• *E. coli* antibiotic resistance is a major problem
  – Only subset of cattle isolates have multidrug-resistance
  – Transfer of resistance concerning (e.g. with resistance elements)
• WGS can cheaply, quickly identify *E. coli* resistance genotypes
  – Superior to PCR tests as indicator of genotypic resistance
  – Need to be careful with truncated genes
• High correlation of genotype with phenotype (approximately 99%)
  – May provide reasonable alternative to phenotypic testing
• Reduction in streptomycin resistance cutoff for *Salmonella, E. coli* may be advisable
  – Provides better correlation with genotypic indicators of resistance
Nutshell

WGS is still work in progress that is being fine-tuned and standardized, however, to achieve its putative potential in:-

- discerning bacterial isolates to single nucleotide level,
- detecting evolutionary changes in infection within a population,
- detecting resistance genes, virulence genes, mutation, etc,
- tracing origins and modes of transmission,

A well annotated fully functional database is indispensable.
Acknowledgements

- Shaohua Zhao DVM, MPVM, PhD and Gregory Tyson PhD of CVM
- Marc Allard PhD, Ruth Timme PhD, and Peter Evans PhD, of CFSAN
- Keith Iams PhD of ORA
- Elaine Yeh of ORA SFL Microbiologist
- David Lau and Teresa Lee Supervisory Microbiologist at SFL
- Tom Sidebottom, ORA SFL Director
- SFL Microbiologist for isolating pure bacterial cultures used in WGS