How Whole Genome Sequencing is Augmenting FDA’s Role in Foodborne Outbreak Investigation

Eric L. Stevens, Ph.D.
FDA Commissioner’s Fellow
U.S. Food & Drug Administration
Center for Food Safety and Applied Nutrition
Office of Regulatory Science
Division of Microbiology
College Park, MD 20740
Safe, Wholesome, Sanitary Foods
The fresh-cut tomato supply chain
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.

A “One Health” approach to food safety—bringing together expertise and resources from the clinical, veterinary, wildlife health, and ecology communities—has the potential to reveal the sources, pathways, and factors driving the outbreaks of foodborne illness and possibly prevent them from occurring in the first place.

NOTE: Countries are listed in alphabetical order and not by volume of export.
Why develop a WGS-based Network?

- Tracking and tracing of food pathogens
  - Global & Domestic
  - Faster identification of the food involved in the outbreak
- Limited number of investigators vs. facilities and import lines
- Insufficient resolution of current tools
  - matching clinical to environmental samples
PFGE discrimination level

- Molecular technique that uses restriction enzymes to cut DNA at specific locations
- The resulting pieces are then run on a gel and separate according to size (with the smaller fragments moving further).
- Use the banding pattern of the gel to group bacterial isolates
WGS: Whole-genome sequencing

- PFGE only gives us information at a restriction site based on the banding pattern
- WGS has the ability to give us information at every position in the bacterial genome
WGS: Very brief overview

Bacterial Genome

Amplify genomic DNA and prepare for sequencing chemistry

“Reads”
(thousands to millions)
(range from 300 bp to 40 kb)

Assembly Method (many to choose from)

Aligned reads

Derived Genome Sequence
WGS:
Not all of the genomic sequence is needed

In fact, we expect many of the sites to be the same

SNP = Single Nucleotide Polymorphism

\[
\begin{align*}
\text{TTGTTGGGCAG} \\
\text{ATGTAGGCAG} \\
\text{ATGTTGGGCAG} \\
\text{ATGTTGGGCAC}
\end{align*}
\]

4 aligned genome sequences reduced to 3 SNPs
Clusters similar sequences together under the assumption that it follows evolutionary history.

WGS Phylogenetic tree
FDA’s WGS program

- Most basic application is identifying pathogens from food or environmental samples and then comparing it to clinical isolates

- WGS allows better resolution than PFGE in differentiating related versus unrelated strains
  - Key for epidemiological insights

- However, one of the most promising public health benefits may come from the combination of a pathogen’s genomic information with its geographic location
FDA’s GenomeTrakr

• First distributed network of labs to use whole genome sequencing
• Consists:
  • 14 federal labs
  • 15 state health and university labs
  • 1 U.S. hospital lab
  • 9 international labs
  • Independent researchers
• Data curation and bioinformatic support/analyses provided by National Center for Biotechnology Information (NCBI)
Total Number of Sequences in the GenomeTrakr Database

- **Salmonella**
- **Listeria**
- **E. coli / Shigella**
- **Campylobacter**

Average Number of Sequences Added Per Month:
- **2013**: 169
- **2014**: 1,076
- **2015**: 2,362

First sequences uploaded in Feb 2013

Public Health England uploads more than 8,000 Salmonella sequences
# Minimal pathogen metadata

<table>
<thead>
<tr>
<th>sample_name</th>
<th></th>
<th>collection_date</th>
<th><strong>What</strong></th>
<th></th>
<th><strong>When</strong></th>
<th><strong>Where</strong></th>
<th><strong>Who</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td>organism</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>strain/isolate</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Category (attribute_package)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1a) Clinical/Host-associated</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1a1) specific_host</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1a2) isolation_source</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1a3) host-disease</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>OR</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1b) Environmental/Food/Other</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1b1) isolation_source</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>OR</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6a) geo_loc_name</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>OR</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6b) lat_lon</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6a) geo_loc_name</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>OR</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6b) lat_lon</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>collected by</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Where</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
NCBI pathogen detection pipeline

NCBI Submission Portal

NCBI Pathogen Pipeline
- QC
- Kmer analysis
- Genome Assembly
- Genome Annotation
- Genome Placement
- Clustering
- SNP analysis
- Tree Construction

Reports

BioProject
BioSamples
SRA
GenBank

Institutions:
- USDA
- MDH
- Florida Health
- CDC
- New York State
- Public Health England
- fera
- Brigham and Women's Hospital

FDA

NCBI

U.S. Food and Drug Administration
Protecting and Promoting Public Health
www.fda.gov
Daily phylogenetic trees

NCBI Kmer tree
Listeria monocytogenes
Example 1: Global source traceback

• In 2012, 410 individuals in the U.S. became sick from ingesting food that contained S. Bareilly

• CDC investigated a multistate (29 states) outbreak

• The illnesses were linked to frozen raw yellowfin tuna (Nakaochi Scrape) which was imported from India

• FDA examined the WGS of these isolates and compared them to previous S. Bareilly events
S. Bareilly Phylogeny
NGS distinguishes geographical structure among closely related *Salmonella* Bareilly strains
WGS supports preventative controls

- Permits deep dive to solve persistent/complex problems in a facility or on a farm
- Comparison of internal WGS results to public database of food/environmental isolates
- Public software and analysis tools readily available to industry for viewing of results
Example 2:
S. Braenderup 2014 pre-outbreak

• In 2014, FDA conducted baseline environmental sampling in nut butter processing facilities

• A few of the samples tested positive for S. Braenderup and a PFGE pattern matched several cases of recent salmonellosis without a common link

• WGS was performed on both environmental and clinical isolates and found to be extremely close (2 SNP differences)
S. Braenderup 2014 pre-outbreak

- reference
- CFSAN023486 FL clinical NA
- CFSAN012866 TN clinical 1/28/14
- CFSAN023961 TX clinical 4/24/14
- CFSAN012311 environmental 2/25/14
- CFSAN022904 environmental 7/16/14
- CFSAN012310 environmental 2/25/14
- CFSAN022903 environmental 7/16/14
- CFSAN023487 CT clinical 3/18/14
- CFSAN012309 environmental 2/25/14
- CFSAN023485 TX clinical 6/5/14
- CFSAN012867 IA clinical 3/4/14
- CFSAN023960 NM clinical 1/22/14

- env. swab
- clinical
Representative* timeline for traditional approach (PFGE)

Contaminated food enters commerce

Identify illnesses and get PFGE pattern from clinical samples

Identify contaminated food and confirm that product or environmental sample PFGE pattern matches the clinical sample pattern

Source of contamination identified too late to prevent most illnesses

* Data is for illustrational purposes and does not represent an actual outbreak
Representative* timeline using WGS

FDA, CDC, FSIS, and States use WGS in real-time and in parallel on clinical, food, and environmental samples

Source of contamination identified early through WGS combined database queries

*Data is for illustrational purposes and does not represent an actual outbreak
Immediate impacts of WGS to industry, growers, and distributors, countries, states.

• Earlier intervention means:
  
  • Reduced amount of recalled product
  
  • Fewer sick patients which means fewer lawsuits
  
  • Less impact overall and minimal damage to brand recognition
Acknowledgements

- **FDA**
  - Center for Food Safety and Applied Nutrition
  - Center for Veterinary Medicine
  - Office of Regulatory Affairs

- **National Institutes of Health**
  - National Center for Biotechnology Information

- **State Health and University Labs**
  - Alaska
  - Arizona
  - California
  - Florida
  - Hawaii
  - Maryland
  - Minnesota
  - New Mexico
  - New York
  - South Dakota
  - Texas
  - Virginia
  - Washington

- **USDA/FSIS**
  - Eastern Laboratory

- **CDC**
  - Enteric Diseases Laboratory

- **INEI-ANLIS “Carolos Malbran Institute,” Argentina**

- **Centre for Food Safety, University College Dublin, Ireland**

- **Food Environmental Research Agency, UK**

- **Public Health England, UK**

- **WHO**

- **Illumina**

- **Pac Bio**

- **CLC Bio**

- **Other independent collaborators**