Managing Workflows in the Era of WGS

Dave Boxrud, Molecular Epidemiology Supervisor
Carlota Medus, Epidemiologist Supervisor Sr.
Use of WGS in Minnesota

- Sequence:
  - All *Salmonella*
  - All STEC
  - All *Campylobacter*
  - All *Listeria*

Summer: Four times per week
Winter: Twice a week minimum

- Serotyping by traditional methods ended Spring 2018

- Analyze routinely for surveillance, cluster detection:
  - *S. Enteritidis*, since 2014
  - *S. Typhimurium*, since 2017
  - *Listeria*
  - *Campylobacter*, since 2018

- Other pathogens upon request
PulseNet Traditional

PFGE Group → Clusters → Results ← Epi
Stat Communication (clusters)

Daily Communication

Enteric Isolates Reported on 15-Mar-2003

<table>
<thead>
<tr>
<th>SPEC</th>
<th>LNAME</th>
<th>FNAME</th>
<th>CITY</th>
<th>AGE</th>
<th>AGENT</th>
<th>SUBTYPE</th>
<th>DATE</th>
</tr>
</thead>
<tbody>
<tr>
<td>01774</td>
<td>Larsen</td>
<td>William</td>
<td>Bloomington</td>
<td>37</td>
<td>Campylobacter jejuni</td>
<td>FLA0</td>
<td>11-mar-03</td>
</tr>
<tr>
<td>01779</td>
<td>Simon</td>
<td>Pauline</td>
<td>Minneapolis</td>
<td>72</td>
<td>Campylobacter jejuni</td>
<td>FLA72</td>
<td>10-mar-03</td>
</tr>
<tr>
<td>01778</td>
<td>Hildren</td>
<td>Sveth</td>
<td>Rosemount</td>
<td>6</td>
<td>Escherichia coli</td>
<td>MN21</td>
<td>10-mar-03</td>
</tr>
<tr>
<td>01773</td>
<td>Hildren</td>
<td>Cody</td>
<td>Rosemount</td>
<td>5</td>
<td>Escherichia coli</td>
<td>MN21</td>
<td>12-mar-03</td>
</tr>
<tr>
<td>01765</td>
<td>Bergstrom</td>
<td>Theresa</td>
<td>Rosemount</td>
<td>31</td>
<td>Escherichia coli</td>
<td>MN21</td>
<td>11-mar-03</td>
</tr>
<tr>
<td>01768</td>
<td>Roberts</td>
<td>Marvin</td>
<td>Bemidji</td>
<td>76</td>
<td>Escherichia coli</td>
<td>MN1</td>
<td>11-mar-03</td>
</tr>
<tr>
<td>01777</td>
<td>Desowitz</td>
<td>Robert</td>
<td>St. Cloud</td>
<td>47</td>
<td>Salmonella saintpauli</td>
<td>14-mar-03</td>
<td></td>
</tr>
<tr>
<td>01779</td>
<td>Beer</td>
<td>Trevor</td>
<td>Minneapolis</td>
<td>3</td>
<td>Salmonella typhimurium</td>
<td>TM43</td>
<td>10-mar-03</td>
</tr>
<tr>
<td>01769</td>
<td>Sampson</td>
<td>Elsa</td>
<td>St. Paul</td>
<td>53</td>
<td>Salmonella typhimurium</td>
<td>TM22</td>
<td>11-mar-03</td>
</tr>
<tr>
<td>01774</td>
<td>Brady</td>
<td>Harold</td>
<td>Bloomington</td>
<td>16</td>
<td>Shigella flexneri</td>
<td>Sa</td>
<td>13-mar-03</td>
</tr>
<tr>
<td>01767</td>
<td>Grandon</td>
<td>Louise</td>
<td>Chicago</td>
<td>24</td>
<td>Shigella sonnei</td>
<td>SS1</td>
<td>10-mar-03</td>
</tr>
<tr>
<td>01764</td>
<td>Petrovich</td>
<td>Helen</td>
<td>Edina</td>
<td>67</td>
<td>Shigella sonnei</td>
<td>SS44</td>
<td>09-mar-03</td>
</tr>
</tbody>
</table>

Names are fictitious.
## WGS Data Types

<table>
<thead>
<tr>
<th>Sample ID</th>
<th>Collection Date</th>
<th>Sample Type</th>
<th>Library Type</th>
<th>Library ID</th>
<th>Library Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>SampleA</td>
<td>03/15/2019</td>
<td>Blood</td>
<td>Normal</td>
<td>LB001</td>
<td>Normal Library</td>
</tr>
<tr>
<td>SampleB</td>
<td>03/16/2019</td>
<td>Tumor</td>
<td>Tumor Library</td>
<td>LB002</td>
<td>Tumor Library</td>
</tr>
</tbody>
</table>

### WGS Data Types

- BG: Barcode
- BC: Barcode
- BS: Barcode
- DR: Detected Reads
- SR: Sequence Reads
- ZU: Zygosity

### WGS Data Analysis

- Q20: Sequences with Phred Quality Score > 20
- Q30: Sequences with Phred Quality Score > 30

### WGS Data Summary

- Total Reads: 10,000,000
- Q20 Reads: 9,500,000
- Q30 Reads: 8,500,000

### WGS Data QC

- Base Call Quality: 98.5%
- Base Call Error: 0.2%

---

*Note: WGS Data Types are summarized in the attached spreadsheet for detailed analysis.*
Current WGS Communication

Cluster 2017001: 2 new isolates are 1-5 SNPs from others in the cluster. Now 8 isolates total. (Marijke, SE11B6)
E2017005289
E2017004460

Cluster 2017008: 2 new isolates are 1-5 SNPs from others in the cluster. Now 4 isolates total. (SE11B116)
I2017005940-1
E2017003821

Cluster 2017010: 1 new isolate is 0 SNPs from others in the cluster. Now 3 isolates total. (SE11B6)
E2017006068

NEW Cluster 2017013: 2 isolates are 0 SNPs from each other.
E2017003849 (SE235B78)
E2017005056 (SE181B93)
### Determining Cluster Range for Cluster Detection

<table>
<thead>
<tr>
<th>Benefit</th>
<th>Narrow Range</th>
<th>Broad Range</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cluster cases more likely to have common exposure</td>
<td>All cluster cases identified</td>
<td></td>
</tr>
<tr>
<td>Excluding true cluster cases</td>
<td>Including not true cluster cases</td>
<td></td>
</tr>
</tbody>
</table>

Additional Considerations:
- Philosophy may adjust during an outbreak (narrow at beginning, broad after source is identified)
- Based range on your experience (retrospective outbreaks helpful)
- Understand exceptions (polymicrobial outbreaks, animal exposures, long time-frame outbreaks)
Cluster Identification Using cgMLST

• Use previous OBs to identify clusters
• Allele code
• LIMS
• Frequent discussion with epi

Defined Outbreak Samples
- Outbreak 1: Sept 2000
- Outbreak 2: May 2001
- Outbreak 3: Aug 2001
- Outbreak 4: Nov 2003
- Outbreak 5: Aug 2008
- Outbreak 6: Spring 2014
- Outbreak 7: Spring 2014

• Still to be determined
• Focus on rapid and easy communication
• Likely will be in a variety of ways-email, LIMS, dendrograms, etc
• Epi will be a partner in determining methods
• Refine process regularly
Post PFGE Testing

• WGS offers advantage of better epi concordance and much more information, however it is slower than current methods

• MDH will have limited *Salmonella* serogrouping
  – Complicated, polyclonal outbreaks-animals, petting zoos, etc
  – Some cluster investigations (this needs discussion)

• MDH may continue limited *Salmonella* PFGE
  – Some cluster investigations (this needs discussion)
### Cases, Select Pathogens, Minnesota, 2013-2017

- For all of these pathogens, submission to the PHL is required
- Population: 5.6 million

<table>
<thead>
<tr>
<th>PATHOGEN</th>
<th>Median n per year</th>
<th>Min. n</th>
<th>Max. n</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Salmonella</strong></td>
<td>861</td>
<td>745</td>
<td>1002</td>
</tr>
<tr>
<td><strong>STEC</strong></td>
<td>286</td>
<td>223</td>
<td>342</td>
</tr>
<tr>
<td><strong>Campylobacter</strong></td>
<td>927</td>
<td>834</td>
<td>1049</td>
</tr>
<tr>
<td><strong>Listeria</strong></td>
<td>12</td>
<td>3</td>
<td>17</td>
</tr>
</tbody>
</table>
Two factors impacting the epi workflow and workload:

• Turn-around time
  – Serotyping
  – Subtyping

• Increase in specificity
Isolation Timeline *Salmonella*

Culture at Clinical Laboratory

- Stool sample collected
- CIDT at Clinical Laboratory

- Case report from health care provider
- Results culture 2 to 3 d
- Isolate Received at PHL*
- Identification of isolated pathogen

- Results CIDT 0 to 1 d
- Specimen Received at PHL*
- Identification of isolated pathogen

Case report from health care provider

*Shipping time = 1 – 3 days*
Characterization Timeline *Salmonella*

- **Identification of isolated pathogen**: 5 to 7 days
- **Traditional Serotyping**: Days 9 to 11
- **PFGE (in parallel to serotyping)**: Days 10 to 12
- **Traditional Methods Sequencing**: Days 12 to 19
- **WGS**: Day 18
Epi Follow-up Timeline *Salmonella*

**Pre-WGS**
- Identification of isolated pathogen (5 to 7 d)
- PFGE or serotyping reported
- Case investigation started

**Now**
- WGS
- Start case investigation?
Interviewing Cases: Basic Philosophy

• Interview all cases in surveillance

• Interview ASAP

• Collect details on specific exposures
  – Dates
  – Restaurant, grocery store names
  – Brand names
  – Open-ended food histories

• Dynamic investigation approach (add questions regarding exposures mentioned by other cases, and call cases back with additional questions)
Interviewing Cases, PFGE Era

- Dynamic investigation approach (a.k.a. iterative; add questions regarding exposures mentioned by cases, and call cases back with additional questions)
Impact on our Basic Approach

Notified of cluster

PFGE

Case #1

Case #2

Case #3

Case #4

WGS

Case #1

Case #2

Case #3

Case #4

Retrospectively dynamic?
### Allocating Cases to Cluster Investigations

<table>
<thead>
<tr>
<th>Case in surveillance</th>
<th>Salmonella serotype</th>
</tr>
</thead>
<tbody>
<tr>
<td>Case 1</td>
<td>INFANTIS</td>
</tr>
<tr>
<td>Case 2</td>
<td>ENTERITIDIS</td>
</tr>
<tr>
<td>Case 3</td>
<td>HVITTINGFOSS</td>
</tr>
<tr>
<td>Case 4</td>
<td>SENFTENBERG</td>
</tr>
<tr>
<td>Case 5</td>
<td>TYPHIMURIUM</td>
</tr>
<tr>
<td>Case 6</td>
<td>TYPHIMURIUM</td>
</tr>
<tr>
<td>Case 7</td>
<td>BRAENDERUP</td>
</tr>
<tr>
<td>Case 8</td>
<td>ENTERITIDIS</td>
</tr>
<tr>
<td>Case 9</td>
<td>NEWPORT</td>
</tr>
<tr>
<td>Case 10</td>
<td>NEWPORT</td>
</tr>
<tr>
<td>Case 11</td>
<td>GEORGIA</td>
</tr>
<tr>
<td>Case 12</td>
<td>4,12:i:-</td>
</tr>
<tr>
<td>Case 13</td>
<td>4,5,12:i:-</td>
</tr>
<tr>
<td>Case 14</td>
<td>CHAILEY</td>
</tr>
</tbody>
</table>

**Clusters**

1. Case 1
2. Case 2
3. Case 3
4. Case 5
5. Case 7

**Non-cluster**

Case 4

---

### Allocation

- **Cluster 1:** Case 1
- **Cluster 2:** Case 2, Case 3
- **Cluster 3:** Case 4
- **Cluster 4:** Case 5, Case 7
- **Cluster 5:** Case 6
- **Non-cluster:** Case 4, Case 7

---

### Notes

- Cases 12 and 13 have non-matching serotypes (4,12:i:- and 4,5,12:i:-) and are not clustered.
- Case 14 has serotype CHAILEY and is not clustered.
# Impact on Allocating Cases to Cluster Investigations

<table>
<thead>
<tr>
<th>Clusters</th>
<th>Case in surveillance</th>
<th>Salmonella serotype</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Case 1</td>
<td>Pending</td>
</tr>
<tr>
<td></td>
<td>Case 2</td>
<td>Pending</td>
</tr>
<tr>
<td></td>
<td>Case 3</td>
<td>Pending</td>
</tr>
<tr>
<td></td>
<td>Case 4</td>
<td>Pending</td>
</tr>
<tr>
<td></td>
<td>Case 5</td>
<td>Pending</td>
</tr>
<tr>
<td></td>
<td>Case 6</td>
<td>Pending</td>
</tr>
<tr>
<td></td>
<td>Case 7</td>
<td>Pending</td>
</tr>
<tr>
<td></td>
<td>Case 8</td>
<td>Pending</td>
</tr>
<tr>
<td></td>
<td>Case 9</td>
<td>Pending</td>
</tr>
<tr>
<td></td>
<td>Case 10</td>
<td>Pending</td>
</tr>
<tr>
<td></td>
<td>Case 11</td>
<td>Pending</td>
</tr>
<tr>
<td></td>
<td>Case 12</td>
<td>Pending</td>
</tr>
<tr>
<td></td>
<td>Case 13</td>
<td>Pending</td>
</tr>
<tr>
<td></td>
<td>Case 14</td>
<td>Pending</td>
</tr>
</tbody>
</table>

Clusters 1, 2, 3, 4, 5, and non-cluster.
Ongoing vs. Completed Transmission, PFGE

- Date of Notification
- Onset date

Graphs showing the number of cases over different onset dates.
Ongoing vs. Completed Transmission, WGS

Date of Notification

Number of cases

Onset date

Date of Notification

Number of cases

Onset date
Impact on Cluster Investigation, WGS

• At the time of interview, serotype and subtype unavailable
  – Rely on epi data for up-front identifications of some outbreaks
    ▪ E.g., restaurants, events, child care, long term care facilities
  – Retrospective identification of clusters using WGS resulting in re-interviewing all or most cases in a cluster

• Benefits of interviewing cases earlier
  – Earlier interventions (child care, work exclusions)
  – Improved recall of cases being interviewed
Factors Affecting Epi Workflow

• Increase in specificity
  – Smaller clusters
  – Cases in the cluster more likely to actually be part of the outbreak
  – Less time investigating clusters of unrelated cases
<table>
<thead>
<tr>
<th></th>
<th>PFGE</th>
<th>WGS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of clusters</td>
<td>39</td>
<td>92</td>
</tr>
<tr>
<td>Median number of cases per cluster</td>
<td>4</td>
<td>2</td>
</tr>
<tr>
<td>Range of number of cases per cluster</td>
<td>2 - 99</td>
<td>2 - 18</td>
</tr>
<tr>
<td>Number of clusters with ≥10 cases</td>
<td>12</td>
<td>7</td>
</tr>
<tr>
<td>Number of clusters with pattern .0004 isolates</td>
<td>7</td>
<td>28</td>
</tr>
<tr>
<td>Number of clusters with pattern .0002 isolates</td>
<td>5</td>
<td>27</td>
</tr>
</tbody>
</table>
### S. Typhimurium Cluster Comparison, 2017

$n = 191$ isolates

<table>
<thead>
<tr>
<th></th>
<th>PFGE</th>
<th>WGS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of clusters</td>
<td>25</td>
<td>17</td>
</tr>
<tr>
<td>Median number of cases per cluster</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>Range of number of cases per cluster</td>
<td>2 - 7</td>
<td>2 - 6</td>
</tr>
<tr>
<td>Number of clusters with ≥3 cases</td>
<td>8 (32%)</td>
<td>8 (47%)</td>
</tr>
</tbody>
</table>
Conclusions - Impact

• The impact of stopping serotyping by traditional methods on epi was not readily apparent until serotyping ended
  – Time frame for case follow-up and interviews
  – Workflows/Logistics
• Longer TAT for both serotyping and sequencing data will impact
  – Timeliness of cluster detection
  – Sequence of events (most cases not interviewed vs. interviewed when cluster detected)
  – Need to rely on epi signals until WGS results available
Conclusions - Work in Progress

• Lab-Epi communication
  – How to communicate results
  – What data to communicate daily vs. as needed

• Clusters
  – Different definitions for cluster detection vs. active investigation
    - Use outbreak data as guide: Evaluate the range of alleles/SNPs among cases who reported the outbreak-vehicle/exposure
  – More vs. fewer clusters TBD, will vary by pathogen/serotype
Conclusions - Benefits

- Specificity
  - Better epi concordance
  - Save time by not investigating unrelated cases (PFGE clusters that are not epi-linked)
- WGS is a very powerful tool for epidemiological investigations
March 15

“It’s the end of the world as we know it
It’s the end of the world as we know it
It’s the end of the world as we know it and I feel fine”
R.E.M.

https://www.youtube.com/watch?v=-OIsr36N1Hs&feature=youtu.be
Thank you
Comments?

Dave.Boxrud@state.mn.us
Carlota.Medus@state.mn.us
Selected Enteric Pathogens Reported to MDH 2000-2017

Year of Report

Number of Cases

Median = 932
Median = 705
Median = 178
Median = 306
Median = 251
Median = 135

MN population: 5.6 million
### Challenges & Concerns: Turn-Around Time

<table>
<thead>
<tr>
<th>Performance Metrics</th>
<th>Salmonella</th>
<th>STEC</th>
</tr>
</thead>
<tbody>
<tr>
<td>Days from isolate receipt/recovery to PFGE upload to PulseNet</td>
<td>2</td>
<td>4</td>
</tr>
<tr>
<td>Days from isolate receipt/recovery to WGS sequence being shared</td>
<td>9</td>
<td>17</td>
</tr>
</tbody>
</table>

- Data as of July 2018 (will improve)
- Longer turn-around time (TAT) for WGS analysis vs. PFGE
- Longer TAT for serotyping using WGS