Public Health Food Safety Applications for Whole Genome Sequencing (WGS)

Peter Gerner-Smidt, MD ScD
Chief, Enteric Diseases Laboratory Branch

WGS Works!
The Experience In The United States

Shared Mission—Different Roles

Non-regulatory

- Disease surveillance
- Outbreak detection and investigation
- Analyzing burden, trends, and effectiveness of prevention efforts
- Attribution to sources
- Education and training
- Information for policy

Regulatory

- FDA and USDA
  - Inspection
  - Enforcement
  - Investigating farm and production facilities
  - Product recall
  - Product traceback
  - Risk assessment and management
Surveillance By WGS In The U.S.

- **Integrated Surveillance**
  - Began September 2013 with Listeria
- Since 2014 almost all multi-state outbreaks by any foodborne pathogen are investigated by WGS

### WGS For Outbreak Investigations

2014 U.S. Caramel Apples *Listeria* Outbreak

- Allele differences at node: median [min–max]
  - (>5,800 loci analyzed by BioNumerics software)

#### Listeria Cluster Metrics Before and After WGS

<table>
<thead>
<tr>
<th></th>
<th>Pre-WGS (Sep 2012- Aug 2013)</th>
<th>WGS Year 1 (Sep 2013- Aug 2014)</th>
<th>WGS Year 2 (Sep 2014- Aug 2015)</th>
</tr>
</thead>
<tbody>
<tr>
<td>No. of clusters detected</td>
<td>14</td>
<td>19</td>
<td>21</td>
</tr>
<tr>
<td>No. of clusters detected sooner or only by WGS</td>
<td>6</td>
<td>6</td>
<td>6</td>
</tr>
<tr>
<td>No. of outbreaks solved (food source identified)</td>
<td>98</td>
<td>2</td>
<td>4</td>
</tr>
<tr>
<td>Median no. of cases per cluster</td>
<td>5</td>
<td>6</td>
<td>3</td>
</tr>
</tbody>
</table>

No. cases linked to food source

13

20

98

Data are preliminary and subject to change.

How WGS Influenced Outbreak Investigations

- **Improved case definitions in outbreaks**
  - Apparent PFGE clusters are not single-source outbreaks or are pseudo-clusters
  - Isolates with same PFGE patterns may be unrelated
  - Isolates with different PFGE patterns may be related
- Increase confidence in the link between human and product isolates
- Link historical cases to a current outbreak investigation
- Characterize the ecology of long-term pathogen reservoirs in the food chain
WGS In Public Health In A Global Perspective

A foodborne infection on one continent may have its source on a different continent.

International outbreaks are common.

Subtyping methods and analyses must be standardized at the global level to ensure global comparability of data.

- Outbreak investigations fast

WGS in Public Health:

The analytical tools must be

• Simple
  - Public health microbiologists are NOT bioinformaticians
  - Standard desktop software

• Comprehensive
  - All characterization incl. analysis in one workflow

• Working in a network of laboratories - STANDARDIZED
  - Everybody analyze data the same way
  - Free sharing and comparison of data between labs
    - Robust QA/QC

“Transforming Public Health Microbiology – PulseNet and Beyond”

- Replacing traditional microbiology with WGS
  - For Food Safety Reference Labs:
    - Cost-efficient consolidation of multiple workflows: Identification – serotyping – virulence profiling – antimicrobial resistance characterization – subtyping

Foodborne Pathogens Know No Borders

http://www.kidsmaps.com/
Partners In System Development

**International Partners:**
- PulseNet International
- eCDC
- EFSA
- Statens Serum Institut
- Public Health England
- Institut Pasteur
- DTU
- Academia
- GMI
- PHAC

**U.S. Partners:**
- PulseNet
- OutbreakNet
- FDA/CFSAN-CVM
- Genome Trakr
- Academia
- USDA/FSIS-ARS
- NIH

1. We neither have the capacity nor the knowledge to make the WGS transformation alone
2. What we do must be in sync with what others do to ensure national and international comparability of data

WGS for Reference Characterization

- Genus/species ID
- Serotype
- Virulence profile
- Antimicrobial resistance
- Plasmid profile
- .... more

WGS implementation in food safety

Foodborne Disease Branches
- Other
- CDC and State programs

‘One Shot’ Characterization Of STEC by WGS

<table>
<thead>
<tr>
<th>Phenotype</th>
<th>Resistant</th>
<th>Not Resistant</th>
</tr>
</thead>
<tbody>
<tr>
<td>Genotype</td>
<td>1129</td>
<td>1166</td>
</tr>
<tr>
<td>1142</td>
<td>1867</td>
<td>3009</td>
</tr>
</tbody>
</table>

Resistance Prediction By WGS Is Very Accurate

ResFinder ~ Salmonella

Measure | Value (%) | 95% CI
---|---|---
Sensitivity | 98.9 | 98.1-99.4
Specificity | 98.0 | 97.3-98.6
Kappa coefficient | 0.97 (very good) | 0.96-0.98

Explanation of Virulence and Resistance Markers:
- *Salmonella* - This test has not been cleared or approved by the FDA. The performance characteristics have not been fully established. The results of this test should not be used for the diagnosis, treatment, or assessment of patient health or management.

NARMS, 2015 - Data are preliminary and subject to change
Two WGS Analytical Approaches

- **Nucleotide level comparison: hqSNP**
- **Gene- gene comparison: cg/wgMLST**

<table>
<thead>
<tr>
<th></th>
<th>hqSNP</th>
<th>cg/wgMLST</th>
</tr>
</thead>
<tbody>
<tr>
<td>Epidemiological concordance</td>
<td>High</td>
<td>High</td>
</tr>
<tr>
<td>Stable nomenclature</td>
<td>(No)</td>
<td>Yes</td>
</tr>
<tr>
<td>Speed</td>
<td>Slow SNP calling, slow analysis</td>
<td>Slow allele calling, fast analysis</td>
</tr>
<tr>
<td>Local computing requirements</td>
<td>Medium-High</td>
<td>Low</td>
</tr>
<tr>
<td>Local bioinformatics expertise</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>Reference used to perform analysis</td>
<td>Sequence of closely related annotated strain</td>
<td>Allele database</td>
</tr>
<tr>
<td>Internationally standardization</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>Cost of analysis</td>
<td>(Expensive)</td>
<td>Cheap</td>
</tr>
</tbody>
</table>

The preferred public health standard is **cg/wg MLST**
- May be globally standardized
  - Direct comparison of data generated in different labs
- Do not require bioinformatics expertise
- Is NOT computer intensive
- BUT few MLST allele databases are currently available in public domain

**hqSNP** is best suited for analysis when specific information that is not provided by cg/wgMLST is needed
- Difficult to standardize
  - Comparison of data generated in different labs will require analysis in each lab after sharing raw sequence data
- Bioinformatics expertise and high capacity computing required
- BUT easy to set up with resources available on the internet

The Gene-Gene (MLST) approach is scalable
PFGE indistinguishable isolates from an outbreak and unrelated to the outbreak (2013)
**WGS Data Flow**

- **LIMS**
- **Sequencer**
- **End users in national/local laboratories**

**Public Health databases**
- Extensive Metadata
- Allele names, Allele code (strain names)

**Databases**
- NO Metadata

**Temporary storage, QA/QC, Data extraction**
- Trimming, mapping, de novo assembly, SNP detection, allele detection
- NO Metadata

**Public Domain**

**How Close Is Close?**

**2014 U.S. Caramel Apples Listeria Outbreak**

- **PFGE**
- Unrelated isolates (hot dog and patient)
- Highly related patient isolate, different PFGE pattern
- Not closely related (minimum 1,628 allele differences)

Cluster 1 (≤6 allele differences)

Cluster 2 (≤10 allele differences)
- 2014L-6577 2014L-6656 2014L-6681

**2012-2013 Salmonella serovar Heidelberg multistate outbreak associated with chicken from producer X**

- **PFGE patterns**
  - Pattern 122
  - Pattern 672
  - Pattern 45
  - Pattern 22
  - Pattern 326
  - Pattern 41
  - Pattern 258

- **Chicken from producer X**
- **Clinical isolate from 2011**

**Data are preliminary and subject to change.**
**Salmonella serovar Heidelberg**

PFGE pattern JF6X01.0022 single state outbreaks in summer 2013

- 2013T-3050-
- 2013T-3042-
- 2013T-3111-
- 2013T-3112-
- 2013T-3040-
- 2013T-3059-
- 2013T-3041-
- 2013K-1037-
- 2013AM-0488
- 2013AM-0486
- 2013AM-0487
- 2013K-1036-
- 2013K-1040-
- 2013K-1039-
- 2013K-1038

Salmonella serovar Heidelberg

Same PFGE Pattern – Multiple Outbreaks With Different WGS Profiles

105-113 SNPs

2-8 SNPs

2-16 SNPs

4-5 SNPs

3-6 SNPs

State A funeral

State B daycare

State C Church Supper

Unrelated Sporadic case

Unrelated Sporadic case

Unrelated Sporadic case

**How Close Is Close?**

- Depends on the question you want answered
  - Outbreak? Attribution? Production ecology? Global Epidemiology?
- Depends on the pathogen
- Depends on the epidemiological context

**WGS Is Here To Stay But Issues Needs to Be Addressed**

- Data sharing
  - With everyone? With public health partners? With industry?
- Politics & Ethics
  - Trade barriers? Liability? Intellectual property? vs Common good
- How do we implement WGS in lesser resourced countries?
  - Some have no basic surveillance capacity

A WGS match between a food isolate and a clinical isolate does NOT mean that the food caused the patient’s illness

*Listeria Cluster*

They likely share an ancestor somewhere in the food production chain

Epidemiological and traceback information remains critical
Acknowledgements

Center for Genomic Epidemiology, DTU
University of Oxford, M. Maiden, K. Jolly
Public Health England, T. Dallman

Disclaimers:

"The findings and conclusions in this presentation are those of the author and do not necessarily represent the official position of the Centers for Disease Control and Prevention"

"Use of trade names is for identification only and does not imply endorsement by the Centers for Disease Control and Prevention or by the U.S. Department of Health and Human Services."