AN EMERGING TECHNOLOGY: THE USE OF NEXT GENERATION SEQUENCING FOR IMPROVING FOOD SAFETY

Peter Gerner-Smidt, MD, DSc, Chief
Enteric Disease Laboratory Branch
Acknowledgements

Disclaimers:
"The findings and conclusions in this presentation are those of the author and do not necessarily represent the official position of the Centers for Disease Control and Prevention"

"Use of trade names is for identification only and does not imply endorsement by the Centers for Disease Control and Prevention or by the U.S. Department of Health and Human Services."

Disclosures:
The presenter is a U.S. federal employee with no conflicts to declare
Next Generation Sequencing (NGS) ~ Whole Genome Sequencing (WGS)

The Whole Genome Sequencing (WGS) Process
WGS is a laboratory procedure that determines the order of bases in the genome of an organism in one process. WGS provides a very precise DNA fingerprint that can help link cases to one another allowing an outbreak to be detected and solved sooner.

1. DNA Extraction
Scientists take bacterial cells from an agar plate and treat them with chemicals that break them open, releasing the DNA. The DNA is then purified.

2. DNA Shearing
DNA is cut into short fragments of known length, either by using enzymes “molecular scissors” or mechanical disruption.

3. DNA Library Preparation
Scientists make many copies of each DNA fragment using a process called polymerase chain reaction (PCR). The pool of fragments generated in a PCR machine is called a “DNA library.”

4. DNA Sequencing
The DNA library is loaded onto a sequencer. The combination of nucleotides (A, T, C, and G) making up each individual fragment of DNA is determined, and each result is called a “DNA read.”

5. DNA Sequencing Analysis
The sequencer produces millions of DNA reads and specialized computer programs are used to put them together in the correct order like pieces of a jigsaw puzzle. When completed, the genome sequence containing millions of nucleotides (in one or a few large pieces) is ready for further analysis.
Next Generation Sequencing for Food Safety

---

**Metagenomics**

- Whole Genome Sequencing (WGS)
- Metabarcoding/Amplicon sequencing
- NGS Platforms
  - Illumina
  - Ion Torrent
  - PacBio
  - Oxford
  - Nanopore

**Foodomics**

- Investigative approaches using combinations of metagenomics, metatranscriptomics, metaproteomics, metabolomics
- For in depth investigation of microbial communities including analysis of microbial interactions and elucidation of microbial roles such as diversity, protein expression and metabolic functions within the entire community

---

**Single colony isolation**

**Microbial population**

Why the hype about WGS?

- WGS is transforming microbiology, replacing numerous traditional methods with one in a single efficient workflow
WGS-based foodborne pathogen surveillance workflow
Reference characterization by WGS: ’One shot’ characterization of STEC

**Genus/Species:** *Escherichia coli*

**Serotype:** O104:H4

**Pathotype:** Shiga toxin-producing and enteroaggregative *E. coli* (STEC/EAEC)

**Virulence profile:** stx2a, aggR, aggA, sigA, sepA, pic, aatA, aaiC, aap

**Sequence Type:** ST678

**Allele code:** 102.45.26.35.3

**Antimicrobial resistance genes:** $\text{bla}_{\text{TEM-1}}, \text{bla}_{\text{CTX-M-15}}, \text{strAB}, \text{sul2}, \text{tet(A)A}, \text{dfrA7}$
Subtyping
WGS superior resolution to the old standard, PFGE

Listeria Outbreak Linked to Artisan Cheese (2013)

All the same PFGE pattern
WGS detects more outbreaks than by PFGE

Number of Salmonella outbreaks detected with laboratory data in Canada

- Enteritidis
- Heidelberg
- Typhimurium

Year 2012 2013 2014 2015 2016 2017

WGS

Courtesy C. Nadon, Public Health Agency of Canada
Real-time WGS improves laboratory surveillance

*Listeria metrics*

- No. of clusters detected:
  - PFGE (1-year pre-WGS): 14
  - 3-Year average WGS: 19

- No. of clusters detected sooner or only by WGS: N/A

- No. of outbreaks solved (food source identified):
  - PFGE (1-year pre-WGS): 6.7
  - 3-Year average WGS: 6.3

- Median no. of cases per cluster:
  - PFGE (1-year pre-WGS): 2
  - 3-Year average WGS: 4

Courtesy Amanda Conrad, CDC Outbreak Response & Preparedness Branch
Foodborne infections is a global issue

- A foodborne infection on one continent may have its source on a different continent
- International outbreaks are common
Challenges to WGS for global food safety

- No international quality standards exist
- No international consensus on the use of specific pipelines
- Limited data sharing
The PulseNet solution to use WGS for foodborne disease surveillance in the global context

Fast, precise, simple communication and easy sharing of data is key in outbreak investigations

- Standardized/harmonized and validated quality standards and analysis
- Results in standardized format
- Low volume format
  - to accommodate slow internet speeds
  - no need to go back to raw data
- Solutions must be PRACTICAL and NOT necessarily PERFECT
  ‘If it works, it is good enough’
- Encourage public sharing of raw sequence data
Data sharing

- WGS data should be publicly available in real time
  - SRA, ENA and the DNA Data Bank of Japan
  - Minimum epidemiological data – time, place and type of isolate

- Barriers
  - Ethics: Personal identifiable information
  - Intellectual property and other legal issues

- Food industry concerns
  - No “statute -of-limitations” on liability
  - No precise definition of “outbreak ”
  - No international interpretation standards
  - Trade implications
Other concerns with WGS

• WGS turnaround time issues
  • Still long (~ 7 work days)
• Cost
• Cluster triage
  • Not resources to investigate all outbreaks
  • Which should be investigated?
• Culture-independent diagnostic testing (CIDT)
  • We are losing the isolates!
Next Generation Sequencing for Food Safety

NGS Platforms
- Illumina
- Ion Torrent
- PacBio
- Oxford
- Nanopore

Metagenomics
- Targeted amplification of marker/phylogenetic genes eg. 16S rDNA of mixed microbial communities
- For culture independent diagnostics, identification & subtyping of specific pathogens; study of microbial diversity of food/environment over time; food fraud detection, quality control of fermentation, identification of spoilage bacteria in food and environment

Metabarcoding/Amplicon sequencing
- Whole Genome Sequencing (WGS)
- SNP
- Gene by gene-based

Foodomics
- Investigative approaches using combinations of metagenomics, metatranscriptomics, metaproteomics, metabolomics
- For in depth investigation of microbial communities including analysis of microbial interactions and elucidation of microbial roles such as diversity, protein expression and metabolic functions within the entire community

Metagenomics

**Targeted sequencing:**
* Amplicon sequencing
* ‘Metabarcoding’
  - Targeted diagnostic and subtyping
  - Population studies
  (30,000 feet view’)

**Random Shotgun Sequencing:**
* ‘Metagenomics’
  - Unbiased diagnostics & subtyping
  - Population studies
  (‘microscope view’)

**DNA preparation**

**Food/Environmental specimen**

**Stool/clinical sample**
The stool microbiome

- **Microbes**
  - Bacteria
  - Viruses
  - Parasites
  - Fungi

- **Other cells**
  - Human
  - Food Animals
  - Plants

Science. 336:8 1246-247
What is a healthy microbiome?

- Three ‘enterotypes’

The gut microbiome of inflammation

- *Proteobacteria/Enterobactericeae/E.coli* are enriched

What does the microbiome tell us?

- No microbiomes are identical
- The human microbiome composition depends on
  - the host
  - health status
  - diet
  - and much more….
- But…..
  - What is cause and effect?
    - Are microbiome changes causing illness or are they a consequence of illness?
Subtyping WITHOUT culture
Approaches to sample/specimen based subtyping

Two CDC examples

Amplicon sequencing

1,000’s of species specific MLST targets

Shotgun metagenomics

• Unbiased sequencing with identification of strain specific signatures
Amplicon sequencing approach for subtyping

Primer design pipeline

ATTTCCGAAACG
Resolution similar between amplicon sequencing and culture based WGS approach for *Salmonella* subtyping
Metagenomics case study:

- Two similar 2013 *Salmonella enterica* outbreaks in Colorado and Alabama
- Originally distinguished using epidemiological and isolate genomic data

- Unbiased sequencing
Isolate based WGS tree

Orange: Colorado
Purple: Alabama

0-5 SNVs
Median: 2.5

1-95 SNVs
Median: 79

0 SNVs
Reference-free metagenomics tree

Loopy
Taciturn
Usual
Dizzy
Florid

Ethereal
Unnamed
Resplendent
Kickstart
Fortitude
Acidic

0 SNVs

0-34 SNVs
Median: 34

Orange: Colorado
Purple: Alabama

NC_011083
Big Data will improve food safety

- Pathogen characterization direct-from-specimen (faster) - Metagenomics
- Linking to other data from different sources = More information to inform policy

But
- Privacy issues
- Regulatory hurdles
- Data capacity issues
Conclusions

- WGS is the new gold standard tool for characterization and tracking of foodborne microorganisms
- WGS of clinical and food specimens shows promise to replace isolate based methods in the future
- Early days of research in the human (and food) microbiome
- Sequencing data will become an essential part of the Big Data to improve food safety (and much more)
- Improvements in sequencing, informatics and bioinformatics technologies are needed
The Food & Stool Metagenomics App & The Ultra-Portable Sequencer

Real-time metagenomics sequencing:
Data for Action!