Using Whole Genome Sequencing to Source Foodborne Outbreaks

Eric L. Stevens, PhD
U.S. Food & Drug Administration
Center for Food Safety and Applied Nutrition
Both Houses of Congress pass a bill. President signs and bill becomes law.

Agency writes regulations/rules based on law, then enforces the new regulations/rules. New regulations/rules have the force of law.

Agency writes guidance documents to explain the rules. Guidance provide recommendations and do not have the force of law.
U.S. Food and Drug Administration

Center for Food Safety and Applied Nutrition

Mission Statement

CFSAN, in conjunction with the Agency's field staff, is responsible for promoting and protecting the public's health by ensuring that the nation's food supply is safe, sanitary, wholesome, and honestly labeled, and that cosmetic products are safe and properly labeled.
U.S. Food and Drug Administration

Center for Food Safety and Applied Nutrition

Scope of Responsibility

• All foods except meat, poultry, processed eggs, and catfish
• 25% of consumer spending in the U.S. is on products regulated by FDA
  • 75% of this amount is spent on foods
• CFSAN regulates $417 billion domestic foods, $49 billion imported foods, $60 billion cosmetics sold in interstate commerce
• 377,000 registered food facilities (154,000 domestic and 223,000 foreign)
  • Excludes restaurants, institutional food service, and supermarkets/grocery stores
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
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.
The fresh-cut tomato supply chain

Contamination occurs at farm

Represents 1 SNP
Goal of Phylogenetic Trees using WGS Data:
Infer evolutionary relationships based on nucleotide differences
And match clinical to food/environmental isolates
FDA’s GenomeTrakr

• First distributed network of labs to use WGS:
  - 25 US State health and university labs
  - 15 federal labs
  - 1 US hospital lab
  - 20 international labs

• Focuses on environmental and food isolates to be a large reference database for clinical samples

• Each genomic sequence has associated metadata
  - Species
  - Geographic location (within US has states, outside US has countries)
  - Date of isolation and Collector
  - Isolation source (food product, environmental swab)

www.fda.gov
### Minimal pathogen metadata

**What**
- sample_name
- organism
- strain/isolate

**When**
- collection_date

**Where**
- Geographic location
  - geo_loc_name
  - lat_lon

**Who**
- collected by

### Category (attribute_package)
- Clinical/Host-associated
  - specific_host
  - isolation_source
  - host-disease
- Environmental/Food/Other
  - isolation_source

---

www.fda.gov
NCBI pathogen detection pipeline

- NCBI Submission Portal
  - QC
    - Kmer analysis
  - Genome Assembly
  - Genome Annotation
  - Genome Placement
  - Clustering
  - SNP analysis
  - Tree Construction

Reports

www.fda.gov
Basic Data Flow for Global WGS Public Access Databases

**DATA ACQUISITION**
Sequence and upload genomic and geographic data

**DATA ASSEMBLY, ANALYSIS, AND STORAGE**
International Nucleotide Sequence Database Collaboration (INSDC)
Shared Public Access Databases
- NCBI – National Center for Biotechnology Information
- EMBL – European Molecular Biology Laboratory
- DDBJ – DNA Databank of Japan

**PUBLIC HEALTH APPLICATION AND INTERPRETATION OF DATA**
- Find clinical links
- Identify clusters
- Conduct traceback
- Develop rapid methods
- Develop culture independent tests
- Develop new analytical software
Total Number of Sequences in the GenomeTrakr Database

Average Number of Sequences Added Per Month Each Year

- 2013: 169
- 2014: 1,076
- 2015: 2,362
- 2016: 4,529
- 2017: 5,826
- 2018: 9,434

First sequences uploaded in February 2013

* Other pathogens: Cronobacter, V. vulnificus, C. botulinum, and C. perfringens
Using NCBI to Access Data

Pathogen Detection

NCBI Pathogen Detection integrates bacterial pathogen genomic sequences originating in food, environmental sources, and patients. It quickly clusters and identifies related sequences to uncover potential food contamination sources, helping public health scientists investigate foodborne disease outbreaks.

Find Isolates now!

Explore the Data

<table>
<thead>
<tr>
<th>Species</th>
<th>New Isolates</th>
<th>Total Isolates</th>
</tr>
</thead>
<tbody>
<tr>
<td>Salmonella enterica</td>
<td>1,060</td>
<td>91,365</td>
</tr>
<tr>
<td>E.coli and Shigella</td>
<td>117</td>
<td>35,293</td>
</tr>
<tr>
<td>Listeria monocytogenes</td>
<td>11</td>
<td>15,717</td>
</tr>
<tr>
<td>Campylobacter jejuni</td>
<td>26</td>
<td>12,600</td>
</tr>
</tbody>
</table>

Data Resources

- Isolates Browser
- Antimicrobial resistance reference gene database
- Isolates with antibiotic resistant phenotypes
- Beta-lactamase resources
- Download analysis results (FTP)
Why Open Data?

Why submit?

- To contribute information on pathogen sequences that may help discover sources of contamination or help in solving outbreaks more quickly
- To provide valuable real-time information about the relationship of an isolate to other isolates and outbreaks
- To enhance the set of pathogen genomes that can be used by the scientific community
- To supply information on the set of resistance genes present in a pathogen

The NCBI Pathogen Detection system is built on the foundation of open data. Data are intended to be submitted and released to the public immediately. Currently four major foodborne pathogens (Campylobacter, Escherichia coli and Shigella, Listeria, and Salmonella) are being analyzed in real time as participating public health agencies submit the sequences. Pathogens are also being analyzed for antimicrobial resistance.
Human migration patterns

https://commons.wikimedia.org/wiki/File:BlankMap-World_1935.png
Understand human genetic variation and create reference populations
Using public records to recreate my ancestry (phylogeny)
Using public records to recreate my ancestry (phylogeny)
DNA plus metadata tell the same story!
Why can’t we do the same for microbes?
Imagine sampling and sharing microbial DNA from all over the world
Building reference populations for Salmonella
Building reference populations for Listeria
Building reference populations for different foods
Building reference populations for different infections
Building reference populations for ANYTHING
And making the data available to everyone!
Example:
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)
Outbreak/Pre-outbreak summer of 2014

A Growing Regulatory Role:
Improving Food Safety

1. Stop Contamination
   • Food Safety Modernization Act (2010) – Preventive Controls, Improve Industry Practices

2. Identify Source of Foodborne Outbreaks More Quickly
   • Whole-Genome Sequencing (WGS) surveillance of bacterial pathogens, environmental testing

3. Resolve Sporadic Illnesses
   • Low level contamination events, Whole-Genome Sequencing (WGS) surveillance of bacterial pathogens, environmental testing
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
Regular testing throughout network

• Identifies specific suppliers that are introducing contaminants

• Identifies whether contaminant is resident to a facility or transient

• Knowledge of where contaminant is coming from allows industry to fix the problem based on scientific evidence.
Importance of a Balanced Approach

Clinical Samples

Maximum WGS Benefit

Food and Environmental Samples

20 years of PulseNet USA
1996 - 2016
**US PulseNet** members collaborating around foodborne outbreaks

- **Clinical** and some food/env isolates submitted to NCBI
- Separate DB with full metadata

**NCBI’s Pathogen Detection**
- Contains ALL WGS data generated from PulseNet and GenomeTrakr labs.
- Public Health England, Argentina, and others.
- Produces daily trees from current SNP clusters

**Food and environmental isolates.**
- All data submitted to NCBI

**State and Local Health Agencies**

**Academic labs**

**International labs**

**Agriculture labs**

**commercial labs**
Where WGS can be used
Note:

- These slides are for teaching purposes only and have been collected from images that I have made, from the CDC and FDA, and from around the web.

- The findings and conclusions in this report are those of the author and do not necessarily represent the official position of the Food and Drug Administration.