Establishing a National Surveillance Network for Foodborne Pathogens Based on Whole Genome Sequencing

Steven Musser, Ph.D.
Deputy Center Director for Scientific Operations
Center for Food Safety and Applied Nutrition, FDA

Next-Generation Sequencing for Food Pathogen Traceability
UCD Institute of Food and Health *in conjunction with* UCD Centre for Food Safety
and the Food Safety Authority of Ireland - March 24, 2014
Each year 9.4 million episodes of foodborne illness in the United States

55,961 hospitalizations

1,351 deaths

Salmonella spp. cause 11% of foodborne illnesses each year

(Scallan et al. 2011 Emerging Infectious Diseases • www.cdc.gov/eid).
The Public Health Need

Clinical ID and fingerprint

Identify Food and confirm Fingerprint

Product enters commerce

Source of contamination identified too late
Some perspective on the food supply

• Tracking and Tracing of food pathogens
  • Almost 200,000 registered food facilities (2/14)
    – 81,574 Domestic and 115,753 Foreign
  • More than 300 ports of entry
  • More than 130,000 importers and more than 11 million import lines/yr
  • In the US there are more than 2 million farms
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.
Is WGS a viable solution?

- Cost
- Increasing ease of operation
- Database longevity
- Sample prep
  - Identical for all pathogens
- Cost savings
  - Resistance, subtyping, virulence factors, more...
- New applications
  - tracking, regulatory/compliance actions, historical trends, more...

**Cost per bacterial genome**

- **Miseq**
  - $70/genome in 2014

<table>
<thead>
<tr>
<th>Year</th>
<th>Cost per Genome</th>
</tr>
</thead>
<tbody>
<tr>
<td>2007</td>
<td>$3,500</td>
</tr>
<tr>
<td>2008</td>
<td>$3,250</td>
</tr>
<tr>
<td>2009</td>
<td>$3,000</td>
</tr>
<tr>
<td>2010</td>
<td>$2,750</td>
</tr>
<tr>
<td>2011</td>
<td>$2,500</td>
</tr>
<tr>
<td>2012</td>
<td>$2,250</td>
</tr>
<tr>
<td>2013</td>
<td>$2,000</td>
</tr>
</tbody>
</table>
Next-generation Lab Response vs. Conventional Lab Response

Next-Generation Lab Response

1. DNA extraction
2. Library prep
3. Sequencing run
4. NCBI Pipeline
5. Data upload
6. Serology

Start of Outbreak

Conventional Lab Response

1. Sample accession
2. Colony isolated
3. PFGE plugs made
4. Plugs digested
5. Gel run
6. Results uploaded

1 = Day 1
11/27/2013
This from 1859, Darwin's, On the Origin of Species

- “It is obvious that the Galapagos Islands would be likely to receive colonists, whether by occasional means of transport or by formerly continuous land, from America; and the Cape de Verde Islands from Africa; and that such colonists would be liable to modification;— the principle of inheritance still betraying their original birthplace”

With WGS, we now have the potential to discern those birthplaces…
Can WGS fill a Public Health role?

• If yes, then...
  • Initiate pilot study
  • Develop collaborations and partnerships
    – NCBI, States, CDC and other Federal partners
  • What infrastructure would be needed?
  • Support multiple sequencing platforms?
    – Multiple data formats
    – How reproducible are the data AND answers?
  • How would data be accessed and stored?
    – Public vs. private **No data hoarding**
  • Metadata
Metadata

- Simple but complete for each Strain

- Clinical or environmental (specific source)
  - Environmental swab or type of food

- Location as accurate as allowable
  - State, Region, Country

- Submitter name – Usually organization

- Date of isolation
Network Requirements

• Well characterized strain sets

• A large database of sequences with accurate metadata

• A network of sequencing labs

• Analytical software

• Somewhere to store the data
FDA, USDA, CDC

State, Local, Federal and Foreign Public Health Agencies

Academia

NCBI, EMBL, DDBJ (Public Access Database)

GenomeTrakr

High Capacity Sequencing of Collections

100K Genome Project

Network of Sequencers

DATA ANALYSIS

DATA ASSEMBLY AND STORAGE

DATA ACQUISITION
FDA provides

- 1 Miseq system
- Sufficient reagents to sequence > 300 genomes per year
- Dedicated scientific staff (bioinformatics and/or laboratory support) through Oak Ridge Institute for Science and Education (ORISE)
- Bioinformatics and laboratory support, analysis pipeline

Network Lab provides

- Minimum ~300 genomes with metadata uploaded to NCBI per annum, minimum 20X coverage
- Food and environmental related bacterial (prefer Salmonella) isolates

Cost to FDA ≈ $200k/lab
7 state health depts. + 10 FDA-ORA

Network of Sequencers
FDA-State Desktop Pilot called GenomeTrakr


Food and Drug Administration, Center for Food Safety and Applied Nutrition: GenomeTrakr Project

Currently encompasses whole genome sequencing of cultured pathogens as part of a surveillance project for the rapid detection of outbreaks of foodborne illnesses.

Project Type: Umbrella project (Subtype: Disease)

Relevance: Agricultural, Medical, Industrial, Environmental

Project Data:

<table>
<thead>
<tr>
<th>Resource Name</th>
<th>Number of Links</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sequence Data</td>
<td></td>
</tr>
<tr>
<td>Nucleotide (total)</td>
<td>31172</td>
</tr>
<tr>
<td>WGS master</td>
<td>194</td>
</tr>
<tr>
<td>Genomic DNA</td>
<td>30978</td>
</tr>
<tr>
<td>SRA Experiments</td>
<td>437</td>
</tr>
<tr>
<td>Protein Sequences</td>
<td>901056</td>
</tr>
<tr>
<td>Publications</td>
<td></td>
</tr>
<tr>
<td>PubMed</td>
<td>12</td>
</tr>
<tr>
<td>PMC</td>
<td>9</td>
</tr>
<tr>
<td>Other Datasets</td>
<td></td>
</tr>
<tr>
<td>BioSample</td>
<td>1104</td>
</tr>
<tr>
<td>GEO DataSets</td>
<td>1</td>
</tr>
</tbody>
</table>

Gene expression:

Highest level of assembly:

- SRA or Trace: 1
- No data links: 4
- Total: 5

Accession: PRJNA183844  ID: 183344

http://www.fda.gov/Food/FoodScienceResearch/WholeGenomeSequencingProgramWGS/ucm363134.htm
### Expanding the network

**Partners with sequencers**
- United Kingdom - FSA
- Canada – CFIA and PHAC
- Argentina - WHO
- Taiwan

**State Partners**
- 6 States have requested funding
- Organizations/Countries joining the network

**Partners with isolates**
- Ireland
- Mexico
- Turkey
- Columbia
- Chile
- Brazil
- Thailand
- Ethiopia

- APHL
- WHO
- USDA
- GMI
- Italy
- Germany
- Denmark
- Australia
- Spain
Now What?

- NGS clearly defines foodborne outbreaks – more than 15 different examples
- NGS network is reliable, efficient and can provide very good location specificity of outbreaks
- We have sequenced about 2900 Salmonella, more than 900 Listeria, and closed 100 genomes. Our current rate is about 500 sequences a month.
- The need for increased number of well characterized environmental (food, water, facility, etc.) sequences may outweigh need for extensive clinical isolates
- Many requests for information or help from other public health labs
Needs/concerns

• Network security issues
  – Sequencers
  – Software

• Improved informatics and software development
  – Widely available commercial solutions
  – Custom solutions
  – Automated identification of AMR, virulence markers, etc

• Cloud computing and access to HPC

• Data presentation to different groups
  – Physicians
  – Epidemiologists
  – Researchers
FDA - CFSAN
Marc Allard          Rebecca Bell
Eric Brown          Andrea Ottesen   James Pettengill
Ruth Timme          Jie Zheng        Charlie Wang
Christine Keys      Cong Li
Errol Strain        Yan Luo
Mark Mammel          Darcy Hanes

FDA Division of Field Sciences  Rebecca Dreisch
NYPH  Bill Wolfgang Kimberly Musser and colleagues
MPH  Alvina Chu and colleagues
FDH  Anita Wright Judy Johnson
ADPH  Victor Waddell  Dave Engelthaller Paul Keim
WDH  Brian Hyatt Chen Li  William Glover
CDC  John Besser, Eija Trees, Duncan MacCannell and colleagues
National Institutes of Health
David Lipman (NCBI)  Martin Shumway (NCBI)
Tatiana Tatusova (NCBI)  William Klimke (NCBI)
Illumina
Lisa Alves  Susan Knowles  Omayma Al-Awar and colleagues
CLC Bio  David Michaels  Cecilia Boysen and colleagues
Questions