How Whole-Genome Sequencing Impacts Outbreak Investigations – A Public Health Perspective

Anna Carlson, PhD

Nebraska Department of Health and Human Services
Foodborne Disease Epidemiology Surveillance Coordinator

BIFSCo March 2016
Overview

- **Background**
  - Surveillance and subtyping for *Listeria*, *Salmonella* and *E. coli*
  - *Listeria* Initiative, *Listeria* CDC WGS pilot project

- **Utility of whole genome sequencing (WGS) in Public Health**
  - What PFGE & WGS are doing for public health surveillance now

- **Real world examples of use of WGS in outbreak detection**

- **Concluding thoughts**
  - WGS benefits and challenges
SURVEILLANCE AND SUBTYPING FOR LISTERIOSIS
Listeria monocytogenes

- **Outbreak detection is currently WGS-centric**
  - Chosen for the CDC WGS pilot project
  - Other subtyping methods not ideal
    - Low diversity, not easy to assess relatedness by PFGE
    - Genome is small and easy to sequence and analyze

- **Great standard epidemiologic data collection available**
  - *Listeria* Initiative adopted in 2004

- **Strong regulatory component**
  - Provides drive for enhanced epi information
Background

**Listeria Initiative**

- **State/local health departments interview all patients**
  - Using a standard and intensive form for all jurisdictions in US
  - Whether or not part of an outbreak
    - Maximize quality of data

- **CDC monitors for clusters**
  - When cluster detected, epidemiologists analyze questionnaire data
  - ID associations of exposure
Listeria Outbreaks and Incidence, 1983-2014

No. outbreaks

Outbreak Incidence

Era
Outbreaks per year Pre-PulseNet
Median cases per outbreak

1983 0.3 69
1985 1
1987 1
1989 1
1991 1
1993 1
1995 1
1997 1
1999 0
2001 0
2003 0
2005 0
2007 0
2009 0
2011 0
2013 1

Incidence (per million pop)
Listeria Outbreaks and Incidence, 1983-2014

- Outbreaks per year:
  - Pre-PulseNet: 0.3
  - Early PulseNet: 2.3

- Median cases per outbreak:
  - Pre-PulseNet: 69
  - Early PulseNet: 11

Incidence (per million pop):
- Pre-PulseNet: 0.3
- Early PulseNet: 2.3
Listeria Outbreaks and Incidence, 1983-2014

No. outbreaks

Incidence (per million pop)

<table>
<thead>
<tr>
<th>Year</th>
<th>Outbreaks</th>
<th>Incidence</th>
</tr>
</thead>
<tbody>
<tr>
<td>1983</td>
<td>1</td>
<td>0.3</td>
</tr>
<tr>
<td>1985</td>
<td>1</td>
<td>0.3</td>
</tr>
<tr>
<td>1987</td>
<td>1</td>
<td>0.3</td>
</tr>
<tr>
<td>1989</td>
<td>1</td>
<td>0.3</td>
</tr>
<tr>
<td>1991</td>
<td>1</td>
<td>0.3</td>
</tr>
<tr>
<td>1993</td>
<td>1</td>
<td>0.3</td>
</tr>
<tr>
<td>1995</td>
<td>1</td>
<td>0.3</td>
</tr>
<tr>
<td>1997</td>
<td>1</td>
<td>0.3</td>
</tr>
<tr>
<td>1999</td>
<td>1</td>
<td>0.3</td>
</tr>
<tr>
<td>2001</td>
<td>1</td>
<td>0.3</td>
</tr>
<tr>
<td>2003</td>
<td>1</td>
<td>0.3</td>
</tr>
<tr>
<td>2005</td>
<td>1</td>
<td>0.3</td>
</tr>
<tr>
<td>2007</td>
<td>1</td>
<td>0.3</td>
</tr>
<tr>
<td>2009</td>
<td>1</td>
<td>0.3</td>
</tr>
<tr>
<td>2011</td>
<td>1</td>
<td>0.3</td>
</tr>
<tr>
<td>2013</td>
<td>1</td>
<td>0.3</td>
</tr>
</tbody>
</table>

Era
Outbreaks per year
Pre-PulseNet: 0.3
Early PulseNet: 2.3
Listeria Initiative: 2.9

Median cases per outbreak
Pre-PulseNet: 69
Early PulseNet: 11
Listeria Initiative: 5.5
CDC *Listeria* WGS Pilot Project

- Goal: Sequence all *L. monocytogenes* isolates in US
- Near real time
- Perfect project to piggy-back onto *Listeria* Initiative
*Listeria* Outbreaks and Incidence, 1983-2014

<table>
<thead>
<tr>
<th>Era</th>
<th>Outbreaks per year</th>
<th>Median cases per outbreak</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pre-PulseNet</td>
<td>0.3</td>
<td>69</td>
</tr>
<tr>
<td>Early PulseNet</td>
<td>2.3</td>
<td>11</td>
</tr>
<tr>
<td><em>Listeria Initiative</em></td>
<td>2.9</td>
<td>5.5</td>
</tr>
<tr>
<td>WGS</td>
<td>9</td>
<td>4.5</td>
</tr>
</tbody>
</table>

No. outbreaks

Incidence (per million pop)
SURVEILLANCE AND SUBTYPING FOR SALMONELLA AND E. COLI
Salmonella and E. coli

- Outbreak detection and investigation still PFGE-centric
  - WGS only used for additional subtyping of selected outbreak isolates for better clarity resolution
    - Mostly for non-O157 and Salmonella

- Limited epi data available
  - Routine interviews are minimal
    - Only extensive interviews on cases in identified clusters

- An epi initiative would be much harder considering the numbers of cases we see annually compared to Listeria
UTILITY OF WGS IN PUBLIC HEALTH

Why there is a drive for WGS and why PFGE is now considered “inferior”?
Characterization of Enteric Pathogens Today

**CDC Specimen ID:** 2014003970  
**CDC Unique ID:** N8K7D8C1  
**CDC Locs ID:** 2014C-3008

- **GENUS/SPECIES:** *Escherichia coli*
- **SERO TYPE:** O104:H4*
- **PATHOTYPE:** Shiga toxin producing and Enterotoxigenic *E. coli* (STEC & Eagg EC)
- **VIRULENCE PROFILE:** stx2a, aggR, aggA
- **SEQUENCE TYPE:** ST678
- **ANTIMICROBIAL RESISTANCE:** Ampicillin, Cefoxitin, Ceftriaxone, Streptomycin, Tetracycline, Sulfamethoxazole/Trimethoprim

**Disclaimer:** 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.

**Explanation of Virulence Markers**

- Biochemical ‘panel’
- O and H agglutination
- Min. 2 PCRs + RFLP
- 7 PCRs + sequencing
- Disc diffusion OR broth micro dilution
- PFGE typing

**TAT:** 1-2 weeks
But I do still love PFGE......

Pulse-Field Gel Electrophoresis (PFGE):

- PulseNet database
- Gold standard for outbreak detection
- Limited resolution because it only interrogates small regions of the microbial genome
**PFGE recap**

**Different Organisms**

- **A**
- **B**
  - Restriction site
  - Bacterial Genome

**Genetically Related Organisms**

- **C**
- **D**
  - Single Nucleotide difference

**PFGE DNA FINGERPRINT**

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

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

- **A**
- **B**
- **C**
- **D**
PFGE vs WGS

- PFGE = world map
- WGS = Google Earth
Why integrate WGS into current surveillance methods?

Pulse-Field Gel Electrophoresis (PFGE):
- PulseNet
- Gold standard for outbreak detection
- Limited resolution because it only interrogates small regions of the microbial genome

Whole Genome Sequencing (WGS):
- High resolution
- Improve discrimination between sporadic and outbreak-associated isolates
- "future proof"
- Context of relatedness
PRACTICAL IMPACT OF WGS ON OUTBREAK INVESTIGATIONS
BREAKING UP PFGE CLUSTERS

And refining case definitions!
All *Salmonella* Enteritidis JEGX01.0004/JEGA26.0002 Isolates Received August-September 2014 (n=19)

- Over 85% of SE isolates can be classified into 5 PFGE types
- Of the SE isolates collected in 1 year in NY, half were denoted PFGE type 4

Levinson et al., NY Wadsworth Center, unpublished data
den Bakker et al., 2014 Rapid Whole Genome Sequencing for Surveillance of SE
**All Salmonella Enteritidis JEGX01.0004/JEGA26.0002 Isolates Received August-September 2014 (n=19)**

The chart illustrates the number of cases of *Salmonella Enteritidis* JEGX01.0004/JEGA26.0002 isolates received by MDH per week in August and September 2014. The chart shows the distribution of isolates by week, with a focus on isolates without SNPs (0 SNPs). The data indicates that most isolates were received in the later part of August and early September.
All *Salmonella* Enteritidis JEGX01.0004/JEGA26.0002 Isolates Received August-September 2014 (n=19)

**Number of Cases**

- August 2014: 1, 4, 11, 18, 25
- September 2014: 1, 8, 15, 22, 29

**Week of Isolate Receipt at MDH**

- 0 SNPs
- Ate chicken Kiev
- May have eaten chicken Kiev
- Secondary case
- Did not report eating stuffed chicken products
TYING PFGE PATTERNS TOGETHER
Possible Listeriosis Cluster in Kansas March 2015

- Kansas reports 2 listeriosis cases same PFGE from 1 hospital
- Review annual data - 3 other listeriosis patients same hospital
  - different PFGE patterns??
- Four isolates highly related by WGS

<table>
<thead>
<tr>
<th>cdc_id</th>
<th>Id</th>
<th>State ID</th>
<th>PFGE-Asci-pattern</th>
<th>PFGE-Apal-pattern</th>
</tr>
</thead>
<tbody>
<tr>
<td>2014L-6727</td>
<td>PNUSAL001200</td>
<td>KS___211123</td>
<td>GX6A16.0026</td>
<td>GX6A12.0227</td>
</tr>
<tr>
<td>2015L-6049</td>
<td>PNUSAL001324</td>
<td>KS___268222</td>
<td>GX6A16.0026</td>
<td>GX6A12.0227</td>
</tr>
<tr>
<td>2014L-6083</td>
<td>PNUSAL000593</td>
<td>ks___42951</td>
<td>GX6A16.0061</td>
<td>GX6A12.0026</td>
</tr>
<tr>
<td>2014L-6129</td>
<td>PNUSAL000635</td>
<td>KS___79910</td>
<td>GX6A16.0020</td>
<td>GX6A12.0227</td>
</tr>
<tr>
<td>2014L-6728</td>
<td>PNUSAL001201</td>
<td>KS___211134</td>
<td>GX6A16.0282</td>
<td>GX6A12.0355</td>
</tr>
</tbody>
</table>

wgMLST (<All Characters>)

<table>
<thead>
<tr>
<th>Allele mec</th>
<th>st node</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.0</td>
<td>100</td>
</tr>
<tr>
<td>12.5 [0-18]</td>
<td></td>
</tr>
<tr>
<td>2571</td>
<td></td>
</tr>
</tbody>
</table>
LINK BETWEEN HUMAN AND PRODUCT ISOLATES
Pre-Packaged Salad Recall, 2014
WGS Links Sporadic Case

- **March 13**: Lettuce recall
  - *Listeria monocytogenes*
  - Did it cause anyone to get sick??

- **April 21**: OH case w/ PFGE pattern matching lettuce

- **April 25**: Patient consumed multiple types of pre-packaged lettuce, possibly the recalled brand.

- **May 13**: Canada reports clinical and lettuce isolates match by WGS methodology
DISCUSSION
What we learned using WGS?

- Use of routine, prospective WGS-enhanced surveillance for *Listeria*
  - Detected more clusters and solved more outbreaks

- Use of WGS as a further sub-typing method for *Salmonella* and *E. coli*
  - Appears likely to be able to break apart common PFGE patterns into more meaningful clusters

- Ascertain relatedness of individual isolates to each other, no matter their source.
Challenges

- We need to harmonize methods
- Where to store the data?
- Difference between good and bad sequencing runs?
  - QC data?

- What’s a match? How to interpret it?
  - Accuracy depends on sequencing methods and how the data is analyzed.
  - No pattern names: up-front effort to decide how best to communicate
    - Trees and heat maps vs. a name

- Collaboration between lab and epi is essential
  - Epi data remain essential to understand sequencing results and to take public health action
Acknowledgments

CDC Enteric Diseases Laboratory Branch (Matt Wise and Brandon Jackson)

CDC Enteric Diseases Epidemiology Branch

State, Local, and Territorial Health Departments
  Minnesota Dept of Health (Kirk Smith, Carlota Medus, David Boxrud)
  NY Dept of Health, Wadsworth Center (Kara Levinson)

Food and Drug Administration

US Department of Agriculture’s Food Safety and Inspection Service

Outbreak Response and Prevention Branch

Epidemic Intelligence Service

The findings and conclusions in this report are those of the author and do not necessarily represent the official position of the Centers for Disease Control and Prevention, or other state health departments mentioned in this presentation.
Thank you!
PRODUCT TESTING-DRIVEN OUTBREAKS
Listeriosis Linked to Recalled Stone Fruits

- July 2014 recall receives extensive media coverage
- Hundreds of inquiries from concerned clinicians and public
- Stone fruit isolates obtained from company
- 4 human isolates in 2014 with PFGE match
- Was it an outbreak?
Stone Fruit WGS: Epi-Lab Concordance

- MA and MN patient isolates highly-related to stone fruit isolates
  - MA patient and MN patient ate nectarines and peaches

- SC and IL patient isolates not highly-related to stone fruit isolates
  - SC patient did not eat recalled fruits and IL patient had no exposure information available
**Listeria Conclusions**

- **Listeria Initiative** - Proactive rather than reactive
  - Importance of having epi data before you need it

- **WGS pilot project** - Blending of new technology with enhanced epi data
  - The proof is in the puddin’ – the numbers show it
    - Increased number of outbreaks detected and the median number of cases in outbreaks were smaller
Salmonella enterica serovar Enteritidis

PFGE JEGX01.0004 contains 45 genomic clusters

PFGE JEGX01.0002 contains 8 genomic clusters

PFGE JEGX01.0034 contains 5 genomic clusters

PFGE JEGX01.0056 contains 4 genomic clusters

PFGE JEGX01.0021 contains 12 genomic clusters

Levinson et al., NY Wadsworth Center, unpublished data
den Bakker et al., 2014 Rapid Whole Genome Sequencing for Surveillance of Salmonella enterica Serovar Enteritidis